UNIVERSITY OF NATIONAL AND WORLD ECONOMY

*The Spirit Makes The Power*

# 4TH INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION

## ICAICTSEE – 2014

***October 24 – 25th, 2014***

***University of National and World Economy***

***Sofia, Bulgaria***

# CONFERENCE PROCEEDINGS

**UNDER THE AUSPICES OF:**

ifip the leading edge of information technology

**PUBLICATIONS INDEXED BY:**

EBSCO HOST

***and***

ProQuest

**SPONSORS AND PARTNERS:**

MusalaSoft кабинама SAP ITHEA IDNC

# PROCEEDINGS OF THE

## INTERNATIONAL CONFERENCE ON APPLICATION OF

## INFORMATION AND COMMUNICATION TECHNOLOGY AND

## STATISTICS IN ECONOMY AND EDUCATION

## ICAICTSEE-2014


**EDITOR:**

**PROF. DR. DIMITER G. VELEV**

**DEPT. OF INFORMATION TECHNOLOGIES AND COMMUNICATIONS**

**UNIVERSITY OF NATIONAL AND WORLD ECONOMY**

**UNSS – STUDENTSKI GRAD**

**1700 SOFIA, BULGARIA**

***DGVELEV@UNWE.BG***



**ISSUED FOR PUBLICATION: SEPTEMBER 15TH 2015**

**SIZE: JIS-B5**

**QUIRES: 46**

**VOLUME: 100**

**ISSN 2367-7635 (PRINT)**

**ISSN 2367-7643 (ONLINE)**

# CONFERENCE COMMITTEES

## CONFERENCE CHAIR

| | | |
|---|---|---|
| Dimiter G. **Velev** | University of National and World Economy | ***Bulgaria*** |

## PROGRAM COMMITTEE

| | | |
|---|---|---|
| César Correa **Arias** | Universidad de Guadalajara | ***Mexico*** |
| Saji **Baby** | GEO Environmental Consultation | ***Kuwait*** |
| Victor **Blagodatskikh** | Moscow State University of Economics, Statistics and Informatics | ***Russia*** |
| Helena **Blazun** | University of Maribor | ***Slovenia*** |
| Robert **Brumnik** | Institute for Corporate Security Studies | ***Slovenia*** |
| Gaetano **Cascini** | Politecnico di Milano | ***Italy*** |
| Donald T. **Chang** | Metropolitan State College of Denver | ***USA*** |
| Sérgio Manuel Serra da **Cruz** | Universidade Federal Rural do Rio de Janeiro | ***Brazil*** |
| Tadeusz **Czachorski** | Institue of Theoretical and Applied Informatics, PAS | ***Poland*** |
| Vladimir **Dimitrov** | Sofia State University | ***Bulgaria*** |
| Ivan **Drogobytsky** | Fiancial University under the Goverment of the Russian Federation | ***Russia*** |
| Francisco Patricio **Esteves** | Higher Institute of Technology and Sciences | ***Angola*** |
| Vasil **Georgiev** | Sofia State University | ***Bulgaria*** |
| Valentin **Goev** | University of National and World Economy | ***Bulgaria*** |
| Bernard **Grabot** | University of Toulouse | ***France*** |
| Christopher Khoo Soo **Guan** | Nanyang Technological University | ***Singapore*** |
| James K. **Ho** | Iniversity of Illinois at Chicago | ***USA*** |
| Aytekin **Isman** | Sakarya University | ***Turkey*** |
| Sergey **Ivanov** | University of the District of Columbia, Washington, DC | ***USA*** |
| Masayasu **Kanno** | Kanagawa University | ***Japan*** |
| Madjid Malikovich **Karimov** | Tashkent University of Information Technologies | ***Uzbekistan*** |
| Sergii **Kavun** | Kharkiv Institute of Banking | ***Ukraine*** |
| Alexander **Khoroshilov** | UNESCO Institute for Information Technologies in Education | ***Russia*** |
| Leszek F. **Korzeniowski** | European Association for Security | ***Poland*** |
| Hee Dong **Kim** | Hankuk University of Foreign Studies | ***S.Korea*** |
| Valentin **Kisimov** | University of National and World Economy | ***Bulgaria*** |

| | | |
|---|---|---|
| Peter **Kokol** | University of Maribor | ***Slovenia*** |
| George **Kovacs** | Computer and Automation Institute | ***Hungary*** |
| Oleksiy E. **Kovalenko** | Institute of Mathematical Machines and Systems Problems, NASU | ***Ukraine*** |
| Anatoly **Krapivensky** | Volgograd branch of Moscow University of Fiance and Law | ***Russia*** |
| Vinod **Kumar** | Sprott School of Business, Carlton University | ***Canada*** |
| Andrew **Kusiak** | The University of Iowa | ***USA*** |
| Ronald S. **Laura** | University of Newcastle | ***Australia*** |
| Patrick **Letouze** | Universidade Federal do Tocantins | ***Brazil*** |
| Havar Amir oglu **Mammadov** | Azerbaijan Technical University | ***Azerbaijan*** |
| Violeta **Manevska** | St. Kliment Ohridski University - Bitola | ***Macedonia*** |
| Christopher **McMahon** | University of Bristol | ***UK*** |
| Pedja M. **Milosavljevic** | University of Nis | ***Serbia*** |
| A.K. **Mishra** | Motilal Nehru National Institute of Technology | ***India*** |
| Igor **Mladenovic** | University of Nis | ***Serbia*** |
| Alexander N. **Moiseev** | Tomsk State University | ***Russia*** |
| Balakrishnan **Muniandy** | Universiti Sains Malaysia | ***Malaysia*** |
| Valentin V. **Nechaev** | Moscow State Institute of Radiotechnics, Electronics and Automatics | ***Russia*** |
| Tetsuo **Noda** | Shimane University | ***Japan*** |
| Serghei **Ohrimenco** | Laboratory of Information Security, AESM | ***Moldova*** |
| Gustav **Olling** | IFIP TC5 | ***USA*** |
| K.R. **Pardasani** | Maulana Azad National Institute of Technology | ***India*** |
| Veselka **Pavlova** | University of National and World Economy | ***Bulgaria*** |
| David Nadler **Prata** | Universidade Federal do Tocantins | ***Brazil*** |
| A. M. **Rawani** | National Institute of Technology | ***India*** |
| Volodymyr **Shkir** | National University Lviv Polytechnic | ***Ukraine*** |
| Yixun **Shi** | Bloomsburg University of Pennsylvania | ***USA*** |
| Santhidran **Sinnappan** | University Tunku Adbul Rahman | ***Malaysia*** |
| Kamelia **Stefanova** | University of National and World Economy | ***Bulgaria*** |
| Karen **Starr** | Alfred Deakin Research Institute | ***Australia*** |
| Olexandr **Trydid** | Kharkiv Institute of Banking | ***Ukraine*** |
| A.F. **Verlan** | Pukhov Institute for Modelling in Energy Engineering, NASU | ***Ukraine*** |
| Hsin-Hung **Wu** | National Changhua University of Education | ***Taiwan*** |
| Shi **Yizhe** | ShenYang University of Chemical Technology | ***China*** |
| Parviz Ali **Zade** | OKAN University | ***Turkey*** |
| Milan **Zorman** | University of Maribor | ***Slovenia*** |

## LOCAL ORGANIZING COMMITTEE

| | | |
|---|---|---|
| Emil **Denchev** | University of National and World Economy | ***Bulgaria*** |
| Rosen **Kirilov** | University of National and World Economy | ***Bulgaria*** |
| Mihail **Konchev** | University of National and World Economy | ***Bulgaria*** |
| Vania **Lazarova** | University of National and World Economy | ***Bulgaria*** |
| Plamen **Milev** | University of National and World Economy | ***Bulgaria*** |
| Alexandrina **Murgeva** | University of National and World Economy | ***Bulgaria*** |
| Monika **Tzaneva** | University of National and World Economy | ***Bulgaria*** |
| Ilko **Velikov** | University of National and World Economy | ***Bulgaria*** |

# CONTENTS

*Dear Colleagues,*

*On behalf of the Conference Committees I have the honor and pleasure to thank you all for your true participation in the work of the International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2014), http://icaictsee-2014.unwe.bg/, which took place on October 24 – 25th, 2013 at the University of National and World Economy (UNWE), Sofia, Bulgaria.*

*I would like specially to express my gratitude to all of the program committee members for their genuine support without which it would never had happened.*

*Due to the current state of scientific development in all spheres of human activity, the constant knowledge and skill actualization of the academia and researchers in the field of Information and Communication Technology (ICT) is an obligatory necessity, especially when the world is in a global economic and financial crisis. The definition of long-term scientific research tasks in this area has a strategic importance, which is even truer for young researchers, lecturers and Ph.D. students.*

*The conference aims were conducting analysis of the current problems and presenting results of the ICT application in different areas of economy, education and related areas of scientific knowledge; outlining the existing possibilities for the application of modern ICT tools, methods, methodologies and information systems in economy and education; discussing advanced and emerging research trends with a long-term importance in the field of ICT application in economy and education.*

*The conference has established an academic environment that fosters the dialogue and exchange of ideas between different levels of academic and research communities.*

*The conference outcome is 101 published research papers (41 from foreign participants), the explosion of fresh ideas and the establishment of new professional relations.*

*I do hope to meet you all again at ICAICTSEE – 2015 ( http://icaictsee.unwe.bg/ )!*

*Dimiter G. Velev*
*ICAICTSEE-2014 Chair*

# ICAICTSEE - 2014
# KEYNOTE SPEECHES

KEYNOTE SPEECH 1

# TAIWAN HI-TECH ALLIANCE, LINKING YOU TO A WORLD OF PROFITS

**Litien CHEN**

*Director of Taiwan Trade Center, Sofia, Bulgaria*


# Country Profile

| TAIWAN | |
|---|---|
| Capital | *Taipei* |
| Location | *East Asia* |
| Land | *36,188 skm* |
| Population | *23,367,320* |
| Labor Force | *11 million* |
| Literacy | *98.3%* |

Source: Directorate-General of Budget, Accounting & Statistics, Executive Yuan; Photographer: Wu Zhi Xue, Ye Ling Zhi, Tourism Bureau

2

# Topics

I. Taiwan's Global Competitiveness
II. Taiwan's Foreign Trade and Economy
III. Taiwan connected to the World

3

## I. Taiwan's Global Competitiveness



4

# Taiwan's Competitiveness

***Importance of Taiwan's Technology and Innovation in the World Market***

| | | |
|---|---|---|
| WORLD ECONOMIC FORUM | WEF Global Competitiveness Report 2012-2013—State of Cluster Development | 1st |
| IMD | 2012 IMD World Competitiveness Yearbook Entrepreneurial Spirit | 1st |
| E·I·U The Economist Intelligence Unit | EIU: 2009-2013 Innovation Competitiveness Report (among 55 countries) | 6th (2nd in Asia) |
| | BERI: Investment Environment Risk Assessment Report 2012 (among 50 countries)<br>Business Environment Risk Intelligence (BERI)<br>Investment Environment Risk Assessment Report 2013-II<br>3rd in Profit Opportunity Recommendation (among 50 countries); 2nd in Asia | 3rd (2nd in Asia) |

5

# Taiwan's Industry Clusters

- Automobiles & Motorcycles
- Automobile Parts
- Bicycles
- Ceramics & China
- Clothing/Fabrics
- Dye Industry
- Fabrics/Cloths
- Faucets
- Fishnets
- Floriculture
- Foods
- Glasses


- Hats
- Hosiery
- Machinery
- Optoelectronics
- Petrochemicals
- Saxophones
- Screws
- Semiconductor
- Shipbuilding & Steel
- Synthetic Fibers & Yarns
- Towels
- Wood-carvings

6


Taiwan's Global Competitiveness
Leading Provider of ICT Products
LCD Monitors
Market Share: 71.9%
#1
IC Foundry
Market Share: 69.3%
#1
Motherboard
(System & Pure MB)
Market Share: 95.6%
#1
Servers
(System & Pure MB)
Market Share: 88.7%
#1
Notebook
Market Share: 94.4%
#1
Netbook
Market Share: 88.3%
#1
WLAN CPE
Market Share: 67.3%
#1
Cable Modem
Market Share: 81.9%
#1
Blank Optical Disk
Market Share: 57%
#1
Source: IDB, Taiwan 2011
7


Taiwan offers...
Think Taiwan, Think High Tech
•World's top supplier of ICT products
•Strategic & emerging industries (WiMAX, Nanotech, Biotech, FPD, Semiconductor…)
Access to Fast Emerging Markets
•Access to Greater China & ASEAN markets
R&D Capability
•R&D centers include: Intel, IBM, Ericsson, MS, NEC, Dell, Motorola, Dow, Fujitsu, Sony, others
Industry Clusters & Integrated Supply Chain Network
•Strong & efficient vertical integration
•Comprehensive manufacturing assembly lines
High Quality Human Resources
•Abundance of highly educated professionals
• Business friendly working environment
8

## II. Taiwan's Foreign Trade and Economy

9

| Economic Facts | | |
|---|---|---|
| GDP | US$ 489 Billion | (2013) |
| Economic Growth Rate | 2.98% | (2013) |
| GDP per capita | US$ 21,557 | (2013) |
| Foreign Trade | US$ 305.8 Billion (Export) | (2013) |
| | US$ 268.5 Billion (Import) | (2013) |
| Foreign Exchange Reserves | US$ 415.6 Billion | (2013) |

Source: Department of Statistics, MOEA

10

Unit: Billion USD


| | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|
| Export | 198.43 | 224.02 | 246.68 | 255.63 | 203.67 | 274.6 | 308.30 | 301.20 |
| Import | 182.61 | 202.70 | 219.25 | 240.45 | 174.37 | 251.24 | 281.61 | 270.50 |

Source: MOEA

11

## Taiwan's Foreign Trade and Economy

Economic Breakthroughs | 6 Emerging Industries | 4 Emerging Intelligent Industries | 10 Key Service Industries | Taiwan Excellence Award 2011 | Taiwan Top 20 Global Brands

### Deregulation

Improved relations across the Taiwan Strait & deregulations for FDI from China since July 1, 2009.

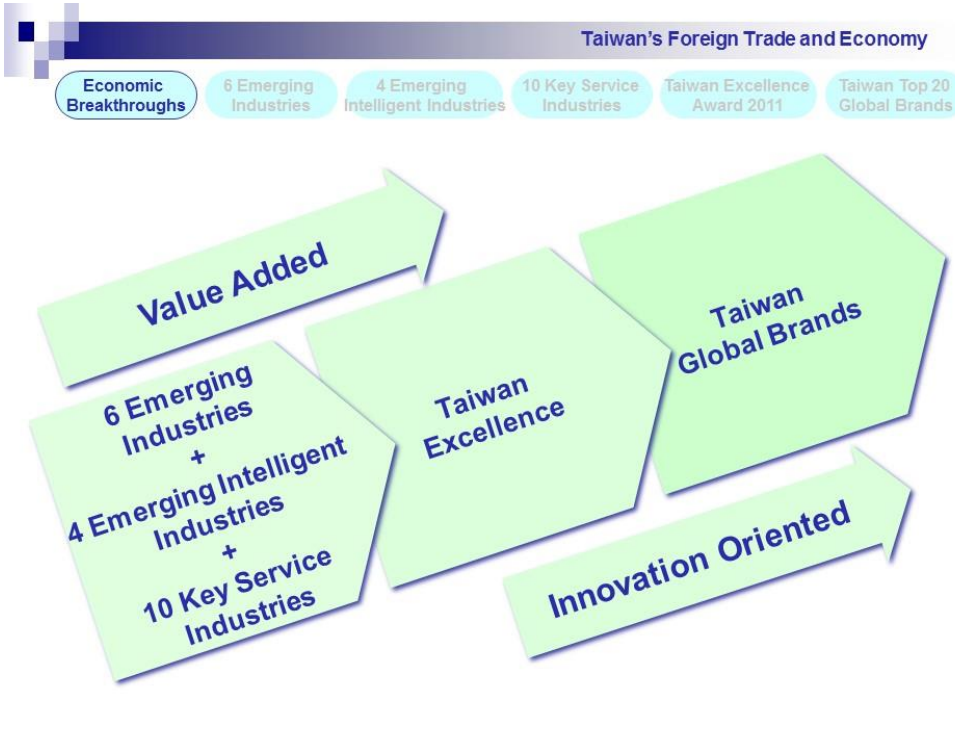
### GPA

In July 2009, Taiwan officially became the 41st member of the WTO-GPA (Government Procurement Agreement)

### FTA & ECFA

- FTA: Panama (Jan 2004), Nicaragua (June 2006), Guatemala (July 2006), El Salvador (May 2007), Honduras (Mar 2008)
- ECFA (Economic Cooperation Framework Agreement) between Taiwan and China (June 2010)

12


Taiwan's Foreign Trade and Economy
Economic Breakthroughs
6 Emerging Industries
4 Emerging Intelligent Industries
10 Key Service Industries
Taiwan Excellence Award 2011
Taiwan Top 20 Global Brands
Value Added
6 Emerging Industries + 4 Emerging Intelligent Industries + 10 Key Service Industries
Taiwan Excellence
Taiwan Global Brands
Innovation Oriented
13


6 Emerging Industries
High-end Agriculture
Biotech
Tourism
Six Emerging Industries
Culture & Creation
Green Energy
Healthcare
14


4 Emerging Intelligent Industries
Intelligent Electric Vehicle
Cloud Computing
Four Emerging Intelligent Industries
Patent Industrializa-tion
Intelligent Green Building
15


10 Key Service Industries
Education
Int'l Cuisine
Int'l Medical Care
Chinese E-commerce
Ten Key Service Industries
Music & Digital Content
WiMax
Conventions & Exhibitions
Urban Renewal
Finance
International Logistics
16


Taiwan's Foreign Trade and Economy
Economic Breakthroughs
6 Emerging Industries
4 Emerging Intelligent Industries
10 Key Service Industries
Taiwan Excellence Award
Taiwan Top 20 Global Brands
TAIWAN EXCELLENCE
Quality
R&D
Design
Marketing
Touch Diamond
For more information, please visit http://brandingtaiwan.org
17


Total value for Top 20 brands amounted to US$10.64 billion in 2013.
2013 Taiwan Top 20 Global Brands 台灣國際品牌
85°C
htc quietly brilliant
SYNNEX 聯強國際
康師傅
D-Link
ADVANTECH
MAXXIS
JOHNSON
Transcend
TREND MICRO
acer
Uni-President
merida
王品集團 wowprime
GIANT
ASUS
DELTA
旺旺
CyberLink
ZyXEL
Source: www.branding-taiwan.tw
18

# Taiwan's Exceptional Design Capability

**Awards won by Taiwan in top international design competitions, including 27 gold or first-place awards**

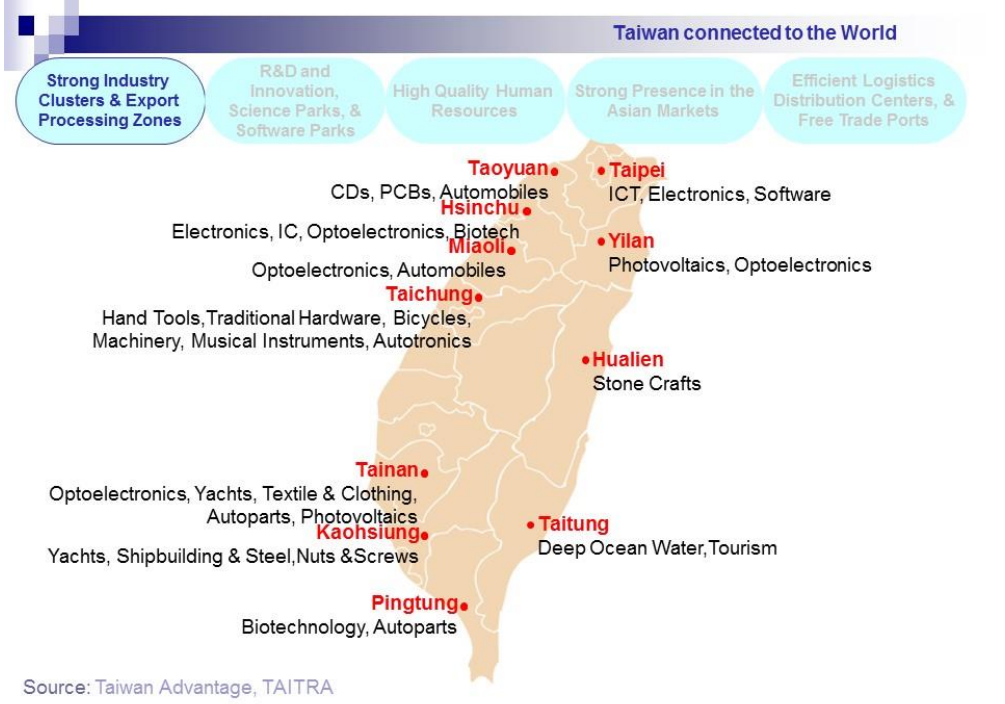
| Awards | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 合計 |
|---|---|---|---|---|---|---|---|---|---|---|
| iF, Germany (Referred to as "Design Oscar") | 5 | 14 | 37 (1 gold) | 72 (3 gold) | 56 (1 gold) | 99 (2 gold) | 66 (1gold) | 105 (3 gold) | 95 (7 gold) | **549** |
| Red Dot, Germany (Design award) | 0 | 2 | 20 (1 first) | 37 (2 first) | 45 | 66 (5 first) | 71 (5 first) | 101 (3 first) | 92 (6 first) | **434** |
| IDEA, U.S. (Industrial design award) | 0 | 1 | 5 | 1 | 4 | 5 | 2 | 6 | 5 | **29** |
| G-Mark, Japan (Product design award) | 11 | 36 | 38 | 38 | 28 | 31 | 21 | 48 | 26 | **277** |
| Total | 16 | 53 | 100 | 148 | 133 | 201 | 160 | 260 | 218 | **1289** |

19

## III. Taiwan connected to the World



20


Source: Taiwan Advantage, TAITRA

22

# Living in Taiwan

Wide range of diversified communities from different countries:
USA, Canada, Korea, Japan, India, EU, among many others.

Foreign representative offices and commercial organizations (AIT, AmCham, ECCT, KOTRA, Japan Desk, etc.)

23

## To sustain global competitiveness,

***" We want to lure as many professionals overseas as possible... "***




24

# About HiRecruit Online

- Active 2-way job matching system
- Free access to career information
- Rich pool of job openings from domestic/foreign companies in Taiwan
- Industry highlights
- Employment updates
- Work permits, resources for living in Taiwan, etc.

25

# TAIWAN WELCOMES YOU!!

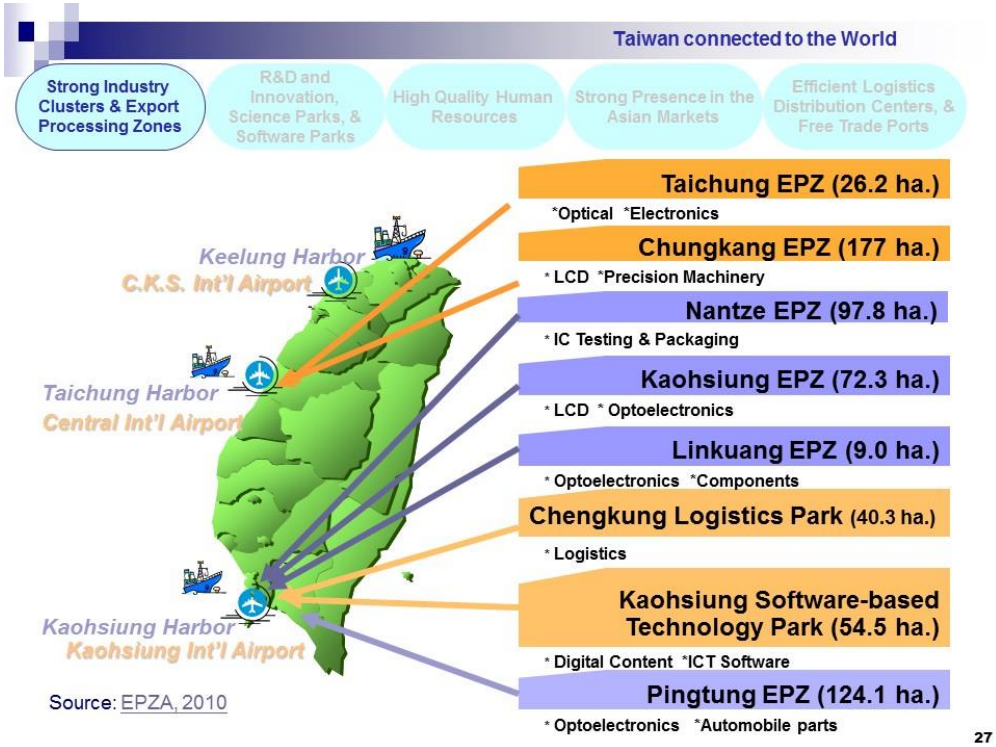
Bachelor's degree w/2+ years of experience in related fields

Masters, Ph.D., or post-graduate degree is a plus

**Talents of:**

R&D, Industrial Design, Technical Engineering, Computer Science, Materials, Physics, Chemistry(cal.), Production, MBA, Sales Channel, Brand Management, Marketing, Logistics, Finance, Accounting, Law, HR

26





**(1) U.S. Patents Granted**


Source: USPTO, 2011

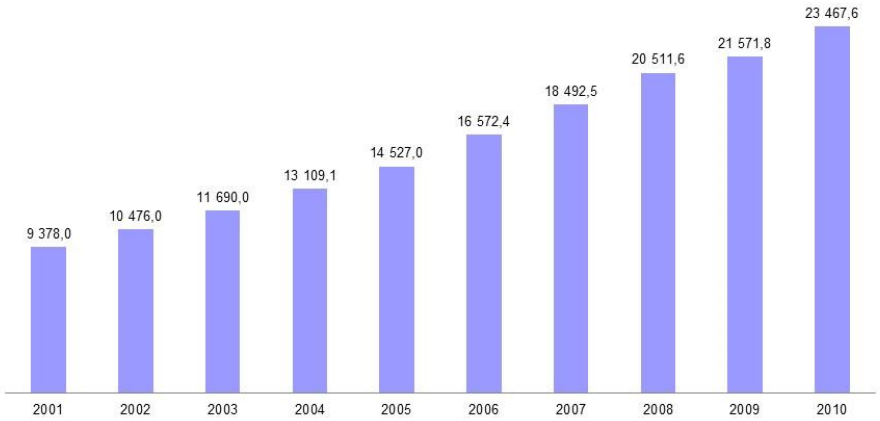
28

**Taiwan connected to the World**

Strong Industry Clusters & Export Processing Zones | **R&D and Innovation, Science Parks, & Software Parks** | High Quality Human Resources | Strong Presence in the Asian Markets | Efficient Logistics Distribution Centers, & Free Trade Ports

**(2) R&D Expenditure Comparison**

Unit: Million USD (current PPP)




Source: NSC, 2011

29

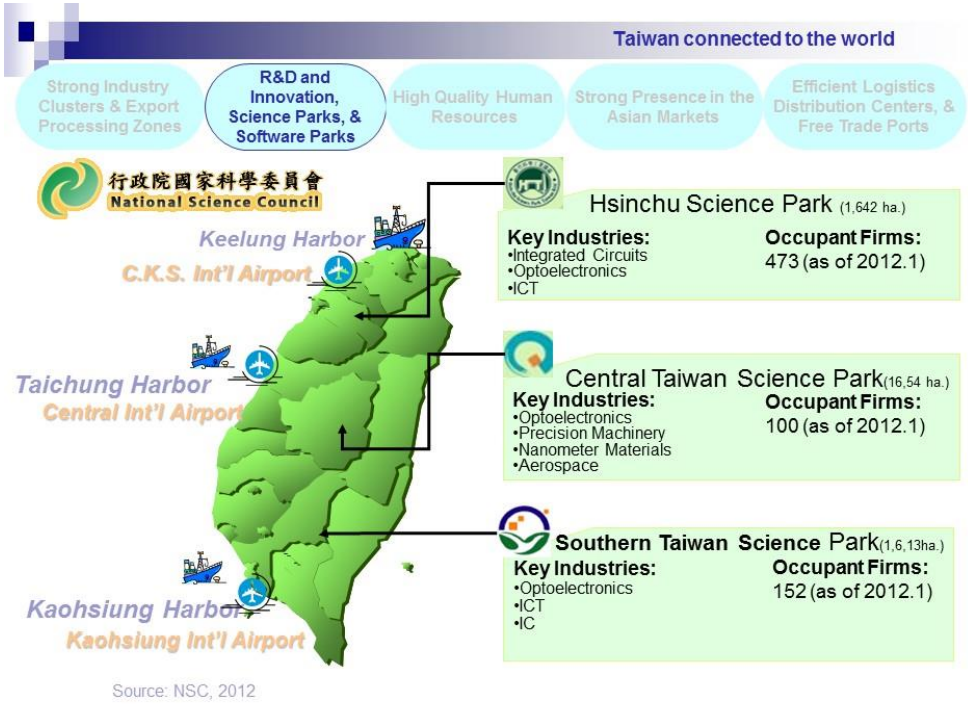
**Taiwan connected to the World**

Strong Industry Clusters & Export Processing Zones | **R&D and Innovation, Science Parks, & Software Parks** | High Quality Human Resources | Strong Presence in the Asian Markets | Efficient Logistics Distribution Centers, & Free Trade Ports

- Department of Industrial Technology, MOEA launched the "Multinational Innovative R&D Centers in Taiwan" Program in 2002 to encourage foreign enterprises to set up R&D centers in Taiwan.
- By July 2009, there were 116 incubation centers in Taiwan and the incubation policy has successfully incubated over 3,600 enterprises which helped the registration of almost 2,000 patents, the transfer of over 840 technologies, and the listing of 45 companies.
- More than 40 companies have set up 43 R&D/Innovation Centers in Taiwan.

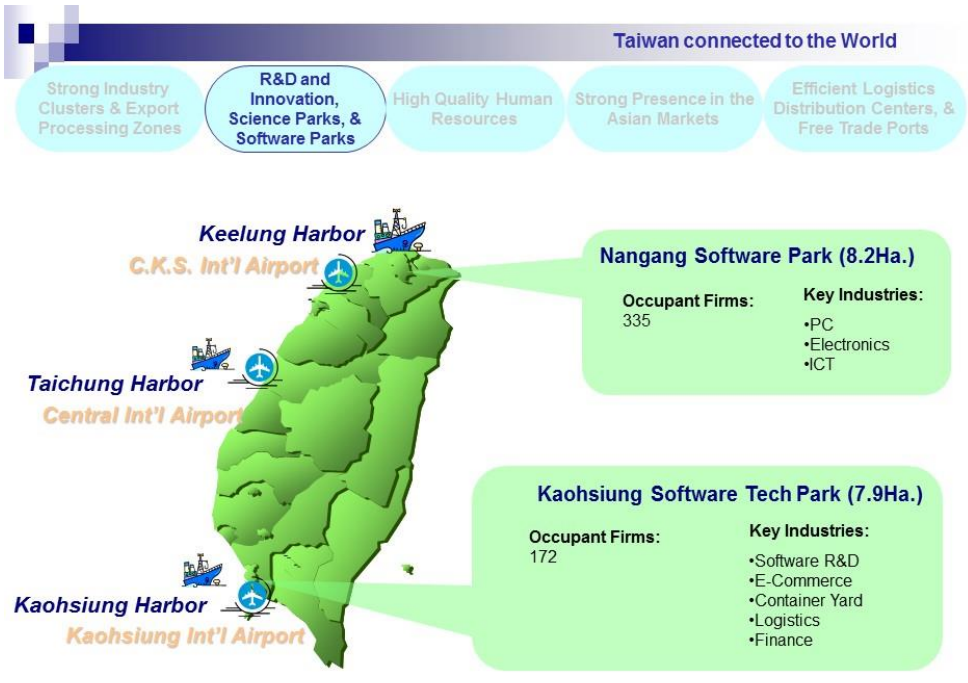
Source: IDB, 2010, DOIT, 2010, SMEA, 2010

30




Source: NSC, 2012

31




Source: NKZONE, KHSP

32

Taiwan connected to the World

Strong Industry Clusters & Export Processing Zones | R&D and Innovation, Science Parks, & Software Parks | **High Quality Human Resources** | Strong Presence in the Asian Markets | Efficient Logistics Distribution Centers, & Free Trade Ports

## *R&D Personnel*

**Researcher per 1,000 Employment in 2009**


Source: NSC, 2011

33

Taiwan connected to the World

Strong Industry Clusters & Export Processing Zones | R&D and Innovation, Science Parks, & Software Parks | **High Quality Human Resources** | Strong Presence in the Asian Markets | Efficient Logistics Distribution Centers, & Free Trade Ports

## International Marketing Talent

**Kenichi Ohmae**
(Known as "Mr. Strategy")

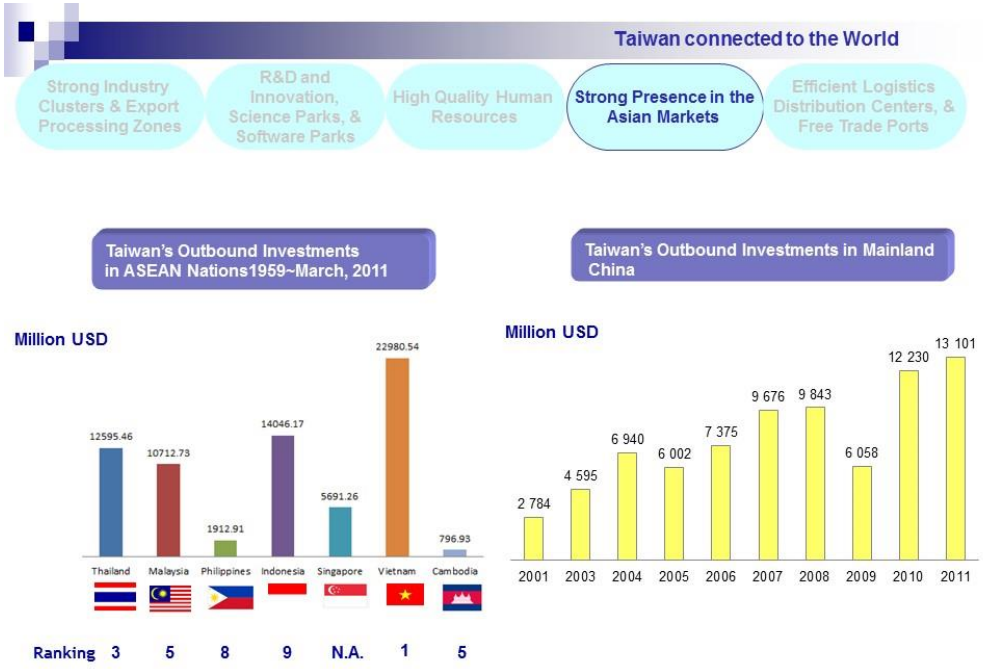
Abundant management talent in financial, international trade industries.

Sufficient internationalized human resources due to long-term foreign trade experience.

Sufficient human resources with multilingual capabilities, namely Chinese, Japanese, and English.

Source: Info Times, 2008

34




Source: Investment Commission, MOEA, DOIS, MOEA

35


**China's Top 10 Exporters in 2009**

Unit:US$ million

| TOP | Company Name | Location | Country | Export Value |
|---|---|---|---|---|
| 1 | Ted-Front Computer | Shanghai | Taiwan (Quanta ) | 21,757 |
| 2 | Hongfujin Precision Industry | Shenzhen | Taiwan (Hon Hai) | 13,206 |
| 3 | Compal Information Technology | Kunshan | Taiwan (Compal) | 8,899 |
| 4 | Futaihua Industrial | Shenzhen | Taiwan (Hon Hai) | 8,797 |
| 5 | Nokia | Beijing | Finland | 8,353 |
| 6 | Wistron Infocomm | Kunshan | Taiwan (Wistron) | 7,163 |
| 7 | Hua Wei Technologies | Shenzhen | China | 6,990 |
| 8 | Maintek Computer | Suzhou | Taiwan (ASUS) | 6,348 |
| 9 | China Shipbuilding | Beijing | China | 6,311 |
| 10 | China National Petroleum | Beijing | China | 5,277 |

Source: Ministry of Commerce of the P.R.O.C.

36

Taiwan connected to the World

Strong Industry Clusters & Export Processing Zones | R&D and Innovation, Science Parks, & Software Parks | High Quality Human Resources | **Strong Presence in the Asian Markets** | Efficient Logistics Distribution Centers, & Free Trade Ports

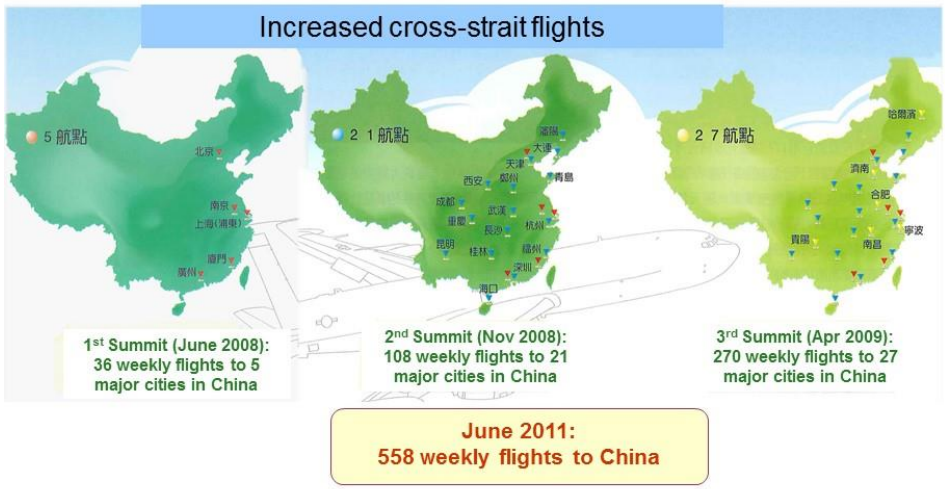
## Taiwan's Strong Economic Position After ECFA


- Import tariff preferences on cross-strait trade
- Opening of direct cross-strait postal, communications, and transportation links facilitates flow of goods and personnel.
- More comprehensive intellectual property rights protection for Taiwan.
- Government incentives for establishment of corporate R&D centers in Taiwan

37

Taiwan connected to the World

Strong Industry Clusters & Export Processing Zones | R&D and Innovation, Science Parks, & Software Parks | High Quality Human Resources | **Strong Presence in the Asian Markets** | Efficient Logistics Distribution Centers, & Free Trade Ports




38

**Taiwan connected to the world**

Strong Industry Clusters & Export Processing Zones | R&D and Innovation, Science Parks, & Software Parks | High Quality Human Resources | **Strong Presence in the Asian Markets** | Efficient Logistics Distribution Centers, & Free Trade Ports

## Success Stories After ECFA

### Qualcomm (USA)

- Currently investing in Taiwan to build Mirasol Display (reflective display technology) factory for mass production.
- Mirasol Display is an emerging display technology, with bistable e-paper features, ability to be clearly seen in sunlight and energy saving benefits.
- Expected to begin production in 2012 with a first- phase investment of US$10bn.

### Hokuto (Japan)

- Pingtung Agricultural Biotechnology Park Branch upgraded into a subsidiary, Compass Biotech Co., Ltd.
- Current annual output of 1,120 tons; second factory being built with an annual output of 2,100 tons. The company produces cultivated edible mushrooms for the Taiwan and Mainland China markets.

### Decathlon (France)

- Expected to establish 15 large-scale stores in Taiwan; first flagship store to open in the Nantun District of Taichung in December.
- Signed contract with the Taichung City Government last year to set up operational headquarters in Taichung. The first store is scheduled to open in the Taichung Eighth-period Redevelopment Zone by the end of the year.

### Costco (USA)

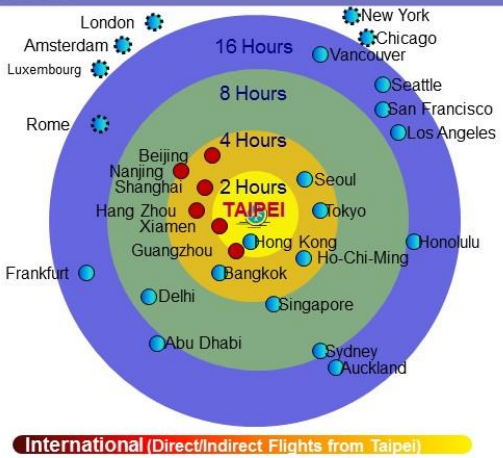
- Already has eight branches in Taiwan; the Neihu store is the 2nd most profitable in the world.
- 5-6 new branches are planned to open in northern Taiwan (such as Sanchong, Xindian, Xinzhuang) and central Taiwan.

39

**Taiwan connected to the World**

Strong Industry Clusters & Export Processing Zones | R&D and Innovation, Science Parks, & Software Parks | High Quality Human Resources | Strong Presence in the Asian Markets | **Efficient Logistics Distribution Centers, & Free Trade Ports**

### *Strategic Location*

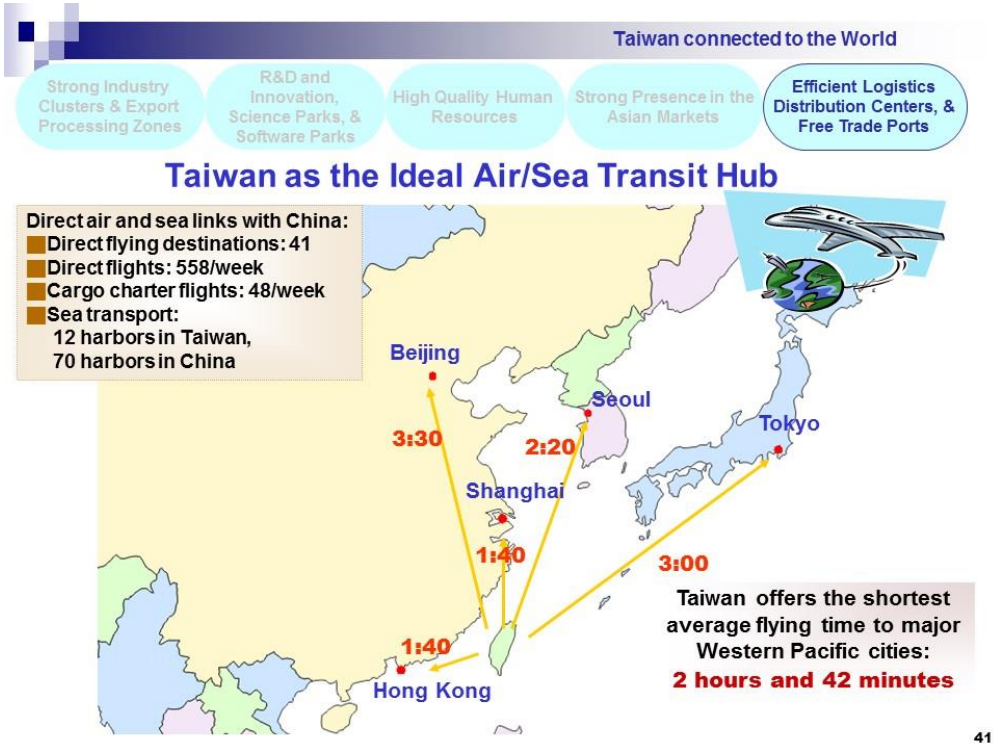



**International** (Direct/Indirect Flights from Taipei)

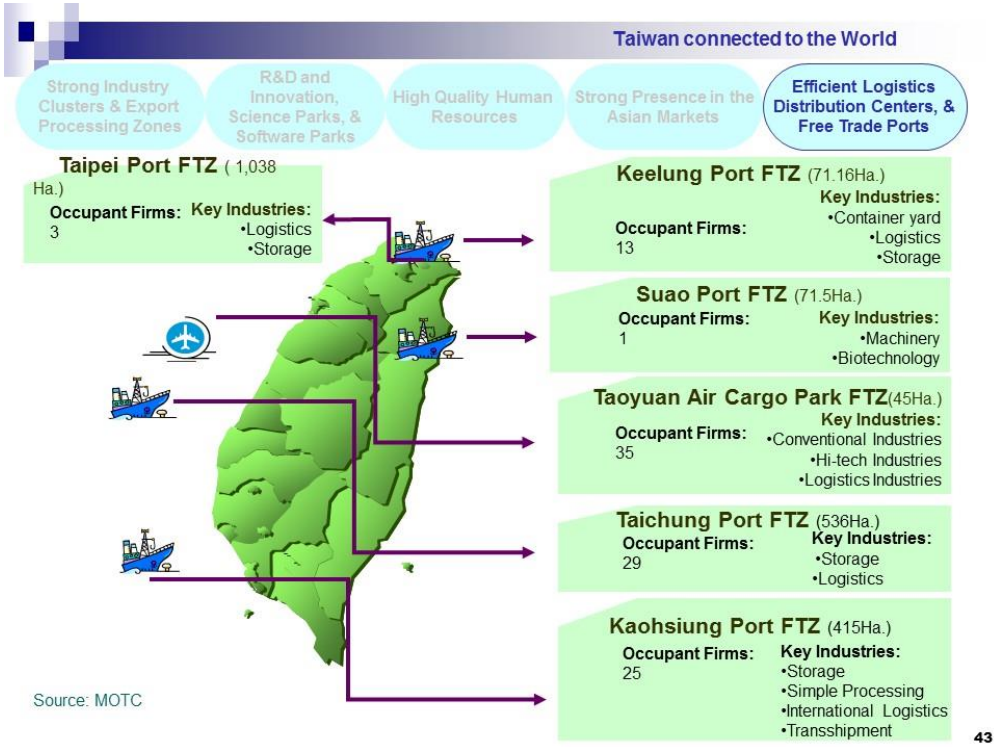
Direct Flights | Indirect Flights | : Int'l Airports

Direct Flights between Taiwan & Mainland China (2010)

40


Taiwan connected to the World
Strong Industry Clusters & Export Processing Zones
R&D and Innovation, Science Parks, & Software Parks
High Quality Human Resources
Strong Presence in the Asian Markets
Efficient Logistics Distribution Centers, & Free Trade Ports
Taiwan as the Ideal Air/Sea Transit Hub
Direct air and sea links with China:
Direct flying destinations: 41
Direct flights: 558/week
Cargo charter flights: 48/week
Sea transport:
12 harbors in Taiwan,
70 harbors in China
Beijing
Seoul
Tokyo
Shanghai
Hong Kong
3:30
2:20
1:40
3:00
1:40
Taiwan offers the shortest average flying time to major Western Pacific cities:
2 hours and 42 minutes
41


Taiwan connected to the World
Strong Industry Clusters & Export Processing Zones
R&D and Innovation, Science Parks, & Software Parks
High Quality Human Resources
Strong Presence in the Asian Markets
Efficient Logistics Distribution Centers, & Free Trade Ports
15 multinational companies set up global distribution centers in Taiwan
samancor
IBM
CREE
FUJITSU
intel
Bayer
BOURNS
HITACHI
Inspire the Next
APPLIED MATERIALS
samag
SmartLink
MEDION
Source: CEPD, IDB, MOEA
42


Taiwan connected to the World
Strong Industry Clusters & Export Processing Zones
R&D and Innovation, Science Parks, & Software Parks
High Quality Human Resources
Strong Presence in the Asian Markets
Efficient Logistics Distribution Centers, & Free Trade Ports
Taipei Port FTZ ( 1,038 Ha.)
Occupant Firms: 3
Key Industries: •Logistics •Storage
Keelung Port FTZ (71.16Ha.)
Occupant Firms: 13
Key Industries: •Container yard •Logistics •Storage
Suao Port FTZ (71.5Ha.)
Occupant Firms: 1
Key Industries: •Machinery •Biotechnology
Taoyuan Air Cargo Park FTZ(45Ha.)
Occupant Firms: 35
Key Industries: •Conventional Industries •Hi-tech Industries •Logistics Industries
Taichung Port FTZ (536Ha.)
Occupant Firms: 29
Key Industries: •Storage •Logistics
Kaohsiung Port FTZ (415Ha.)
Occupant Firms: 25
Key Industries: •Storage •Simple Processing •International Logistics •Transshipment
Source: MOTC
43


Shift
Sell
Made-by-Taiwan
Quality products that global enterprises can trust
Manufacturing expertise
Professional and reliable services
Competitive prices
Efficient supply chain management

www.taiwantrade.com.tw

## e-Trade Services

**Value-added trade databases**
- Taiwan Importers/Exporters Database
- Taiwan Business News
- Taiwan Trade & Industry Associations Directory

Personalized

**Pre-trade**

Taiwan Suppliers
- e-Portals
- e-Catalogs
- e-News/Speical Offers
- Integrated Promotion

**Buyers**
- **e-Sourcing**
- **Inquire Function**
- **Request for Quotation**
- **Add to Preferred List**

**Trade matching**
- Trade Matching
- Biz Exchange
- Search Assists

Interactive

**Post-trade**
- e-Services (logistics management….)

45

## HiRecruit Online

http://hirecruit.nat.gov.tw

46


Taiwan -
Your Ideal
Partner


TAITRA
全球拓展，貿協相伴
Your Partner in World Trade
Thank you!
TAITRA
中華民國對外貿易發展協會
ticc
台北國際會議中心
TAIPEI INTERNATIONAL CONVENTION CENTER
台北世界貿易中心
TAIPEI WORLD TRADE CENTER
TWTC Nangang
台北世貿中心南港展覽館
TWTC Nangang
Exhibition Hall
48

KEYNOTE SPEECH 2

# SIMPLIFY EVERYTHING, DO ANYTHING

**Dimitar Tsvetkov**

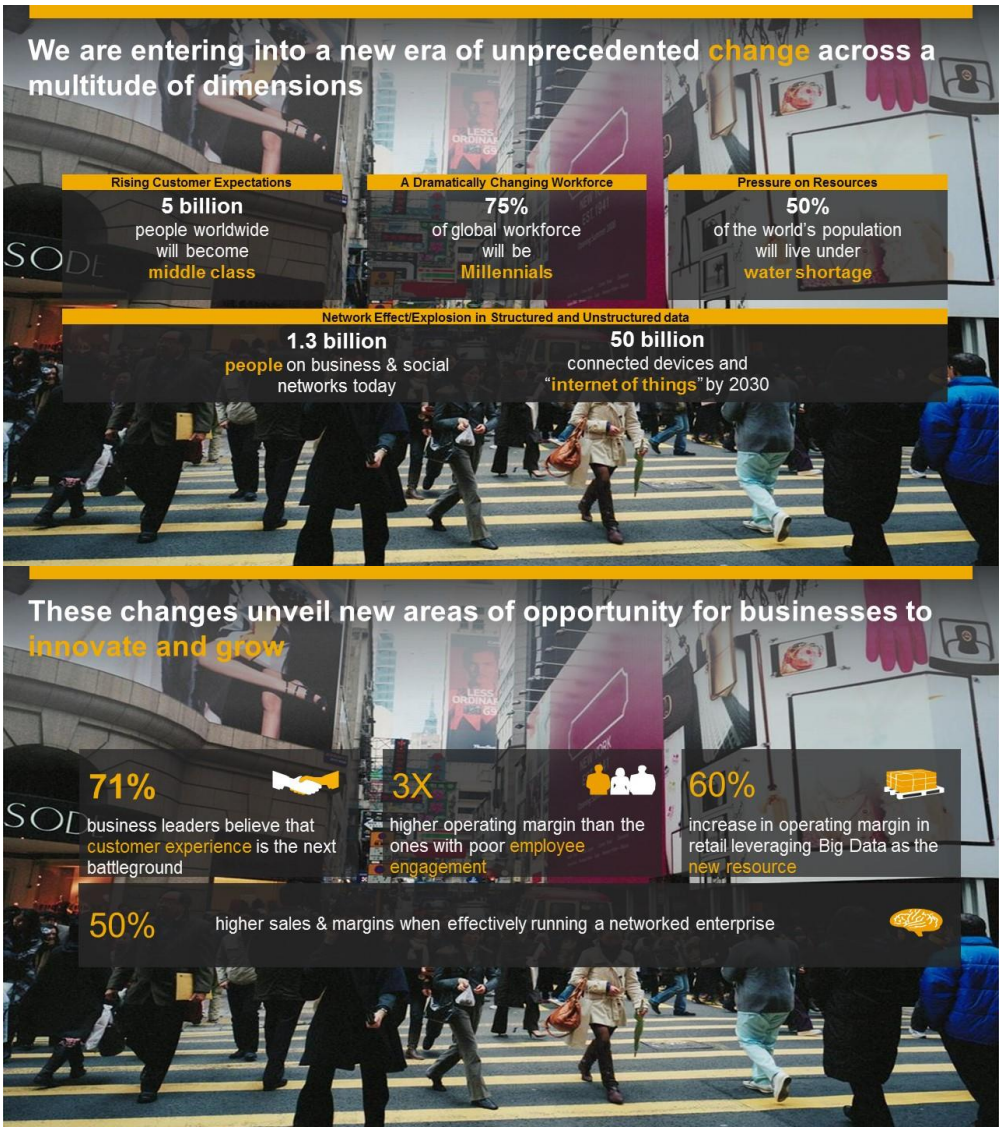
*SAP Solution Consultant, SAP Bulgaria*


"I have always wished for my computer to be as easy to use as my telephone; my wish has come true because I can no longer figure out how to use my telephone."

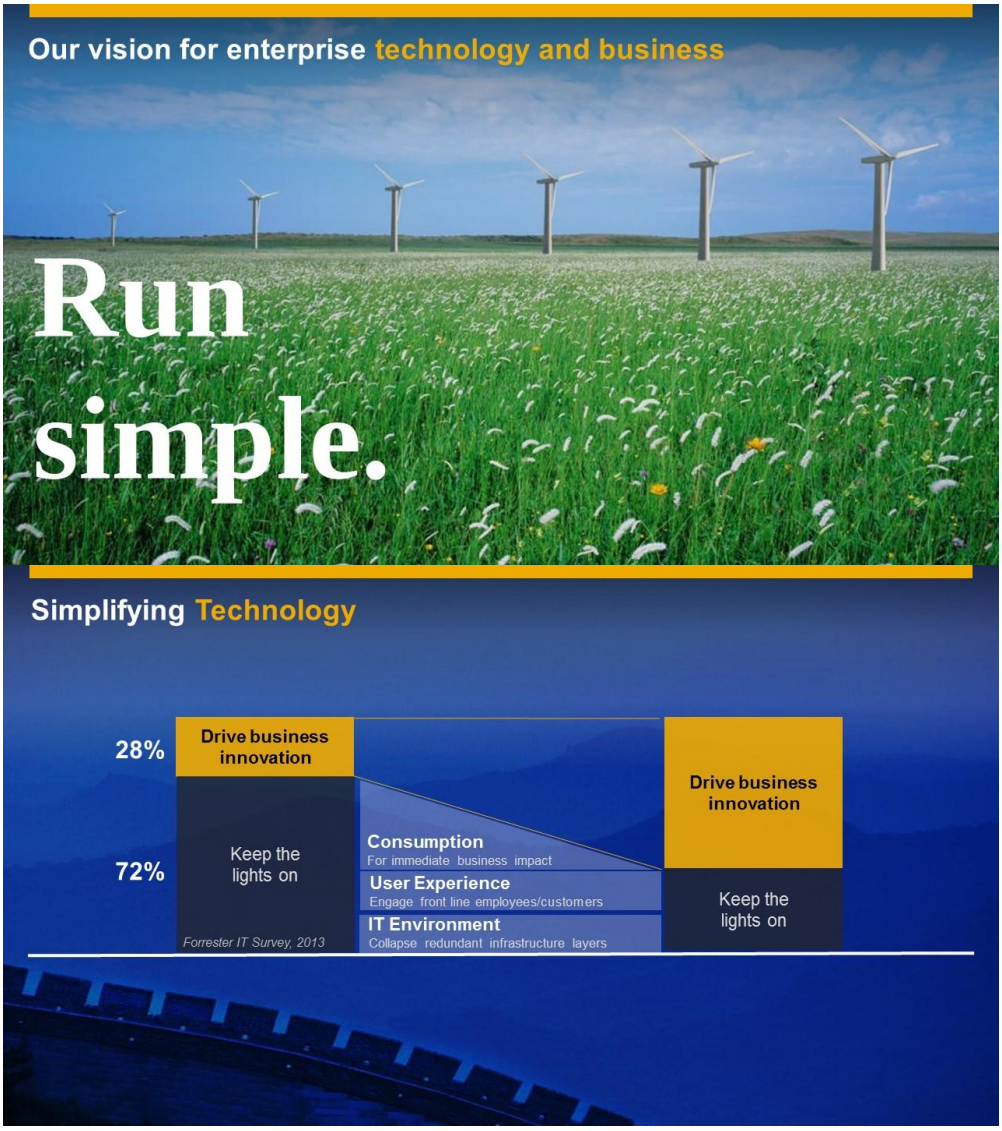
*Danish computer scientist Bjarne Stroustrup*

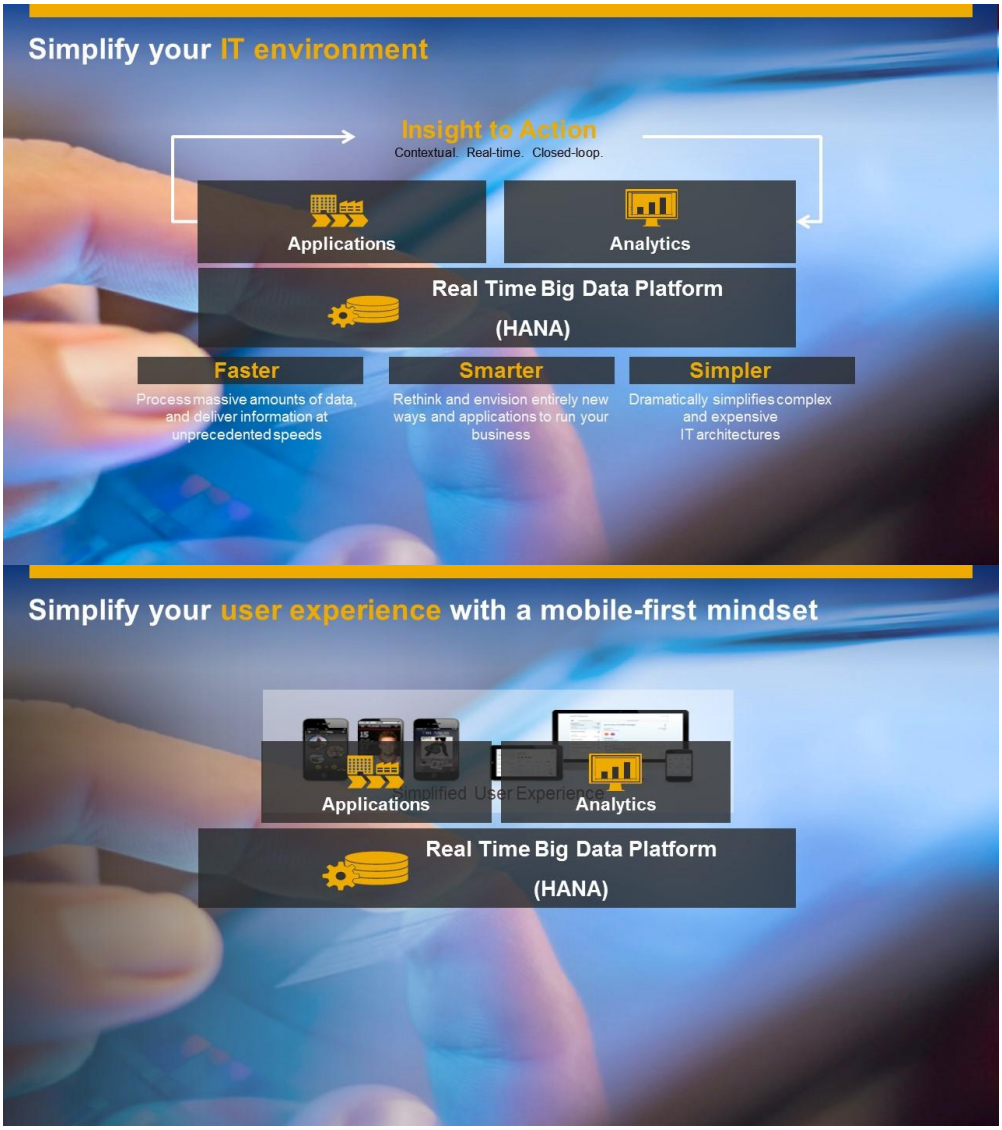
CLARIFICATION: Bjarne Stroustrup is the inventor of the C++ programming language

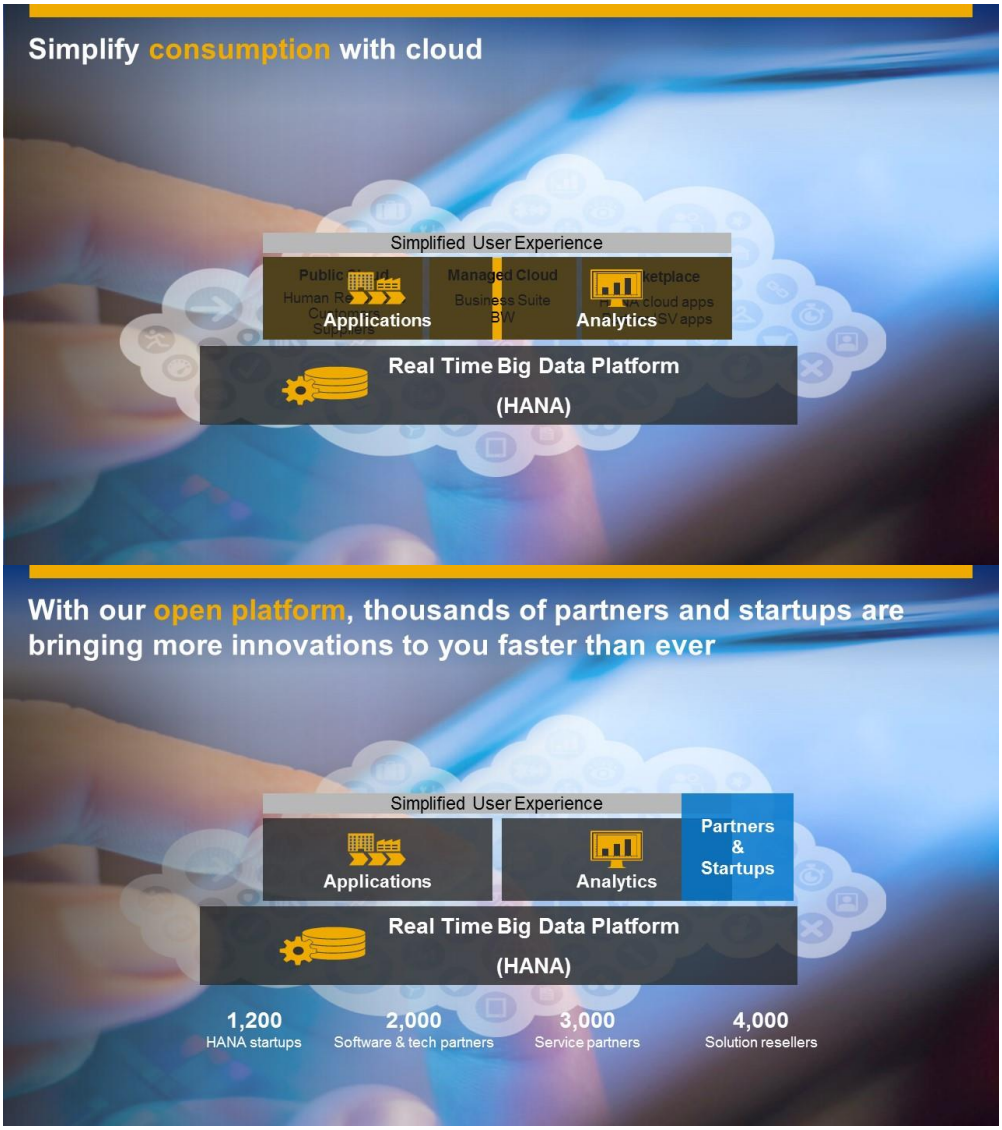
SAP


We are entering into a new era of unprecedented change across a multitude of dimensions
Rising Customer Expectations
5 billion
people worldwide will become middle class
A Dramatically Changing Workforce
75%
of global workforce will be Millennials
Pressure on Resources
50%
of the world's population will live under water shortage
Network Effect/Explosion in Structured and Unstructured data
1.3 billion
people on business & social networks today
50 billion
connected devices and "internet of things" by 2030
These changes unveil new areas of opportunity for businesses to innovate and grow
71%
business leaders believe that customer experience is the next battleground
3X
higher operating margin than the ones with poor employee engagement
60%
increase in operating margin in retail leveraging Big Data as the new resource
50%
higher sales & margins when effectively running a networked enterprise


Transformational technologies offer significant potential to drive business innovation
Cloud
Abundance of cost-effective computational power and storage
Mobile
The new de facto standard in business interaction
Social
Connected business and social networks
Big Data
Real-time analytic for behavioral and predictive insights
More than 60% of CEOs expect 15-50% of their earnings growth in the next 5 years to come from technology-enabled business innovations.
– McKinsey study, 2013
Complexity – most intractable CEO challenge
Built up over decades, it limits the ability to innovate and grow
70% of executives who said complexity is increasing costs
- Bain 2005 Survey
18 months – pace at which data doubles
90% of executives who said business is becoming more complex
- Economist 2010
1% of available data is analyzed
50% of executives unsure who is accountable
- Economist 2010
50 apps per billion in revenue
40% executives worry that their organizations will not keep pace with technology change and lose their competitive edge.
– McKinsey study, 2013


Our vision for enterprise technology and business
Run simple.
Simplifying Technology
28%
Drive business innovation
72%
Keep the lights on
Forrester IT Survey, 2013
Consumption
For immediate business impact
User Experience
Engage front line employees/customers
IT Environment
Collapse redundant infrastructure layers
Drive business innovation
Keep the lights on


Simplify your IT environment
Insight to Action
Contextual. Real-time. Closed-loop.
Applications
Analytics
Real Time Big Data Platform
(HANA)
Faster
Process massive amounts of data, and deliver information at unprecedented speeds
Smarter
Rethink and envision entirely new ways and applications to run your business
Simpler
Dramatically simplifies complex and expensive IT architectures
Simplify your user experience with a mobile-first mindset
Simplified User Experience
Applications
Analytics
Real Time Big Data Platform
(HANA)


Simplify consumption with cloud
Simplified User Experience
Public Cloud
Managed Cloud
Business Suite
BW
Applications
Analytics
Real Time Big Data Platform
(HANA)
With our open platform, thousands of partners and startups are bringing more innovations to you faster than ever
Simplified User Experience
Applications
Analytics
Partners & Startups
Real Time Big Data Platform
(HANA)
1,200
HANA startups
2,000
Software & tech partners
3,000
Service partners
4,000
Solution resellers

## SAP University Alliances

SAP Student Entrepreneurship Program

SAP Millennial Crowdsourcing Community

SAP HANA Innovation Platform for Universities

Enabled via the SAP University Alliances Innovation Lab



## SAP Startup Focus

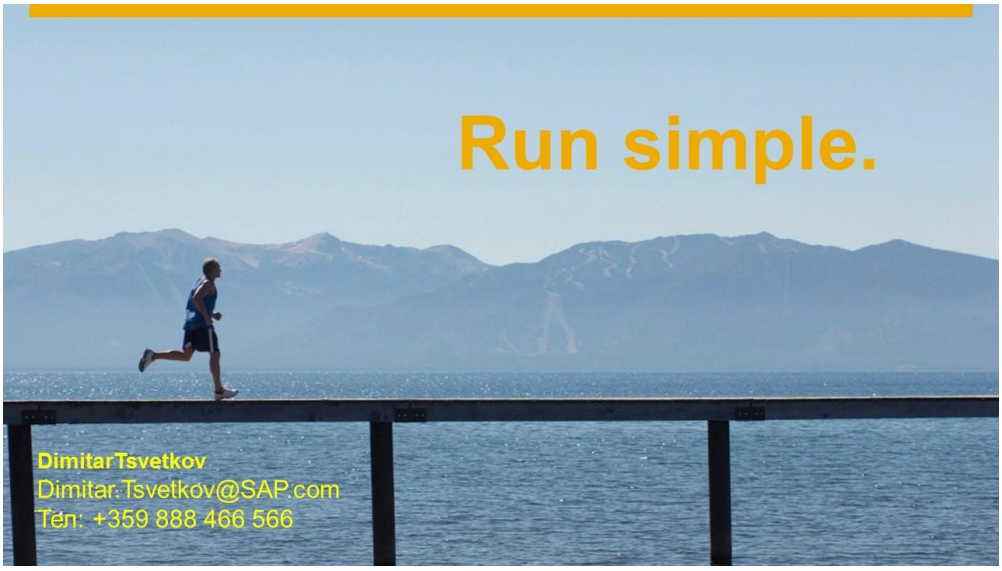
KEY BENEFITS:

- Technology: Free and immediate access to SAP HANA.
- Resources: Free access to technical HANA experts and marketing & sales enablement support
- Customers: Access to SAP's global customer base to sell market ready solutions to
- Money: Opportunity to pitch to the venture community via the SAP HANA Real Time Fund and SAP Ventures.
- Community: Of entrepreneurs, partners, investors and thought leaders in the Big Data space
- SAP launched Startup Focus as a 12-month global program to help big data, predictive and real-time analytics startups to develop new applications on top of the SAP HANA platform and accelerate market traction.

WHAT ARE WE TALKING ABOUT?

- ✓ The Startup Focus program exists to foster innovation outside of the traditional SAP ecosystem and encourage, inspire and energize startups to use the power of the SAP HANA platform to develop their own compelling solutions.
- ✓ The program has helped more than 1,100 to date, with very distinct SAP HANA-based use-cases, from gaming to retail to finance to manufacturing and more.
- ✓ The majority of startups in the program works with non-SAP data and migrated from other platforms, including open source platforms like Cassandra, MongoDB.


Simplifying Business
Hyper-personalize
Customer Experience
Engage & empower
Workforce of the Future
Plan & optimize
Resources in Real-time
Harness the intelligence of
Networked Economy
Business
Outcomes
SAP strategy
SAP Cloud powered by SAP HANA
Run simple. = Cloud + SAP HANA


Run simple.
DimitarTsvetkov
Dimitar.Tsvetkov@SAP.com
Тел: +359 888 466 566

KEYNOTE SPEECH 3

# ARDC METHODOLOGIES FOR KNOWLEDGE ACCUMULATION

**Vassil Angelov and Nickolay Todorovsky**

*Musala Soft, Bulgaria*

## Musala Soft

- Founded in year 2000
- Bulgarian company at 100%
- A team of more than 270 specialists
- Huge amount of successful projects for more than 200 clients in more than 25 countries around the globe
- Technical specialization and partners:

Premier Business Partner IBM | Microsoft Partner Gold Application Development | ORACLE Gold Partner Specialized Oracle Database | vmware PARTNER

## Musala Soft

- Amongst our international clients are:

DHL Danzas Air & Ocean | IBM | hp invent | SAP | OMV | Hamilton Data Services | CISCO | mobilkom austria | vmware | Avery Berkel | Deutsche Telekom | swisscom mobile | aktionaersforum | AVAYA | BUCHANAN | collanos | LIFESCAN Scotland Ltd. Johnson & Johnson company | British American Tobacco

- Amongst our clients in Bulgaria are:

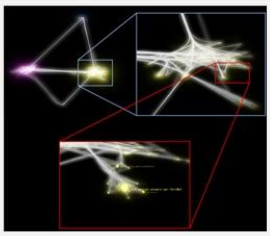
Национална агенция за приходите | Агенция Митници | Министерство на вътрешните работи | tel | VIVACOM | SIEMENS | PHILIP MORRIS | Allianz | dskbank | Адрес недвижими имоти | Atos | PORSCHE | Sofica Group partnership beyond outsourcing

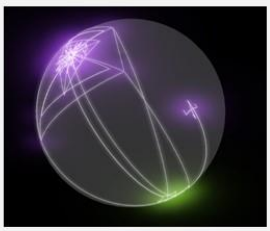
## Applied Research and Development Center (ARDC)

- Founded in the beginning of year 2009
- 5 projects
- Past and current members- around 40

## Applied Research and Development Center (ARDC)

- Software visualization
- Information visualization
- Software modernization
- Automated testing of mobile software





## Applied Research and Development Center (ARDC)


- Operational program „Development of the Competitiveness of the Bulgarian Economy" 2007-2013
- National Innovation Fund
- Scientific Research Fund

## Process – Defining the objectives

- Problems at the workplace and/or during the everyday life
- Analysis of the problems
- Searching of a suitable financing

## Process - Conception

- Research and evaluation of the ideas for the solution
- Presentation of the conception
- Preparation and evaluation of the prototype

## Process - Realization

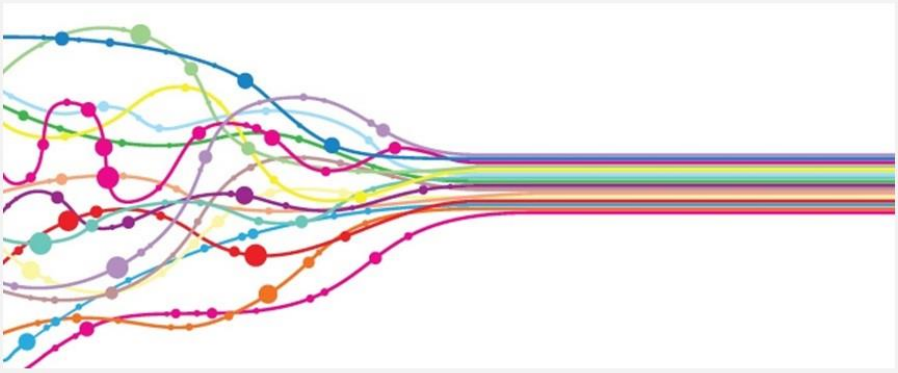
- Design of an end product
- Implementation of an end product
- Interaction with users
- Evaluation and safe-keeping of the solutions
- Follow-up development

## Problems

- Presentation of various kinds of information spaces
- Decreased information overload

## Task

- Bringing a big amount of information into order.



## Objectives

- Create something new, different and easy for integrating, in different products, solution, providing:
  - Intuitive interaction with the information
  - Effective and comfortable navigation

## CORIS

Context Oriented Representation of an Information Space

## Information space?

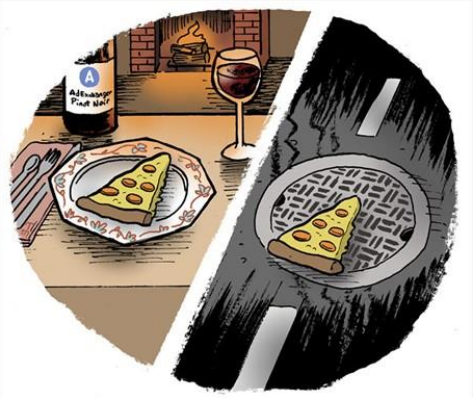
When we look through a web-site...


## Information Space?

… or a restaurant's menu.

# Context

## Context oriented representation?




# Preliminary research

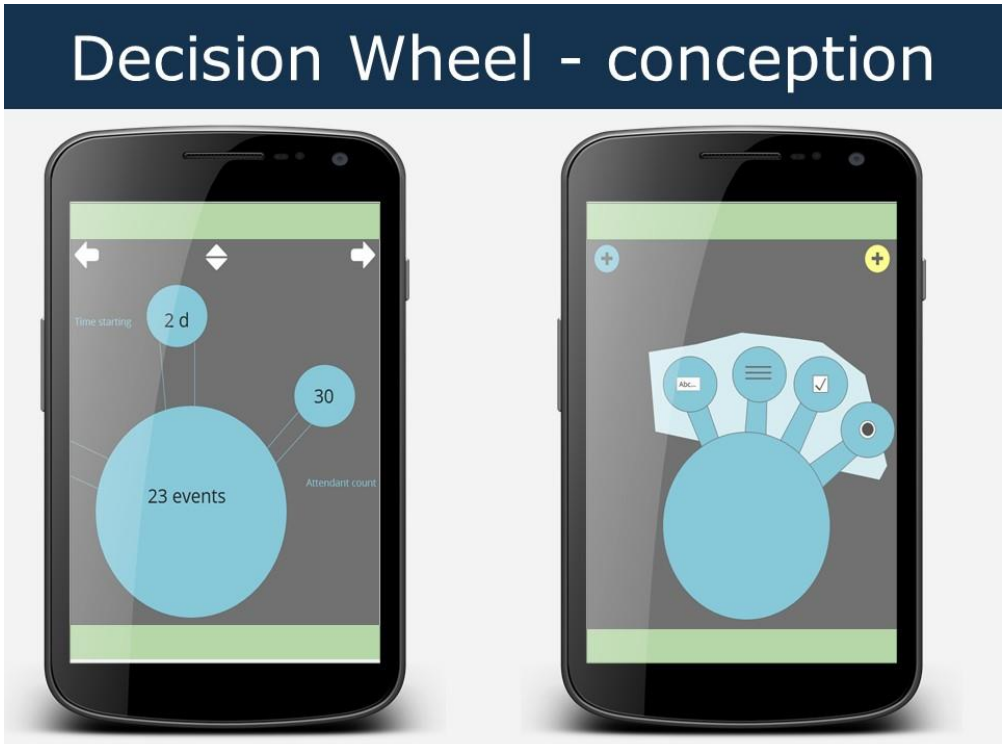
- Small screen presentation techniques
- Navigation
- Focus plus Context
- User experience and usability

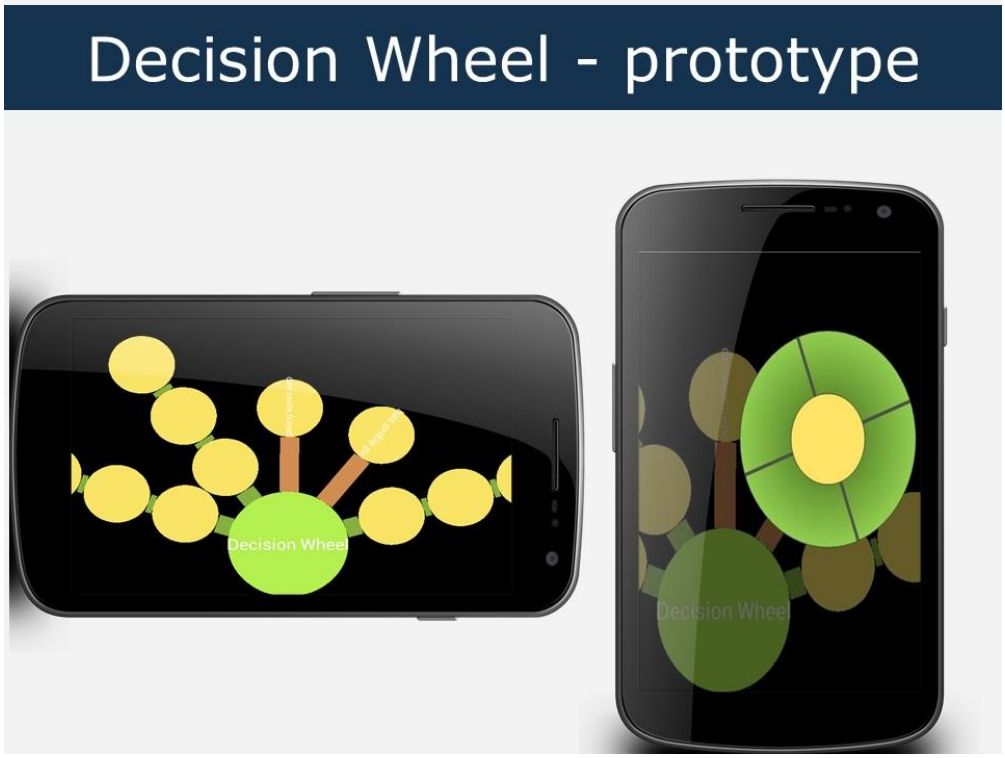
## Research of potential solutions

- Number of gathered articles - 151
- Logged ideas and suggestions- 95

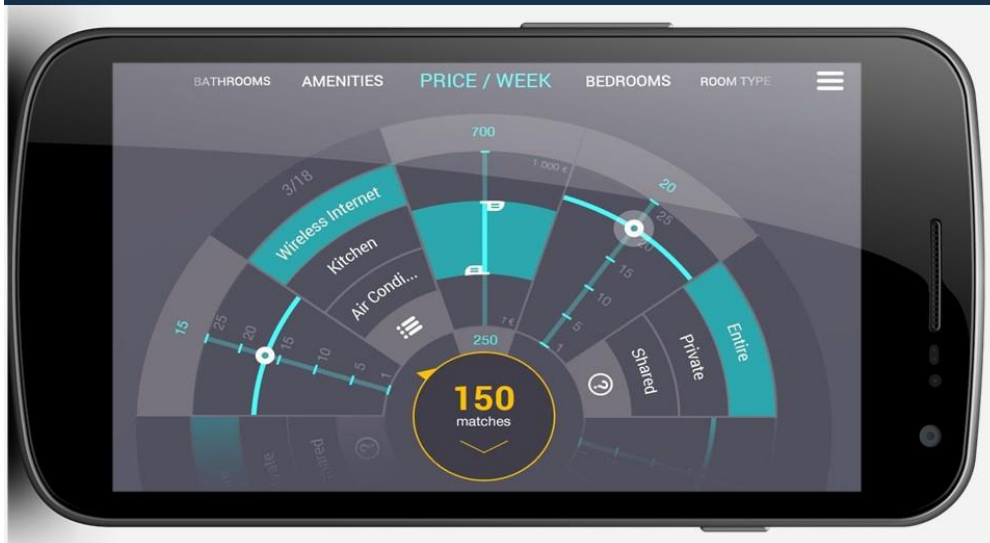
## Conceptual frame

- Innovative and interactive set of filtering criteria
- Presentation of the spinning wheel form
- Preservation of the interaction with familiar filters


Decision Wheel - conception
Time starting
2 d
30
23 events
Attendant count


Decision Wheel - prototype
Decision Wheel
Decision Wheel

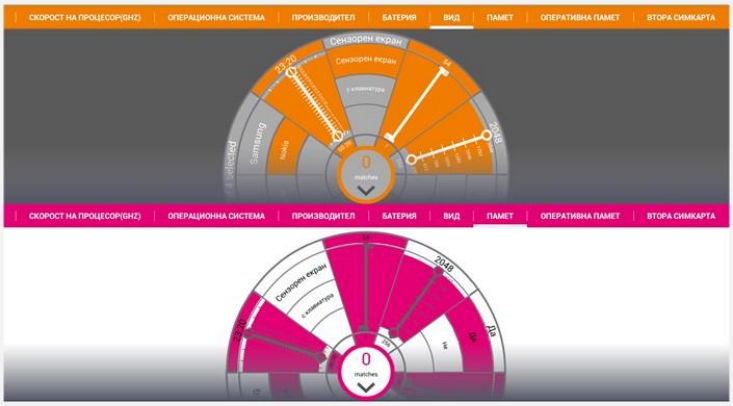
## Decision Wheel - design




## Decision Wheel - realization

# Decision Wheel - realization

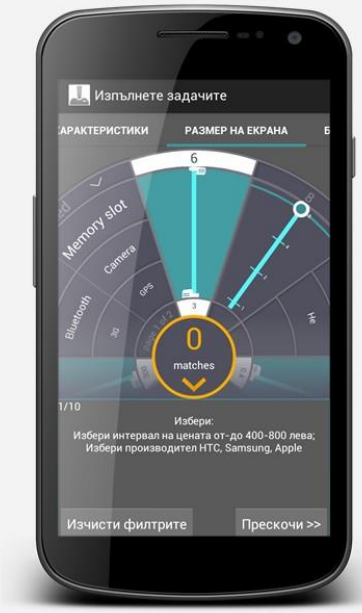
- The opportunity to be easily integrated in different products.




# Usability test

- Conducted over the Decision Wheel
- Real life simulation
  - Game elements
  - Collection of a complex feedback

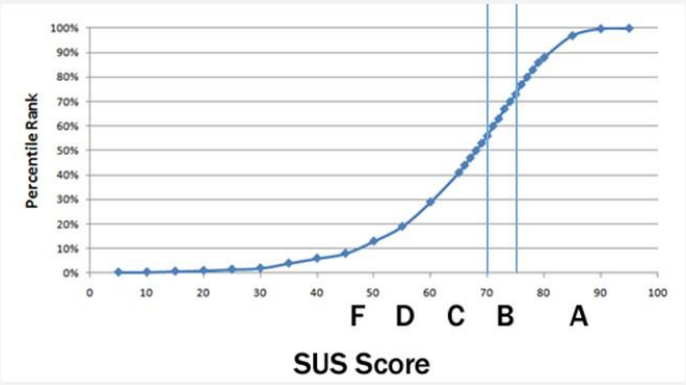
# Usability test





# Usability test

- 1st SUS score = 70 (55%)
- 2nd SUS score = 75 (74%)



## Decision Wheel - score

- Easy integration in different applications.
- Even though the way to select filters is not always the fastest, this way is interactive and pleasing for the users

## Contacts


- www.musala.com
- musala@musala.com

# ICAICTSEE - 2014
# PAPERS

# Cloud Computing and Security of Industrial Systems

**Gavar Mamedov[1], Mustafa Babanli[2], Zafar Jafarov[3]**

[1] Azerbaijan Technical University, Baku, Azerbaijan
aztu@aztu.edu.az

[2] Azerbaijan Technical University, Baku, Azerbaijan
mustafababanli@yahoo.com

[3]Azerbaijan Technical University, Baku, Azerbaijan
c.zafar@mail.ru

**Abstract.** For automation industry, Information Technology (IT) is an essential production factor, but also a major expense post. Because cloud computing promises to deliver IT services more flexibly and cost-efficiently, it for automation industry prospectively. However, given the high degree of regulation, concerns regarding security and compliance requirements arise. In this work, provide a theoretical analysis of security problems in the context of cloud computing. This analysis is complemented by the initial results of an ongoing case study concerning the practical relevance of these problems in the industry. The analysis confirms that security issues pose notable obstacles for the adoption of cloud computing in practice, but also points to appropriate countermeasures

**Keywords.** industrial systems, automation industry, cloud technology, security, compliance

## 1. Introduction

To define corporate sustainability is not an easy task. Even till now terms like CSR (Corporate social responsibility) and corporate citizenship are continue widely to be used. Nowadays, however, all these concepts are replaced by the broader term corporate sustainability. In the beginning it was not customary to define sustainability at corporate level.

The purpose of this report is to review the most common concepts of corporate sustainability in order to outline basic qualitative aspects of the most popular views. The ultimate goal is to find promising directions for building quantitative models of corporate sustainability at a later stage.

The most significant consideration is that these quantitative models had to be built on the basis of information available to the organization. This is information that is collected on a daily basis in corporate databases as a result of current activities.

For automation industry, the collection, processing, and dissemination of large amounts of information constitute the basis for all processes [1]. Accordingly, industry institutions are among the most intensive users of Information Technology (IT) across various service industries. Following some controversy about the “productivity paradox” of IT in the, it has been empirically validated since that investments in IT may result in substantial returns, cost savings. Due to the constant decline in the absolute cost of IT equipment, many industry institutions have historically replaced other production factors with IT. This has led to a further increase in the relative importance of IT, compared to other

production factors such as human labor. As a result, IT constitutes a major expense post for most industrial institutions today.

Lately, cloud computing has arrived as a novel IT paradigm that promises to "revolutionize" the way IT services are provisioned and consumed. The essential idea of cloud computing is to deliver IT services – such as compute infrastructure or storage – in a utility-like manner , thus making these services ultimately more flexible and cost-efficient. Given the role of IT in the critical infrastructure industrial , as an essential production factor, but also a major expense post, cloud computing may seem as a "perfect match" for this industry [2].

However, there appears to be one major obstacle: security. In fact, analysis confirmed that general security concerns and data privacy concerns constitute the most severe reasons for not using cloud computing. It is reasonable to conclude that these concerns are especially pronounced in the critical infrastructure, which is among the most regulated sectors, and thus subject to rigid security requirements.

Based on these observations, we examine the following research question in the work at hand: "Do security concerns pose an obstacle for the adoption of cloud computing in the critical infrastructure industry and, if yes, which concerns specifically"? We provide a theoretical analysis of potential security problems in conjunction with the application of cloud computing.

The remainder of this work is structured as follows: In the following section, we provide an introduction into the fundamental concepts of cloud computing. In the subsequent section, analyse potential security with the use of cloud computing in the industrial systems. Thereafter, we present the methodology and results of our ongoing empirical research. The paper concludes with a brief summary and outlook.

## 2. A brief overview of the concept of cloud technology

The two main sources for precise definition of corporate sustainability are the Brundtland Commission's Report [1] and John Elkington's writings [2]. The first source defines the sustainable development as: "Development that meets the needs of the people today without compromising the ability of future generations to meet their own needs." The second source defines the sustainability at corporate level using the concept of Triple Bottom Line (TBL, 3BL or BL3). This concept is also known as the three sustainable business pillars or the three P's – People, Planet and Profit. That means that to be sustainable as a whole, the business has to be sustainable in its basic three aspects – economic (financial), social and ecologic (environmental).

While the term cloud computing is currently very popular in research and practice today, no commonly accepted definition exists so far. Recently, however, the definition by the National Institute of Standards and Technology (NIST) has emerged as a de-facto standard; it states that "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources". Such resources may include virtually any type of computing capabilities, including "networks, servers, storage, applications, and services". In addition, the definition names five essential characteristics: First, the ability for consumers to commission and decommission capacities in an autonomous manner (on-demand self-service); second, the provision of capacities to heterogeneous end-user devices via the Internet (broad network access); third, the pooling and subsequent provision of resources according to a multi-tenant model, often based on virtualization techniques (resource pooling); fourth, the ability to rapidly add or remove capacities (rapid elasticity); fifth, a metered service provision, often based on a pay-per-use

model (measured service). A basic taxonomy is cloud systems, which distinguishes four common deployment and three service models [3]. An overview is depicted in Figure 1. The deployment models essentially refer to the relationship between service provider and service consumer (also referred to as service user). In the case of *a private cloud*, a service is offered to one exclusive consumer, either by a provider from the same organization or by an external party. A *community cloud*, in contrast, is restricted to a pre-defined set of consumers, rather than an individual consumer. Lastly, a *public cloud* is operated by a specialized vendor; it is open to the general public or a large group of consumers. The term *hybrid cloud* can refer to any combination of aforementioned deployment models. In general, economies of scale increase moving from a private to a public cloud, whereas the control of the user over the overall cloud system decreases [5].

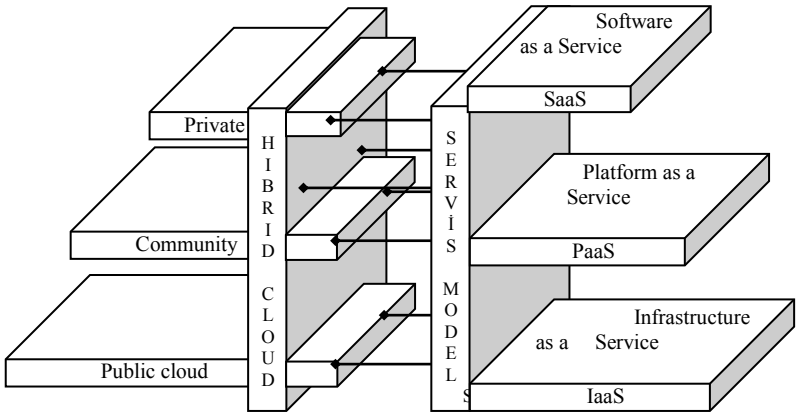



Figure 1: Common cloud deployment and service models

The service models refer to the level of complexity that a cloud service provides. Infrastructure as a Service (IaaS) includes the provision of rather low-level IT capabilities, such as storage or compute power. Platform as a Service (PaaS) refers to somewhat more sophisticated offers, such as programming and execution environments. Lastly, Software as a Service (SaaS) includes complex applications, which are often operated on the basis of IaaS or PaaS services. In general, the complexity of the services increases moving from IaaS to SaaS, whereas the degree of (technical) standardization appears to decrease [6].

We need to strongly emphasize that the economic and the social aspects of the business intersect in the field of the marketing activity of the company. As a sophisticate tool for achieving the corporate goals through exchange, the marketing is responsible for two basic issues: developing strong company brands and building positive perceptions of the company as a whole in the society. Very often, in order to address these complex issues, marketers need help from experts outside the field of marketing. The development of sophisticate CRM systems (Customer Relationship Management Systems) is a step in this direction.

There is another field in the company's activity, where two basic aspects of the corporate sustainability intersect. The economic and the environmental aspects intersect in the field of the corporate real estate management (CREM). On the one hand, corporate real estate constitutes the main part of the company assets and has a decisive impact on the final financial results. On the other hand, buildings and building facilities are the main source of company's carbon footprint.

That is why when discussing corporate sustainability we have to pay special attention to the CRM systems and CREM (corporate real estate management) of the organization.

## 3. Security Problems for Cloud Technology

Recently, a growing number of researchers have been concerned with the issue of security in cloud computing. In our theoretical analysis, we pursued the aim of reviewing the existing literature and consolidating these findings in a structured manner, thus giving us a basis for the subsequent empirical investigation. As the guideline of analysis, used the ten security domains of the well-known Certified Information Systems Security Professional (CISSP) certificate [7]. These domains cover diverse aspects of security, ranging from physical security of computing facilities, business continuity planning for disaster scenarios, to the application of cryptography techniques. Most of these domains can be found in identical or comparable form in other IT security guides, e.g., by the Cloud Security Alliance or the Information Systems Audit and Control Association. Thus, while not specifically tailored to cloud computing, but rather IT in general, the CISSP domains provide a comprehensive scheme for the classification of security issues. Accordingly, mapped each security problem or risk that found in the literature onto one of the ten CISSP domains. In addition, identified specific security objectives that may be threatened by each problem or risk. In this context, we focused on the three "classic" security objectives of confidentiality, integrity, and availability, which have been well-known for many years as part of the so-called "CIA triad". Confidentiality describes that information may only be read by authorized parties; in this context, authorization refers to the fact that these parties possess appropriate access rights [8]. Lastly, availability means that information or a system is accessible in a timely and reliable way whenever needed.

## 4. Empirical analysis in the industry systems

The consideration of the above leads us to the possibility to develop models of the corporate sustainability.

In order to empirically assess our research question and examine the practical significance of the previously identified, cloud computing-related security issues, chose the qualitative research approach of a case study.

With respect to this instrument, different designs are described in the literature, which exhibit specific advantages and disadvantages. In work, pursue a holistic, multi-case design. In this context, holistic means that industry institutions as a whole – and not their individual units or departments – constitute the matter of examination. Chosen a multi-case design due to the potentially higher robustness and explanatory power of such design. As primary data source, selected the instrument of personal interviews with domain experts. As major strengths, this instrument permits a targeted examination of the case study topic and can be highly insightful. However, due to different forms of bias in the responses, the results should also be subject to careful interpretation.

As the basis for the interviews, compiled a questionnaire consisting of roughly 40 individual questions. Using this questionnaire, conducted interview with two representatives of a industry institutions Azerbaijan. Both interviewees work in the IT department of their institute, with a specific focus on IT security, and have previously gained professional experience with respect to cloud computing. In the following, we will refer to the interviewees as A and B.

Each interview lasted approximately one hour in time. Both interviews were digitally recorded and subsequently transcribed into written text.

The transcripts and notes were analyzed using the method of qualitative content analysis. This method is among the recommend procedures for the analysis of expert interviews. The analysis process involves five steps, including a summary and codification of statements, and ultimately results in deduction of scientific concepts. In contrast to more complex analysis procedures, such as the coding method, the qualitative content analysis requires less initial effort and is thus [9]. very well suited for the deduction of preliminary results

The entire document should be in Times New Roman or Times font. Type 3 fonts must not be used. Other font types may be used if needed for special purposes.

Given the restricted number of interviews that have been conducted to date, the following results should be considered preliminary. In addition, because both interviewees are associated with a industry, the discussion of results focuses on the critical infrastructure. However, initial results can provide valuable insights with respect to the research question [10].

To begin with, the interviews confirmed the notion that the term cloud computing is interpreted very widely. Specifically, interviewee B stated that "[cloud computing is] a dazzling term, which is used for a multitude of things and not clearly defined." In general, the interviewees agreed, however, that cloud computing involves the provision of services from a shared environment via a (public) network. These services are standardized and ready-to-use.

All three service models in cloud computing are of relevance to the automation industry. However, according to the interviewees, cloud computing will not be the dominant delivery model in any critical infrastructure, because some processes cannot be outsourced to third parties. Automation industry which are subject to constant change, are specifically suited for the application of cloud computing.

Among the delivery models, private cloud computing has the highest relevance. A private cloud can serve as central service, which is used by the critical infrastructure itself or its outsourced partners. According to both interviewees, IT in the critical infrastructure is subject to a broad range of risks. These risks relate to people, processes, systems, and external factors. Damages may not only concern virtual goods, but also physical goods, such as data centers, facilities, or employees. Interviewee A also pointed to non-financial damages, for instance "risks in terms of reputation". Interviewee B sees the inability to judge the reliability of cloud services as a major problem, given that processing failures may lead to substantial risks for the institute.

With respect to the first CISSP domain, "information security governance", both interviewees saw the application of appropriate monitoring as an essential prerequisite for the use of cloud services. Thus, according to interviewee B, a cloud provider will have to comply with "extensive manuals and policies" and certify its adherence to those rules. As a specific technical mechanism, interviewee A named the use of trusted platform modules, which guarantee that "only warranted operations can be conducted on specific data". In addition, the encryption of critical data is seen as an appropriate measure.

Interviewee B did not see the risk of a vendor lock-in with standardized services in the critical infrastructure industry. Still, critical infrastructure should be very aware about the dependence on specific vendors.

Concerning the second CISSP domain, "Access Control", interviewee B saw the risk of an abuse of administrative privileges as valid scenario, with enormous potential for

attacks. Thus, administrative personnel should "underlie a detailed monitoring". Both interviewees pointed out that critical data, including authentication credentials, should always be transferred via secured channels. With respect to the theft of user accounts, "two-factor authentication" was named as a technical countermeasure.

In the context of the third CISSP domain, "Cryptography", both interviewees stressed the significance of appropriate monitoring solutions to address potential security issues. As a technical measure, interviewee A further pointed to the use of client-based encryption mechanisms, which ensure that data "is [exclusively] written back to the cloud in encrypted form" (i.e., end-to-end encryption is enforced). Both interviewees further stressed that the distinction between data-at-rest and data-in-transit is important for the choice of appropriate security mechanisms, such as channel encryption.

With respect to the fourth CISSP domain, "Physical Security", the use of appropriate monitoring solutions and enforcement of policies at the cloud provider is seen as important aspect again. Both representatives agreed that potential damages do not only concern virtual, but also physical goods, such as data centers or facilities. In this respect, interviewee A believes that "cyber war and cyber terrorism will play a certain role [in the future]".

Concerning the fifth CISSP domain, "Security Architecture and Design", interviewee B identified the comprehensive training of employees as an important measure to raise the awareness of security problems. This does not only concern "the employees [of the institute itself], but everyone you has access to a company's systems, because today, multiple [external] service providers are employed". In this context, interviewee A specifically pointed to the risk through cloud-based man-in-the-middle attacks, saying that a lack of awareness at both the user and provider side would create "completely new opportunities", specifically with respect to eavesdropping. Once again, trusted platform modules are perceived as a viable technical countermeasure to address many security problems.

With respect to the sixth CISSP domain, "Business Continuity Planning (BCP)", the failure of communication channels, namely access to the Internet, was acknowledged as a relevant risk. Interviewee A named legal agreements with the network providers as appropriate countermeasure, saying that "depending on the risk of the processed data there has to exist a disaster recovery scenario in the case of a cloud setting; you have to be able to substitute [network capabilities]".

In the context of the seventh CISSP domain, "Telecommunications and Network Security", the interviewed representatives stressed the importance of monitoring to detect attacks that aim at an exploitation of network or computing capacities. Interviewee A further stated that redundant resources should be made available, depending on the criticality of systems or data.

With respect to the eighth and ninth CISSP domains, "Application Development Security" and "Operations Security", both interviewees referred to the same mechanisms that were previously discussed with respect to the first domain. This includes the use of appropriate monitoring mechanisms and the enforcement of policies by the critical infrastructure as service user.

Lastly, concerning the tenth CISSP domain, "Legal, Regulations, Investigations, and Compliance", interviewee B acknowledged that the inability to localize data in clouds may constitute an important obstacle for their adoption, given that this situation may result in judicial or regulatory problems. Once again, monitoring is seen as a potentially appropriate countermeasure. The physical location of data most notably plays a major role due to different jurisdictions.

In summary, we found that many of the security issues with cloud computing that we identified in our theoretical analysis are also acknowledged by practitioners from the industrial practice. In many cases, appropriate monitoring or trusted platform solutions are named as appropriate technical countermeasures. In addition, it appears that critical infrastructure tend to use the instrument of legal agreements and compliance rules to mitigate risks and shift the financial and legal responsibility for security issues to the cloud providers.

## 5. Conclusion

In the automation industry, IT is one of the substantial production factors, and its relative importance has steadily increased in recent decades. However, IT also poses a major expense post. Cloud computing is a novel architectural paradigm that promises to revolutionize the way IT services are provisioned and consumed. Due to its potential for cost-savings and its flexibility, cloud computing may appear as a perfect match to the industrial systems. However, security concerns have recently been named as a potential sticking point for the adoption of cloud computing, specifically in the critical infrastructure which is subject to a multitude of regulatory requirements.

In this work, to analyze whether security concerns pose an obstacle for the application of cloud computing in the critical infrastructure. For that matter, we identified a set of potential cloud-related security risks, based on a survey of current literature. Subsequently, empirically verified findings through an ongoing case study in the industrial systems.

According to the interviews with two experts, security concerns do in fact constitute an obstacle for the adoption of cloud computing. This is specifically true with respect to the public cloud computing deployment model. Accordingly, this model is only applied to a very limited extent at present. However, potential for the application of cloud computing is seen across all critical infrastructure in the industry area.

On the basis of the case study analysis, it appears that critical infrastructures focus on both legal and technical measures to address potential security problems. The former include legal agreements with cloud providers and the enforcement of security-related manuals and policies; the latter include the use of encryption and trusted platform modules.

## References

1. Ardelt, M., Dölitzscher, F., Knahl, M. and Reich, C. (2011) Sicherheitsprobleme für IT-Outsourcing durch Cloud Computing, HMD - Praxis der Wirtschaftsinformatik, 48, 281, 62-70.
2. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. and Zaharia, M. (2010) A View of Cloud Computing, Communications of the ACM, 53, 4, 50-58.
3. Brynjolfsson, E. and Hitt, L.M. (2000) Beyond Computation: Information Technology, Organizational Transformation and Business Performance, The Journal of Economic Perspectives, 14, 4, 23-48.
4. Lenk, A., Klems, M., Nimis, J., Tai, S. and Sandholm, T. (2009) What's Inside the Cloud? An Architectural Map of the Cloud Landscape, In: Kamal Bhattacharya, Martin Bichler, Stefan Tai (Eds.), Proceedings of the ICSE Workshop on Software Engineering Challenges of Cloud Computing, May 23, Vancouver, Canada, IEEE, 23-31.
5. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J. and Brandic, I. (2009) Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, Future Generation Computer Systems, 25, 6, 599-616.
6. Leymann, F. (2009) Cloud Computing: The Next Revolution in IT, In: Dieter Fritsch (Ed.), Proceedings of the Photogrammetric Week '09, September 7-11, Stuttgart, Germany, Wichmann Verlag, 3-12.
7. [Cloud Security Alliance (2011) Security Guidance for Critical Areas of Focus in Cloud Computing V3.0. [Online] https://www.cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf
8. Conrad, E., Misenar, S. and Feldman, J. (2010) CISSP Study Guide, Elsevier, Burlington.
9. Cropley, A.J. (2005) Qualitative Forschungsmethoden: Eine praxisnahe Einführung (2nd ed.), Klotz, Magdeburg.
10. Darke, P., Shanks, G. and Broadbent, M. (1998) Successfully Completing Case Study Research: Combining Rigour, Relevance and Pragmatism, Information Systems Journal, 8, 4, 273-289.

# A Longitudinal Study on Patient Safety Culture of Physicians – A Case of A Regional Hospital in Taiwan

**Yii-Ching Lee[1,2,3], Li-Ching Wei[4], Shao-Jen Weng[5], Chih-Hsuan Huang[6], Wan-Lin Hsieh[7], Liang-Po Hsieh[8], Hsin-Hung Wu[9]**

[1] School of Health Policy and Management, Chung Shan Medical University, Taichung City, Taiwan
[2] Human Resource Management, Cheng Ching General Hospital-Chung Kang Branch, Taichung City, Taiwan
[3,8] Department of Medical Quality Management, Cheng Ching General Hospital-Chung Kang Branch, Taichung City, Taiwan
[3] yiiching.lee@gmail.com
[8] lphsieh624@yahoo.com.tw

[4,9] Department of Business Administration, National Changhua University of Education, Changhua, Taiwan
[4] wei5514@hotmail.com
[9] hhwu@cc.ncue.edu.tw

[5,7] Department of Industrial Engineering and Enterprise Information, Tunghai University, Taichung City, Taiwan
[5] sjweng@thu.edu.tw
[7] hsiehwl@thu.edu.tw

[6] Department of Business Administration, Tunghai University, Taichung City, Taiwan
tititacer@gmail.com

**Abstract.** The purpose of this study is to evaluate the attitudes from physicians toward patient safety culture from a longitudinal viewpoint. This study uses multivariate analysis of variance to examine how different demographic variables including gender, age, and experience in position impact nine dimensions of patient safety culture. In addition, one variable, year of data, has been added to demographic variables. That is, the physicians' data from 2011, 2012, and 2013 are listed as 1, 2, and 3 in the column from a longitudinal viewpoint. Bonferroni test is performed for post hoc analysis except for gender with only two levels. The major findings are below. The interaction effect of gender and experience in position is significant in stress recognition but Bonferroni test does not show any significant difference among twelve combinations. The interaction effect of gender and year of data is significant in stress recognition as well. Bonferroni test shows that male physicians in 2013 perceive higher satisfaction than female physicians in 2011 in stress recognition. Moreover, the interaction effect of age and year of data is significant in stress recognition. Physicians with 21-30 years old in 2011 perceive lower satisfaction in stress recognition than physicians with 31-40 years old in 2013 and physicians with 51-60 years old in 2013. With the above information, the hospital management can pay much attention to stress recognition and the interaction effects that perceive lower satisfaction in stress recognition.

**Keywords.** Patient safety culture, multivariate analysis of variance, a longitudinal analysis, demographic variable, Bonferroni test.

## 1. Introduction

Patient safety culture plays an important role in positive influences on shorter stay, fewer prescription errors, less ventilator-associated pneumonia, fewer blood-stream and urinary tract infections, and lower mortality [1-3]. Wagner et al. [4] pointed out that hospitals with a more open culture and reflective attitude toward errors and patient safety could reduce the number of accidents and failures. Deilkas and Hofoss [5] stated that patient safety culture can be measured by safety culture surveys.

The hospital survey on patient safety culture with good validity and reliability developed by Sexton et al. [6] has been widely used to assess perceptions of patient safety culture from the entire healthcare organization's staff [7,8]. In hospitals, the core staffs are medical physicians and registered nurses [9]. However, in this study, we focus on physicians solely. In addition, it is of interest for hospital management to observe the attitude from physicians toward patient safety culture with different demographic variables and their combinations from time to time. Therefore, the data in 2011, 2012, and 2013 are to be used and multivariate analysis of variance (MANOVA) is applied for analyses when demographic variables are taken into account.

## 2. Patient Safety Culture

The hospital survey on patient safety culture has six dimensions and each dimension are defined below [9,10]. Teamwork climate is the perceived quality of collaboration between personnel. Safety climate is defined as the perceptions of a strong and proactive organizational commitment to safety. Perception of management is the approval of managerial actions. Job satisfaction is the positivity about the work experience. Stress recognition is to measure how performance is influenced by stressors. Finally, working condition is defined as the perceived quality of the work environment and logistical support such as staffing and equipment.

Taiwan Joint Commission on Hospital Accreditation initially developed the Chinese version of patient safety culture survey in 2008 with six dimensions and 30 questions based on the short form of the hospital survey on patient safety culture in 2006 [4,11]. Later, three hospital-level aspects of safety culture from Agency for Healthcare Research and Quality have been included. Hospital management support for patient safety is defined as the hospital management provides a work climate that promotes patient safety and shows patient safety is a top priority. Teamwork across hospital units is the extent to which hospital units cooperate and coordinate with one another to provide the best care for patients. Hospital handoffs and transitions is the extent to which important patient care information is transferred across hospital units and during shift changes. That is, there are three, four, and four question items for hospital management support for patient safety, teamwork across hospital units, and hospital handoffs and transitions, respectively [12]. Therefore, the final form of the Chinese version consists of nine dimensions and 41 questions.

Lee et al. [13] summarized that physicians are required to fill out all of the questions based on a five-point Likert scale ranging from strongly agree to strongly disagree or frequency such as never, rarely, sometimes, most of the time, and always. Besides, each hospital is required to conduct the survey in a yearly basis. Therefore, a longitudinal analysis can be performed based upon the data sets from different years.

## 3.A Case Study

The purpose of this study is to evaluate the attitudes from physicians of a regional hospital in Taiwan toward patient safety culture from a longitudinal viewpoint. This study uses multivariate analysis of variance to examine how different demographic variables including gender, age, and experience in position influence patient safety culture in terms of dimensions rather than individual questions. In addition, one variable, year of data, has been added to demographic variables. That is, the physicians' data from 2011, 2012, and 2013 are listed as 1, 2, and 3 in the column from a longitudinal viewpoint. Bonferroni test is performed for post hoc analysis except for gender with only two levels.

The numbers of effective questionnaires of physicians from this case regional hospital in 2011, 2012, and 2013 are 46, 59, and 53, and the total number of physicians for analyses in this study is 158. The demographic variables including gender, age, and experience in position are summarized in Table 1. The majorities of physicians are males with the age of 21-60 years old along with 3 to 20 years in experience in position.

Table 1 Demographic Variables of Physicians

| Gender | Frequency |
|---|---|
| Male | 102 |
| Female | 56 |
| Age | Frequency |
| Less than 20 | 1 |
| 21-30 | 29 |
| 31-40 | 47 |
| 41-50 | 57 |
| 51-60 | 23 |
| 61 or over | 1 |
| Experience in Position | Frequency |
| Less than 6 months | 28 |
| 6 to 11 months | 4 |
| 1 to 2 years | 15 |
| 3 to 4 years | 31 |
| 5 to 10 years | 30 |
| 11 to 20 years | 49 |
| 21 years or more | 1 |

In multivariate analysis of variance, there are four common approaches to test if $H_0$: $\mu_1 = \mu_2 = \ldots = \mu_k$, where $k$ is the number of multivariate normal populations, is to be rejected with different levels of powers, including Wilk's likelihood ration test, Pillai test, Lawley-Hotelling test, and Roy's largest root [14]. When the mean vectors of dependent variables are collinear, Roy's largest root is more power than the other three test statistics [15]. Conditional index (CI) can be an index to examine if the mean vectors of dependent variables are collinear [16]. In this study, the range of CI values falls between 10 and 46, indicating the collinearity exists [16]. Therefore, Roy's largest root is more powerful to detect if $H_0$ is to be rejected with $\alpha = 0.05$. Table 2 depicts the MANOVA test statistics of using Roy's largest root, where the p-values of experience in position and five interaction

effects including gender * experience in position, gender * year of data, age * experience in position, age * year of data, and experience in position * year of data are less than 0.05.

Table 2 MANOVA Test Statistics

| Effect | Value | F | Sig. |
|---|---|---|---|
| Experience in Position | .421 | 4.402 | .000 |
| Gender * Experience in Position | .234 | 2.396 | .017 |
| Gender * Year of Data | .205 | 2.068 | .041 |
| Age * Experience in Position | .298 | 2.658 | .005 |
| Age * Year of Data | .353 | 3.682 | .001 |
| Experience in Position * Year of Data | .331 | 3.601 | .001 |

The next step is to examine how demographic variables and interaction effects affect these nine dimensions in patient safety culture. The results show that experience in position has significant impacts on hospital management support for patient safety (p-value = 0.039) and hospital handoffs and transitions (p-value = 0.039). The post hoc analyses should be conducted. However, from Table 1, there is only one physician falling in 21 years or more in experience in position. Thus, this item should be removed in order to perform Bonferroni test. In our study, we perform one-way MANOVA again and conclude that experience in position has no longer impacts on hospital management support for patient safety (p-value = 0.374) and hospital handoffs and transitions (p-value = 0.088) statistically.

From Table 2, in order to perform Bonferroni test if necessary in interaction effects, the number of questionnaire has been reduced from 158 to 157 by excluding one physician with 21 years or more experience in position. In addition, the levels of gender, age, experience in position, and year of data are 2, 5, 6, and 3, respectively. To further examine the interaction effects, the possible combinations of gender and experience in position, gender and year of data, age and experience in position, age and year of data, and experience in position and year of data are 12 (2×6), 6 (2×3), 30 (5×6), 15 (5×3), and 18 (6×3), respectively.

For the interaction effect of gender and experience in position, stress recognition is the only dimension with the p-value of 0.011 less than 0.05 when one-way MANOVA is performed on nine dimensions as shown in Table 3. From descriptive analyses, male physicians with 11-20 years of experience in position have the highest satisfaction value in stress recognition. In contrast, female physicians with 6-11 months of experience in position have the lowest satisfaction in stress recognition. However, Bonferroni test does not show any significant differences among twelve combinations. For the interaction effect of gender and year of data with six combinations, stress recognition is the only significant dimension depicted in Table 4. From descriptive statistics, male physicians in 2013 (Combination 3) have the highest satisfaction value in stress recognition, whereas male physicians in 2011 (Combination 1) have the lowest satisfaction value in stress recognition. By further applying Bonferroni test, Table 5 shows that male physicians in 2013 (Combination 3) are more satisfied than female physicians in 2011 (Combination 4) in stress recognition statistically.

Table 3 The Interaction Effect of Gender and Experience in Position in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Gender * Experience in Position | Teamwork Climate | 159.087 | 11 | 14.462 | 1.120 | .350 |
| | Safety Climate | 210.127 | 11 | 19.102 | 1.337 | .210 |
| | Job Satisfaction | 237.886 | 11 | 21.626 | 1.319 | .219 |
| | Stress Recognition | 319.409 | 11 | 29.037 | 2.340 | .011* |
| | Perception of Management | 128.801 | 11 | 11.709 | 1.349 | .203 |
| | Working Condition | 101.090 | 11 | 9.190 | 1.250 | .260 |
| | Hospital Management Support for Patient Safety | 23.105 | 11 | 2.100 | 0.796 | .643 |
| | Teamwork across Hospital Units | 32.181 | 11 | 2.926 | 1.100 | .365 |
| | Hospital Handoffs and Transitions | 69.987 | 11 | 6.362 | 1.006 | .444 |

Table 4 The Interaction Effect of Gender and Year of Data in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Gender * Year of Data | Teamwork Climate | 27.886 | 5 | 5.577 | 0.420 | .834 |
| | Safety Climate | 77.713 | 5 | 15.543 | 1.065 | .382 |
| | Job Satisfaction | 44.436 | 5 | 8.887 | 0.522 | .759 |
| | Stress Recognition | 173.925 | 5 | 34.785 | 2.701 | .023* |
| | Perception of Management | 13.406 | 5 | 2.681 | 0.295 | .915 |
| | Working Condition | 16.500 | 5 | 3.300 | 0.433 | .825 |
| | Hospital Management Support for Patient Safety | 13.204 | 5 | 2.641 | 1.016 | .410 |
| | Teamwork across Hospital Units | 3.501 | 5 | 1.700 | 0.627 | .679 |
| | Hospital Handoffs and Transitions | 34.598 | 5 | 6.920 | 1.097 | .365 |

Table 5 Multiple Comparison for Gender and Year of Data in Stress Recognition

| (I) Gender * Year of Data | (J) Gender * Year of Data | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 3 | 4 | 2.36 | .037* |

No any significant difference is found on nine dimensions for the interaction effect of age and experience in position with 23 combinations as shown in Table 6. For the interaction effect of age and year of data, there are only thirteen combinations. In addition, there is one combination with the sample of size one. Thus, we remove that sample and then perform one-way MANOVA again with twelve combinations. The result shows that stress recognition is the only significant dimension depicted in Table 7. From descriptive analyses, physicians with 31-40 years old in 2013 (Combination 13) have the highest satisfaction value in stress recognition among twelve combinations followed by physicians with 51-60 years

old in 2013 (Combination 15) and physicians with 51-60 years old in 2012 (Combination 10). In contrast, physicians with 51-60 years old in 2011 (Combination 5) has the lowest satisfaction value followed by physicians with 21-30 years old in 2011 (Combination 2). By Bonferroni test, Table 8 summarizes the statistically significant multiple comparisons of age and year of data in stress recognition. Bonfernoi test shows that physicians with 21-30 years old in 2011 (Combination 2) have lower satisfaction value in stress recognition than physicians with 31-40 years old in 2013 (Combination 13) and physicians with 51-60 years old in 2013 (Combination 15) statistically. Finally, the interaction effect of experience in position and year of data with eighteen combinations does not show any significant differences on nine dimensions. Table 9 summarizes the interaction effect values.

Table 6 The Interaction Effect of Age and Experience in Position in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Age * Experience in Position | Teamwork Climate | 234.178 | 22 | 10.644 | 0.794 | .729 |
| | Safety Climate | 197.251 | 22 | 8.966 | 0.576 | .933 |
| | Job Satisfaction | 473.775 | 22 | 21.535 | 1.348 | .153 |
| | Stress Recognition | 442.015 | 22 | 20.092 | 1.606 | .054 |
| | Perception of Management | 202.307 | 22 | 9.196 | 1.040 | .421 |
| | Working Condition | 173.061 | 22 | 7.866 | 1.060 | .398 |
| | Hospital Management Support for Patient Safety | 78.780 | 22 | 3.581 | 1.469 | .095 |
| | Teamwork across Hospital Units | 75.143 | 22 | 3.416 | 1.336 | .160 |
| | Hospital Handoffs and Transitions | 169.603 | 22 | 7.709 | 1.263 | .208 |

Table 7 The Interaction Effect of Age and Year of Data in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Age * Year of Data | Teamwork Climate | 213.275 | 12 | 17.773 | 1.408 | .169 |
| | Safety Climate | 216.476 | 12 | 18.040 | 1.258 | .250 |
| | Job Satisfaction | 288.532 | 12 | 24.044 | 1.488 | .135 |
| | Stress Recognition | 376.471 | 12 | 31.373 | 2.593 | .004* |
| | Perception of Management | 109.893 | 12 | 9.158 | 1.032 | .423 |
| | Working Condition | 67.802 | 12 | 5.650 | 0.740 | .711 |
| | Hospital Management Support for Patient Safety | 21.062 | 12 | 1.755 | 0.657 | .789 |
| | Teamwork across Hospital Units | 24.732 | 12 | 2.061 | 0.755 | .695 |
| | Hospital Handoffs and Transitions | 60.119 | 12 | 5.010 | 0.778 | .672 |

Table 8 Multiple Comparison for Age and Year of Data in Stress Recognition

| (I) Age * Year of Data | (J) Age * Year of Data | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 2 | 13 | -5.13 | .014* |
| | 15 | -4.97 | .021* |

Table 9 The Interaction Effect of Experience in Position and Year of Data in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Experience in Position * Year of Data | Teamwork Climate | 268.152 | 17 | 15.774 | 1.244 | .239 |
| | Safety Climate | 211.633 | 17 | 12.449 | 0.836 | .650 |
| | Job Satisfaction | 318.818 | 17 | 18.754 | 1.135 | .327 |
| | Stress Recognition | 349.011 | 17 | 20.530 | 1.613 | .069 |
| | Perception of Management | 198.242 | 17 | 11.661 | 1.363 | .164 |
| | Working Condition | 160.963 | 17 | 9.468 | 1.308 | .196 |
| | Hospital Management Support for Patient Safety | 47.825 | 17 | 2.813 | 1.093 | .366 |
| | Teamwork across Hospital Units | 55.121 | 17 | 3.242 | 1.243 | .240 |
| | Hospital Handoffs and Transitions | 141.977 | 17 | 8.352 | 1.373 | .159 |

## 4. Conclusion

This study uses a longitudinal study which focuses solely on physicians to analyze how physicians in different years (2011, 2012, and 2013) perceive the patient safety culture in terms of nine dimensions from the Chinese version of Hospital Survey on Patient Safety Culture from a regional hospital in Taiwan by multivariate analysis of variance. Six effects including five interaction effects have been found to be significantly different. The major findings are summarized below. Though stress recognition is the only significant dimension for the interaction effect of gender and experience in position, Bonferroni test does not show any significant difference among twelve combinations. In addition, stress recognition is the only significant dimension for the interaction effect of gender and year of data. Bonferroni test shows that male physicians in 2013 are more satisfied than female physicians in 2011 in stress recognition statistically. Moreover, stress recognition is the significant dimension for the interaction effect of age and year of data. By Bonferroni test, physicians with 21-30 years old in 2011 perceive lower satisfaction in stress recognition than physicians with 31-40 years old in 2013 and physicians with 51-60 years old in 2013 statistically. With the above information, the hospital management can pay much attention to stress recognition and the interaction effects that perceive lower satisfaction in stress recognition.

## Acknowledgement

This study was partially supported by Cheng Ching General Hospital – Chung Kang Branch, Taichung City, Taiwan with grant number of CH10200159.

## References


1. Schutz, A.L., M.A. Counte, S. Meurer, Development of a patient safety culture measurement tool for ambulatory health care settings: analysis of content validity, *Health Care Management Science*, 10(2), 139-149, 2007.
2. Shie, H.G., W.C. Lee, H.F. Hsiao, H.L. Lin, L.L. Yang, F. Jung, Patient safety attitudes among respiratory therapists in Taiwan, *Respiratory Care*, 56, 1924-1929, 2011.
3. Zuniga, F., D. Schwappach, S. De Geest, R. Schwendimann, Psychometric properties of the Swiss version of the nursing home survey on patient safety culture, *Safety Science*, 55, 88-118, 2013.
4. Wagner, C., M. Smits, J. Sorra, C.C. Huang, Assessing patient safety culture in hospitals across countries, *International Journal for Quality in Health Care*, 25(3), 213-221, 2013.
5. Deilkas, E., D. Hofoss, Patient safety culture lives in departments and wards: multilevel partitioning of variance in patient safety culture, *BMC Health Services Research*, 10, 85, 2010.
6. Sexton, J.B., R.L. Helmreich, T.B. Neilands, K. Rowan, K. Vella, J. Boyden, P.R. Roberts, E.J. Thomas, The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging research, *BMC Health Services Research*, 6, 44, 2006.
7. Chen, I.C., H.H. Li, Measuring patient safety culture in Taiwan using the hospital survey on patient safety culture (HSOPSC), *BMC Health Services Research*, 10, 152, 2010.
8. Fujita, S., K. Seto, S. Ito, Y. Wu, C.C. Huang, T. Hasegawa, The characteristics of patient safety culture in Japan, Taiwan and the United States, *BMC Health Services Research*, 13, 20, 2013.
9. Lee, Y.-C., C.-L. Hsu, H.-H. Wu, L.-P. Hsieh, S.-J. Weng, C.-H. Huang, Evaluating patient safety culture from viewpoints of physicians and registered nurses – a case of a regional hospital in Taiwan, *3rd International Conference on Application of Information and Communication Technology and Statistics in Economy and Education*, 667-676, 2013.
10. Ulrich, B., T. Kear, Patient safety and patient safety culture: foundations of excellent health care delivery, *Nephrology Nursing Journal*, 41(5), 447-456 & 505, 2014.
11. Lee, W.C., H.Y. Wung, H.H. Liao, C.M. Lo, F.L. Chang, P.C. Wang, A. Fan, H.H. Chen, H.C. Yang, S.M. Hou, Hospital safety cCulture in Taiwan: a nationwide survey using Chinese version safety attitude questionnaire, *BMC Health Services Research*, 10, 234, 2010.
12. Lee, L.J., C.H. Hsu, J.C. Hong, Explore the impact of hospital restructuring medical practitioners of patient safety culture attitudes, *Journal of Health Management*, 10(2), 189-199, 2012.
13. Lee, Y.-C., C.-H. Huang, S.-J. Weng, L.-P. Hsieh, H.-H. Wu, Identifying critical factors of patient safety culture – a case of a regional hospital in Taiwan, *International Journal of Innovation, Management and Technology*, 5(3), 183-188, 2014.
14. Wu, H.-H., H.-W. Hsueh, C.-H. Tsai, Analyzing service process satisfaction of a tourism factory by multivariate analysis of variance - a case of Brand's Health Museum, *Technics Technologies Education Management*, 9(2), 402-409, 2014.
15. Rencher, A.C., *Multivariate Statistical Inference and Applications*, John Wiley & Sons, New York, 1998.
16. Belsley, DA, E. Kuh, R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York, 1980.

# Using Multivariate Analysis of Variance to Evaluate Patient Safety Culture – A Case of A Regional Hospital in Taiwan

**Yii-Ching Lee[1,2,3], Xian-Yin Liao[4], Shao-Jen Weng[5], Chih-Hsuan Huang[6], Wan-Lin Hsieh[7], Liang-Po Hsieh[8], Hsin-Hung Wu[9]**

[1] School of Health Policy and Management, Chung Shan Medical University, Taichung City, Taiwan

[2] Human Resource Management, Cheng Ching General Hospital-Chung Kang Branch, Taichung City, Taiwan

[3,8] Department of Medical Quality Management, Cheng Ching General Hospital-Chung Kang Branch, Taichung City, Taiwan

[3]yiiching.lee@gmail.com

[8]lphsieh624@yahoo.com.tw

[4,9] Department of Business Administration, National Changhua University of Education, Changhua, Taiwan

[4]ohohinin@gmail.com

[9]hhwu@cc.ncue.edu.tw

[5,7] Department of Industrial Engineering and Enterprise Information, Tunghai University, Taichung City, Taiwan

[5]sjweng@thu.edu.tw

[7]hsiehwl@thu.edu.tw

[6] Department of Business Administration, Tunghai University, Taichung City, Taiwan

tititacer@gmail.com

**Abstract.** This study is to observe the attitudes from physicians and nurses toward patient safety culture. Multivariate analysis of variance is applied to evaluate how different demographic variables including gender, age, experience in organization, and education impact nine dimensions of patient safety culture. Bonferroni test is performed for post hoc analysis except for gender with only two levels. The major findings are summarized below. Both age and experience in organization are critical demographic variables to affect patient safety culture. However, these two variables do not have significant influences on any dimensions. In contrast, the interaction effect of gender and experience in position has significant impacts on seven out of nine dimensions except for stress recognition and hospital management support for patient safety. Generally, females with less than 6 months perceive better than females with 3-4 years in teamwork climate, job satisfaction, perception of management, working condition, and teamwork across hospital units but perceive poorly than females with 3-4 years in hospital handoffs and transitions statistically.

**Keywords.** Patient safety culture, multivariate analysis of variance, Bonferroni test, demographic variable, Chinese version of patient safety culture.

## 1.Introduction

Sexton et al. [1] stated that healthcare quality and safety should be evaluated within the system and contextual factors including organizational factors, work environment factors, team factors, and staff factors. Shie et al. [2] concluded that better attitude toward patient

safety could result in shorter stay, fewer prescription errors, less ventilator-associated pneumonia, fewer blood-stream infections, and lower mortality. In fact, hospitals with a more open culture and reflective attitude toward errors and patient safety could reduce the number of accidents and failures [3].

The hospital survey on patient safety culture with good validity and reliability developed by Sexton et al. [1] has been widely used to assess perceptions of patient safety culture from the entire healthcare organization's staff [4,5]. Lee et al. [6] pointed out that the core staffs are medical physicians and registered nurses in the hospital. Thus, it is of interest for hospital management to observe the attitude from both physicians and nurses toward patient safety culture with different demographic variables and their combinations from time to time. In this study, multivariate analysis of variance (MANOVA) is applied for analyses when demographic variables are taken into account.

## 2.Patient Safety Culture

There are six dimensions in the hospital survey on patient safety culture originally developed by Sexton et al. [1], including teamwork climate, safety climate, job satisfaction, stress recognition, perception of management, and working condition. The definitions of six dimensions are as follows [1]. Teamwork climate is defined as the perceived quality of collaboration between personnel. Safety climate is the perceptions of a strong and proactive organizational commitment to safety. Perception of management is the approval of managerial actions. Job satisfaction is defined as the positivity about the work experience. Stress recognition is to measure how performance is influenced by stressors. Finally, working condition is the perceived quality of the work environment and logistical support such as staffing and equipment. In 2008, the Chinese version of hospital survey on patient safety culture was developed based upon the translation of the hospital survey on patient safety culture [2,5]. The survey has six dimensions and thirty questions.

Later, three hospital-level aspects of safety culture borrowed from Agency for Healthcare Research and Quality were incorporated into the Chinese version of hospital survey on patient safety culture, i.e., hospital management support for patient safety, teamwork across hospital units, and hospital handoffs and transitions [7]. Hospital management support for patient safety is defined as the hospital management provides a work climate that promotes patient safety and shows patient safety is a top priority. Teamwork across hospital units is that hospital units cooperate and coordinate with one another to provide the best care for patients. Hospital handoffs and transitions is the extent to which important patient care information is transferred across hospital units and during shift changes. Therefore, the Chinese version has become nine dimensions and 41 questions.

In hospitals, the core staffs are physicians and nurses [6]. In addition, nurses are the biggest workforce in the healthcare organizations and can possess enough power to move the underlying organizational culture to a patient safety culture [8]. Thus, it is worth to observe the attitude from physicians and nurses toward patient safety culture for hospital management. During the survey, respondents who are health organizations' staffs are required to answer questions by rating each question based on a five-point Likert scale, ranging from strongly agree to strongly disagree, or frequency such as never, rarely, sometimes, most of the time, and always [1].

## 3.A Case Study

This study is to observe the attitudes from physicians and nurses toward patient safety culture. This study uses multivariate analysis of variance to examine how different demographic variables including gender, age, experience in position, and education influence patient safety culture in terms of dimensions rather than individual questions. Bonferroni test is performed for post hoc analysis except for gender with only two levels. The data are from a regional hospital in Taiwan in 2013, and the effective number of questionnaire is 470. The demographic variables including gender, age, experience in position, and education are summarized in Table 1. The majorities of respondents are females falling with 21-50 years old with college/university degree and have one to twenty years in experience in position.

Table 1 Demographic Variables of Physicians and Nurses

| | Frequency | | Frequency |
|---|---|---|---|
| Gender | | Education | |
| Male | 64 | Junior high school and below | 1 |
| Female | 406 | Senior high school | 3 |
| | | College/University | 439 |
| | | Graduate school | 27 |
| Experience in Position | | Age | |
| Less than 6 months | 60 | Less than 20 | 10 |
| 6 to 11 months | 25 | 21-30 | 189 |
| 1 to 2 years | 94 | 31-40 | 180 |
| 3 to 4 years | 87 | 41-50 | 75 |
| 5 to 10 years | 127 | 51-60 | 15 |
| 11 to 20 years | 71 | 61 or over | 1 |
| 21 years or more | 6 | | |

In performing multivariate analysis of variance, four common approaches have been used to test if $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$, where $k$ is the number of multivariate normal populations, is to be rejected with different levels of powers, including Wilk's likelihood ration test, Pillai test, Lawley-Hotelling test, and Roy's largest root [9]. Rencher [10] stated that Roy's largest root is more powerful than the other three test statistics when the mean vectors of dependent variables are collinear. Conditional index (CI) can be an index to examine whether or not the mean vectors of dependent variables are collinear [11]. The range of CI is this study falls between 9 and 46, indicating the collinearity exists [11]. Thus, Roy's largest root is chosen to determine if $H_0$ is to be rejected with $\alpha = 0.01$. Table 2 summarizes the MANOVA test statistics of using Roy's largest root, where the p-values of age, experience in position, gender * experience in position, and age * experience in position are less than 0.01.

Table 2 MANOVA Test Statistics

| Effect | Value | F | Sig. |
|---|---|---|---|
| Age | 0.064 | 2.890 | .003 |
| Experience in Position | 0.111 | 5.041 | .000 |
| Gender * Experience in Position | 0.058 | 2.629 | .006 |
| Age * Experience in Position | 0.137 | 4.308 | .000 |

The next step is to examine how demographic variables and interaction effects affect these nine dimensions. From Table 3, age does not have any significant impacts on nine dimensions statistically. In addition, from Table 4 experience in position does not have any significant influences on nine dimensions.

Table 3 Age in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Age | Teamwork Climate | 53.797 | 5 | 10.759 | 0.955 | .445 |
| | Safety Climate | 81.631 | 5 | 16.326 | 0.905 | .478 |
| | Job Satisfaction | 62.597 | 5 | 12.519 | 0.730 | .601 |
| | Stress Recognition | 18.774 | 5 | 3.755 | 0.315 | .904 |
| | Perception of Management | 64.603 | 5 | 12.921 | 1.544 | .175 |
| | Working Condition | 48.041 | 5 | 9.608 | 1.192 | .312 |
| | Hospital Management Support for Patient Safety | 9.981 | 5 | 1.996 | 0.760 | .579 |
| | Teamwork across Hospital Units | 22.324 | 5 | 4.465 | 1.941 | .087 |
| | Hospital Handoffs and Transitions | 41.872 | 5 | 8.374 | 1.268 | .277 |

Table 4 Experience in Position in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Experience in Position | Teamwork Climate | 51.194 | 6 | 8.532 | 0.757 | .604 |
| | Safety Climate | 68.276 | 6 | 11.379 | 0.631 | .706 |
| | Job Satisfaction | 134.092 | 6 | 22.349 | 1.303 | .254 |
| | Stress Recognition | 66.447 | 6 | 11.074 | 0.930 | .473 |
| | Perception of Management | 93.888 | 6 | 15.648 | 1.870 | .085 |
| | Working Condition | 105.025 | 6 | 17.504 | 2.171 | .045 |
| | Hospital Management Support for Patient Safety | 38.058 | 6 | 6.343 | 2.415 | .026 |
| | Teamwork across Hospital Units | 38.212 | 6 | 6.369 | 2.768 | .012 |
| | Hospital Handoffs and Transitions | 95.906 | 6 | 15.984 | 2.420 | .026 |

The levels of gender, age, and experience in position are 2, 6, and 7, respectively. To examine the interaction effects, the possible combinations of gender and experience in position and age and experience in position are 14 (2×7) and 42 (6×7), respectively. For the interaction effect of gender and experience in position, there are 14 combinations but one combination (Combination 2, i.e., males with 6-11 months in experience in position) has

only one sample. Thus, this sample should be removed in order to perform Bonferroni test. In our study, we perform one-way MANOVA again with 13 combinations and conclude that the interaction effect of gender and experience in position is significant on seven out of nine dimensions except for stress recognition and hospital management support for patient safety as shown in Table 5.

Table 5 The Interaction Effect of Gender and Experience in Position in Nine Dimensions

| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Gender * Experience in Position | Teamwork Climate | 616.523 | 12 | 51.377 | 4.591 | .000* |
| | Safety Climate | 494.845 | 12 | 41.237 | 2.359 | .006* |
| | Job Satisfaction | 717.434 | 12 | 59.786 | 3.582 | .000* |
| | Stress Recognition | 294.477 | 12 | 24.540 | 1.959 | .026 |
| | Perception of Management | 308.150 | 12 | 25.679 | 3.000 | .000* |
| | Working Condition | 289.549 | 12 | 24.129 | 2.995 | .000* |
| | Hospital Management Support for Patient Safety | 47.685 | 12 | 3.974 | 1.525 | .111 |
| | Teamwork across Hospital Units | 105.856 | 12 | 8.821 | 3.689 | .000* |
| | Hospital Handoffs and Transitions | 325.707 | 12 | 27.142 | 4.129 | .000* |

For interaction effect of gender and experience in position, females with 21 years or more (Combination 14) have the highest perception in teamwork climate but females with 3-4 years (Combination 11) have the lowest perception by descriptive statistics. Bonferroni test, depicted in Table 6, shows males with 21 years or more (Combination 6) perceive significantly better teamwork climate than females with 3-4 years (Combination 11) and females with 5-10 years (Combination 12). Females with less than 6 months (Combination 8) perceive better teamwork climate than females with 3-4 years (Combination 11) and females with 5-10 years (Combination 12). In addition, females with 21 years or more (Combination 14) have better teamwork climate than females with 3-4 years (Combination 11).

Table 6 Multiple Comparison for Gender and Experience in Position in Teamwork Climate

| (I) Gender * Experience in Position | (J) Gender * Experience in Position | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 6 | 11 | 3.86 | .001* |
| | 12 | 3.42 | .006* |
| 8 | 11 | 2.95 | .001* |
| | 12 | 2.51 | .008* |
| 14 | 11 | 4.43 | .004* |

In safety climate, males with 21 years or more (Combination 7) and females with 3-4 years (Combination 11) have the respective highest and lowest satisfaction values. However, Bonferroni test does not show any significant differences among thirteen combinations. In job satisfaction, males with 21 years or more (Combination 7) and females with 3-4 years (Combination 11) have the respective highest and lowest satisfaction values. Bonferroni test

illustrated in Table 7 shows that females with less than 6 months (Combination 8) have significant better satisfaction than females with 3-4 years (Combination 11).

Table 7 Multiple Comparison for Gender and Experience in Position in Job Satisfaction

| (I) Gender * Experience in Position | (J) Gender * Experience in Position | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 8 | 11 | 3.35 | .005* |

Females with less than 6 months (Combination 8) and females with 3-4 years (Combination 11) have the respective highest and lowest satisfaction values in perception of management by descriptive statistics. From Table 8, Bonferroni test states that females with less than 6 months (Combination 8) perceive significant better perception of management than females with 3-4 years (Combination 11) and females with 5-10 years (Combination 12). In working condition, descriptive statistics show that males with 21 years or more (Combination 7) have the highest satisfaction, and females with 3-4 years (Combination 11) have the lowest satisfaction. Bonferroni test, depicted in Table 9, indicates females with less than 6 months (Combination 8) perceive significant better working conditions than females with 3-4 years (Combination 11).

Table 8 Multiple Comparison for Gender and Experience in Position in Perception of Management

| (I) Gender * Experience in Position | (J) Gender * Experience in Position | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 8 | 11 | 2.76 | .000* |
| | 12 | 2.35 | .003* |

Table 9 Multiple Comparison for Gender and Experience in Position in Working Condition

| (I) Gender * Experience in Position | (J) Gender * Experience in Position | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 8 | 11 | 2.24 | .008* |

In teamwork across hospital units, descriptive statistics show that females with 21 years or more (Combination 14) and females with 6-11 months (Combination 9) have highest and lowest satisfaction values, respectively. Table 10 which summarizes Bonferroni test indicates females with less than 6 months (Combination 8) have significant better perceptions in teamwork across hospital units than females with 3-4 years (Combination 11). In hospital handoffs and transitions, females with 5-10 years (Combination 10) and females with 21 years or more (Combination 14) have highest and lowest perceptions, respectively. Table 10 further indicates that both females with less than 6 months (Combination 8) and females with 21 years or more (Combination 14) perceive significantly lower than females with 5-10 years (Combination 12) by Bonferroni test. In addition, females with less than 6 months (Combination 8) perceive lower than females with 5-10 years (Combination 10) and females with 3-4 years (Combination 11).

Table 10 Multiple Comparison for Gender and Experience in Position in Teamwork across Hospital Units

| (I) Gender * Experience in Position | (J) Gender * Experience in Position | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 8 | 11 | 1.31 | .003* |

Table 11 Multiple Comparison for Gender and Experience in Position in Hospital Handoffs and Transitions

| (I) Gender * Experience in Position | (J) Gender * Experience in Position | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| 8 | 10 | -2.13 | .004* |
| | 11 | -2.36 | .000* |
| | 12 | -2.38 | .000* |
| 14 | 12 | -3.22 | .007* |

For the interaction effect of age and experience in position, there are 27 combinations, where 15 combinations have the sample of size zero. Table 12 summarizes these twenty seven combinations along with their respective sample of sizes. Eight out of twenty seven combinations have the sample of size 30 or more, indicating that nineteen combinations might not follow the underlying assumption that samples are from normal distributions. Due to smaller sample of sizes, the interaction effect of age and experience in position is not going to be analyzed.

Table 12 Twenty Seven Combinations of the Interaction Effect of Age and Experience in Position

| Combination | Sample Size | Combination | Sample Size | Combination | Sample Size |
|---|---|---|---|---|---|
| 1 | 9 | 17 | 15 | 27 | 46 |
| 5 | 1 | 18 | 35 | 28 | 9 |
| 8 | 30 | 19 | 59 | 29 | 3 |
| 9 | 12 | 20 | 61 | 31 | 1 |
| 10 | 65 | 22 | 4 | 32 | 1 |
| 11 | 45 | 23 | 1 | 33 | 1 |
| 12 | 37 | 24 | 2 | 34 | 6 |
| 15 | 6 | 25 | 4 | 35 | 3 |
| 16 | 4 | 26 | 9 | 42 | 1 |

## 4.Conclusions

This study uses multivariate analysis of variance to observe the attitudes from physicians and nurses toward patient safety culture with different demographic variables and combinations from a regional hospital in Taiwan. The effective sample of size in 2013 in this case hospital is 470. In addition, Bonferroni test is performed for post hoc analysis except for gender with only two levels. The results show that different ages do not result in significant differences on nine dimensions statistically. In addition, different experiences in position do not have any significant differences among nine dimensions. On the other hand, the interaction effect of gender and experience in position has significant impacts on seven out of nine dimensions. Specifically, both males with 21 years or more and females with less than 6 months perceive significantly better teamwork climate than females with 3-4 years

and females with 5-10 years. Besides, females with 21 years or more have better teamwork climate than females with 3-4 years.

In addition to teamwork climate, females with less than 6 months perceive better than females with 3-4 years statistically in job satisfaction, perception of management, working condition, and teamwork across hospital units. In hospital handoffs and transitions, both females with less than 6 months and females with 21 years or more perceive significantly lower satisfaction than females with 5-10 years. Furthermore, females with less than 6 months perceive lower satisfaction than females with 5-10 years and females with 3-4 years. With the above information, the hospital management can pay much attention to the interaction effect of gender and experience in position that has significant impacts on teamwork climate, safety climate, job satisfaction, perception of management, working condition, teamwork across hospital units, and hospital handoffs and transitions.

## Acknowledgement

This study was partially supported by Cheng Ching General Hospital – Chung Kang Branch, Taichung City, Taiwan with grant number of CH10200159.

## References

1. Sexton, J.B., R.L. Helmreich, T.B. Neilands, K. Rowan, K. Vella, J. Boyden, P.R. Roberts, E.J. Thomas, The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging research, *BMC Health Services Research*, 6, 44, 2006.
2. Shie, H.G., W.C. Lee, H.F. Hsiao, H.L. Lin, L.L. Yang, F. Jung, Patient safety attitudes among respiratory therapists in Taiwan, *Respiratory Care*, 56, 1924-1929, 2011.
3. Wagner, C., M. Smits, J. Sorra, C.C. Huang, Assessing patient safety culture in hospitals across countries, *International Journal for Quality in Health Care*, 25(3), 213-221, 2013.
4. Chen, I.C., H.H. Li, Measuring patient safety culture in Taiwan using the hospital survey on patient safety culture (HSOPSC), *BMC Health Services Research*, 10, 152, 2010.
5. Fujita, S., K. Seto, S. Ito, Y. Wu, C.C. Huang, T. Hasegawa, The characteristics of patient safety culture in Japan, Taiwan and the United States, *BMC Health Services Research*, 13, 20, 2013.
6. Lee, Y.-C., C.-L. Hsu, H.-H. Wu, L.-P. Hsieh, S.-J. Weng, C.-H. Huang, Evaluating patient safety culture from viewpoints of physicians and registered nurses – a case of a regional hospital in Taiwan, *3rd International Conference on Application of Information and Communication Technology and Statistics in Economy and Education*, 667-676, 2013.
7. Lee, L.J., C.H. Hsu, J.C. Hong, Explore the impact of hospital restructuring medical practitioners of patient safety culture attitudes, *Journal of Health Management*, 10(2), 189-199, 2012.
8. AbuAlrub, R.F., E.H.A. Alhijaa, The impact of educational interventions on enhancing perceptions of patient safety culture among Jordanian senior nurses, *Nursing Forum*, 49(2), 139-150, 2014.
9. Wu, H.-H., H.-W. Hsueh, C.-H. Tsai, Analyzing service process satisfaction of a tourism factory by multivariate analysis of variance - a case of Brand's Health Museum, *Technics Technologies Education Management*, 9(2), 402-409, 2014.
10. Rencher, A.C., *Multivariate Statistical Inference and Applications*, John Wiley & Sons, New York, 1998.
11. Belsley, DA, E. Kuh, R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York, 1980.

# The Search for Adaptive Methods Filtering Out High-Noises and Their Application on Oil Complexes' Control Telemetry Systems

**[*]Ozgur USTUN, [*]Murat YILMAZ, [**]Parviz Ali ZADA, [**]Ramazan Nejat TUNCAY**
[*]Istanbul Technical University, Electrical Engineering Dept., 34469, Istanbul, Turkey
[**]OKAN University, Tuzla Campus, 34959, Istanbul, Turkey

**Abstract.** This paper's domain is the application of well-known modern applied mathematical methods in a control structure of oil industry. Different types of Electrical Submersible Pumps (ESP) are running on worldwide oil fields. The ESP telemetry system allows for the obtaining of information in the neighborhood of different heavy electromagnetic noises. The telemetry system needs to receive accurate information on the pump unit's intake pressure, temperature and most importantly for the submersible electric motor, the stator cooling oil insulation resistance, for the successful exploitation of the ESP. The change from analog to digital telemetry techniques is moving quickly in almost all technical fields. But increasing disturbance levels with the corrupted analog telemetry signal are resulting in increasing noise levels; however, often it is still audible or the control system is still reliable. Though, beyond a certain disturbance level, the so-called digital cliff, the digital telemetry signal and control may stop abruptly. In this paper, an analog signal processing implementation was researched for the detection of the most efficient adaptive noise-cancelling filters among dozens of recognized ones for oil industry ESP telemetry systems of under severely noisy conditions. From ten applied adaptive filter algorithms, only four had shown successfully good results in the early prediction of the ESP electric motor real insulation disruption.

**Key Words:** Submersible motor, adaptive filtering, oil industry, noise, signals, communication-telemetry channels.

## 1.Introduction

More than a thousand switchboards of Electro-submersible Pump (ESP) under different trademarks are running in the oil fields, representing a wide spectrum of varied machinery which is working on the problems in oil production and its optimization. Not only are there simple devices of ESP electric motor control, but also there are complicated electronic complexes for installing an oil well borehole for operational duty as well. The main problem is to preset supporting parameters during oil production at several well borehole clusters simultaneously. The submersible telemetry system usually allows for the obtaining of information on the pump unit's intake pressure, temperature and most importantly for a submersible electric motor stator coil, its insulation resistance, for the successful exploitation of the oil complexes in the neighborhood of different types of heavy electromagnetic noises such as: random, pulsing, harmonic, etc. [1–3].

The change from analog to digital communication techniques is moving quickly in almost all technical fields. The emission limits for radiated and conducted disturbances, as prescribed by international standards, are based on their possible impact on analog telemetry. When the existing analog telemetry technologies are substituted with digital technologies, interference could occur even though all equipment complies.

Increasing disturbance levels with corrupted analog telemetry result in an increasing noise level, but it's often still "audible" or the control is still reliable. Digital telemetry technology is much better in suppressing disturbances up to a certain level. With increasing disturbance levels, the analog signal will remain low, but audible. However, beyond a certain disturbance level, the digital telemetry and control stops abruptly. This, the so called "digital cliff" at point C (Figure 1), makes it more complicated with digital telemetry to know the "headroom," (or the disturbance critical level), before the digital telemetry and control unexpectedly stops. The analog signal after point C will be lower, but still audible, and the control system is still reliable.

The success of these noise-cancelling methods and filters is depending mostly on the noise factor (signal/noise ratio) and also on the control signal character under consideration: close to random, exponential, voice, sinusoidal, etc. Most of the publications in the field of noise cancellation methods and their applications deal with rather big signal/noise ratio (that is – noise << signal) and show good achievements for cell phones, radio/TV technique, tape recorders, apparatus for people with hearing loss and concert hall equipment. There are some cases when the signal/noise ratio is around 1 (noise ≈ signal): command-and-control telemetry systems between centers and operators of winches, lorries, compressors workshops, metro, railways, controlled AC/DC drivers, as in the submersible pump telemetry case and etc. Not all of the noise-cancelling methods show good results, especially in real-

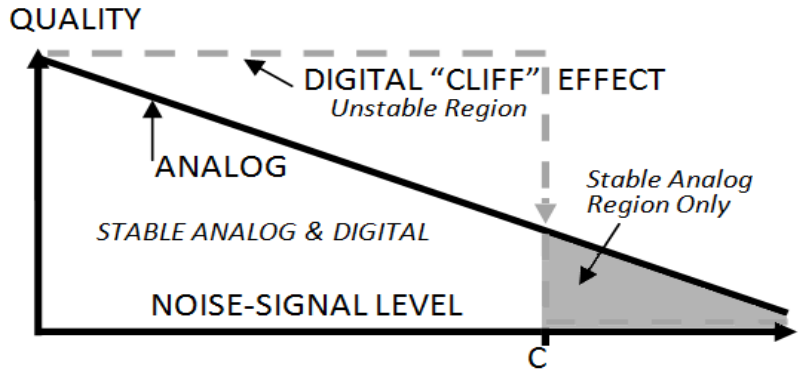



**Figure 1.** Analog transmissions experience gradual degradation; digital ones can experience sudden degradation-cliff.

time noise filtering, and need a special study to find the best filtering method for each engine's particular noise situation [4, 5].

The special cases are some severe ones which deal with the signal/noise ratios ≈ 0.2 ÷ 0.1 (noise ›≈ 5 ÷ 10 higher than useful signals): military command-and-control telemetry between centers, jet pilots and jets service teams of aircraft carriers battleships, in different metallurgical and especially arc furnaces telemetry between dispatcher and operators team, etc. There is an understandable shortage and inadequacy of any information and publications

not only in military fields, but also in the modern technological companies as well. A similar situation of low signal/noise ratio is presented in the field of the power systems, powerful controlled electrical machines and high voltage power electronics control-feedback signals. Any wrong interpretation of the signals due to high noise can bring unexpected accidents or malfunctions of large power systems or heavy duty electromechanical installations. Due to the mentioned digital cliff, the manufacturers of powerfully controlled electrical machines unfortunately still have to sometimes work with analog signals and equipment. This paper deals with the oil industry ESP electric motor, in particular with its control which is working close to, or in the neighborhood of, different heavy of electromagnetic noises such as: random, pulsing, harmonic and so on, that overwhelm the useful signals. Only a few of the MATLAB noise-cancelling methods - Adaptive Filtering Methods - present good real-time noise filtering results for the mentioned ESP severe noise cases [6-8]. Every other apparatus case needs a special study to find the best filtering method for the particular equipment [9–11]. Inductor-capacitive and build-in electronic active filters in some cases cannot fully protect the ESP telecommunication control system from different heavy "jam" of the power convertor electromagnetic noises. Thus, it is decided to search and employ the adaptive filters algorithms to add them into display/control system to improve filtration process from the different heavy "jam" of the electromagnetic noises. They can really help to raise higher reliability and survivability of the ESP. But, there are dozens and dozens of the adaptive filters algorithms. The authors have been applied and tested more than twenty Adaptive Filter algorithms to solve in real time the problem of early prediction of disruptions in the oil industry ESP induction motor. The best one among the analyzed adaptive filter algorithms for application in ESP telemetry system is recognized as normalized LMS method in the paper. The heavy and randomly changing data input requires a normalization process for limiting the values within a norm and a time-varying solution step size feature which is enhancing the accuracy of filtering calculations.

## 2. The Oil Industry Electrical Submersible Pumps (ESP)

ESP induction motors are applied as a drive for electrical centrifugal oil, deep-well pumps. They are usually put on the market in diameter/dimension groups of 103 to 180 mm, and power ranges from 12 to 500 kW, and even more. There are more than 60 different modern types of ESP of various capacities, which allow for picking up the optimal motor-pump to get the maximum possible efficiency for the particular oil field [12].

The techniques manufacturers have to keep are of the highest quality and performance reliability of the ESP motors because of:

1. Very expensive equipment often has to work in inaccessible places or far away from their technical service centers (the Arctic tundra locations, hot or cold deserts, sea jack-up platforms, swamps or lakes, etc.). In some cases, there are high requirements for reliability and survivability of the machines, since the equipment often may be serviced only once a year by a mobile team of technicians (using helicopters, ships or all-terrain track-type vehicles, etc.).

2. Any unexpected temporary equipment outage even for 10-20 minutes for any reason (mechanical, electrical, electronic, etc.), can lead to an array of problems for extended periods, especially during the winter season, such as frozen well liquid in output pipes, lubricating oil, etc.

3. In the so-called reach of sand (or clay) wells, the scenario can be even worse as any stoppage can be dramatic and sand may gradually settle down by gravity and block the ESP

subterranean reverse valve at the bottom of the well. This makes the automatic 'self-start' of the pump more difficult and longer, or even impossible, which in the winter may lead to a forced outage up until the next service team's visit. Moreover, it requires an expensive and complicated lift of the pump exploitation column and the cleaning of the ESP pipes and reverse valve from the sand and other debris.

To raise working survivability and to keep the highest quality, some ESP manufactures apply new techniques and measures, for example: 1) a stator made with the closed groove that raises the cleanliness of the motor's internal room of the engine and allows for the successful application of the winding grooves' firm insulation in the tube form; 2) the electric motor rotor has original bearings, having the mechanical fixings from any cranking; 3) the application of special, modern electro-technical materials allows for maintaining ESP motors at temperature strata liquids up to 120 °C and with superheat-resistant materials up to 160 °C; 4) 100% of ESP's have to be disassembled for all elements to be checked, and, after all of these have been checked, the ESP parts are thoroughly re-assembled and the ESP must be carefully tested in conditions similar to the real ones, including heating the ESP up to working temperatures, as would be in a well.

The control of the normally working motor resistance of the system "transformer-cable-ESP" isolation must not be less than ~ 350 kΩ. This isolation resistance usually decays very slowly over several months from ~ 350 kΩ to ~ 30 kΩ – but not less. The decaying process can take several months, even years, but the resistance of less than ~ 30 kΩ is potentially risky and can be a provocateur of any heavy internal short circuit in the stator winding and destroying some of its section. This means that it will be an expensive and complicated lift of the exploited column, removal of the burnt section of the motor, cleaning of the burnt section and then the re-wiring of it. The ESP motor is the most sensitive and expensive element of the ESP and control of its working motor insulation resistance must be very strong and effective [13, 14].

### 2.1. Structural Cart of the ESP Unit Equipment Complex

The ESP is usually fed from a 400 V network and consumes currents in-between 160–1200 A. ESP has a very complicated telemetry system. The ESP system consists of the following main parts: switchboard, step-up transformer, variable speed driver (VSD) convertor with built-in electronic active filter (FSA) and electrical submersible oil pump (ESP), as shown in Figure 2.

Among other ESP parameters, only the special ESP electric motor telemetry controls the following motors main parameters: insulation resistance (R, Ω) and temperature of the ESP motor

cooling oil ($t_0$, °C), the ESP motor axial and radial vibrations and also some oil pump technological parameters like the formation liquid temperature on the pump discharge ($t_p$, °C), pressure on the pump discharge and the formation of the liquid flow (P), etc.

ESP electrical motors are the most expensive parts of any ESP. To run the ESP under risky conditions is dangerous. Therefore, down-hole submersible telemetry units must control its electrical resistance in real-time very cautiously. It should be mentioned here, that usually telemetry input/output signals of down-hole units are around standards of 0–10 V or 4–20 mA. The ESP electric motors are very sensitive to their regime parameters. It is the cooling oil-filled motor (seldom distilled water), separated by a strong gasket from the pumping liquid, which is row oil mixed with very aggressive salty underground water. But the above-mentioned sand or clay in the underground water very slowly destroys the gasket;

so this salty and aggressive water may gradually enter into the motor cooling oil and slowly spoil its electrical insulation (resistance) from normal at the beginning - more than 350 kΩ, - to a risky one of less than 30 kΩ (and in a bad case - exponentially throughout one-to-three or more months). In bad case after one-two months the oil companies have to look for new ESP supplier with better gasket construction. Therefore, the bad cases after one-two months are very seldom, normal ESP regimes cases may be longer than 5-6 months (record for some ESP working under good sand/clay factor is 2-5 years, records up 10-12).

Telemetry signals are sent up from the down-hole unit to the display/control box on the earth surface by means of the same powerful motor supply cable between the ESP motor and the variable speed driver PWM convertor with a built-in electronic active filter. The PWM convertor waveform distortion factors of output voltage and currents correspondingly are: without filter - $K_{UOutput}$ = 43.73%, $K_{IOutput}$ = 6.75%, and after the filter is applied - $K_{UOutput}$ = 2.55%, $K_{IOutput}$ = 0.58% [3]. Theoretically, harmonics may affect such equipment in several ways but mainly: 1. Notches in the sinusoidal voltage can cause malfunctions due to a misfiring thyristor or IGBT; 2. Voltage harmonics may cause ignition beyond the required time; 3. A resonance in the presence of different types of equipment can lead to overstrain and machines shacking.

Unfortunately, inductor-capacitive and build-in electronic active filters in some cases cannot fully protect the ESP telecommunication control system from different heavy “jam” of the powerful PWM convertor electromagnetic noises: random, pulsing, harmonic and so on, which overwhelm the sensors useful feedback signals [27]. Thus, the adaptive filters algorithms must be added into display/control system to improve filtration process from the different heavy “jam” of the electromagnetic noises. They can really help to raise higher reliability and survivability of the ESP. But there are dozens and dozens of the adaptive filters algorithms. Which of them is the best for the ESP telecommunication control system, good, weak, bad or useless among approximately 20 adaptive filter methods? Can it be recommended for the ESP telecommunication control system to protect from different heavy “jam” of the powerful PWM convertor electromagnetic noises? It was heavy work of these adaptive filters multi-variant comparison in details.

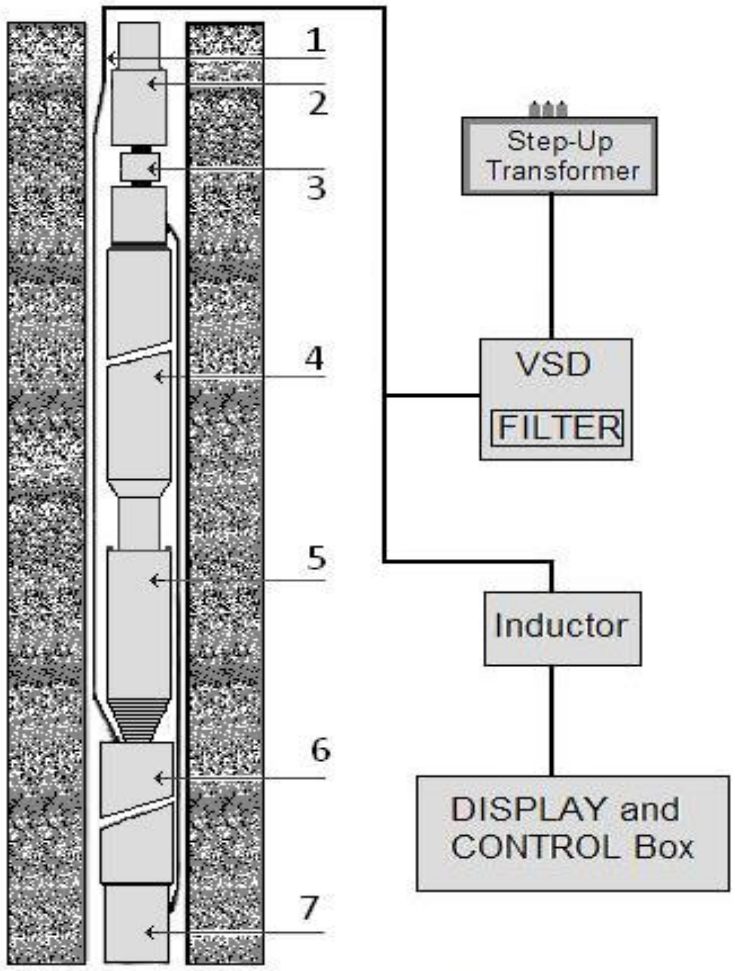



**Figure 2.** Structure cart of the ESP unit equipment complex: step-up transformer, display/control and inductor boxes, VSD convertor with build-in FSA, down-hole unit submersible telemetry system (TMSD Flow/TMSD or TMSD-2); ESP parts: 1) Power cable; 2) Tubing; 3) TMSDF/TMSDP system; 4) Pump; 5) Seal; 6) Electric Motor; 7) TMSD.

For ESP motors, waveform distortion factors of the output voltage and current are not affecting the ESP motor-pump performance activity. But voltage and current residual harmonics may be higher than some of the telemetry signals during some regimes and may confuse the display/control system and cause malfunctions in the equipment of the ESP (data errors, failures, even short circuits in the motor, etc.). The harmonics in power circuits make noises in the chain of the telemetry and control lines. This small noise leads to a certain discomfort, but as it increases, the transmitted information misinterpretation may appear; or it will decrease and in limited cases, the telemetry becomes impossible in totality. In the case of any technological changes in the electrical and telemetry, there should be considered the impact on telemetry lines, and importantly, on the controlled equipment. All of these regime

parameters controlled by ESP telemetry are very important for the ESP normal exploitation. But from the very sensitive ESP electric motor viewpoint, the most important parameter of the ESP motor is the resistance of its cooling oil, which depending on the sand factor; the gasket may wear out and provoke the leaking of salty underground liquid through the gasket and into the motor. This signal is transmitted from the down-hole unit upward to the display/control box of the ESP using also the same powerful motor feeding cable. This cable fills up by interference induction from currents, internal and external networking communication stray pick-ups, VSD convertor powerful PWM pulse interference and residual after-filter harmonics [15, 16]. Thus, this paper deals with an attempt to elevate the accuracy in the continuous interpreting of the R-signal from the above-mentioned corruptions by jams of interferences, harmonics and noises through the help of well-known modern adaptive methods. Any error in the interpretations of the R-signal may bring the wrongful early prediction of the critically low R-signal and the unreasonable expensive lift of the ESP for the motor cooling oil removal and its renewal service. In the worst case of late prediction, may cause a short circuit inside the motor leading to an emergency lift of the ESP for service and a more expensive restoration of the motor.

### 2.2 Interferences and Noises in ESP Telemetry

Thus, the jam of interferences and noises in all ESP telemetry, which accompany the useful telemetry signals, consist of: electromagnetic interference induction from cable currents, internal and external network communications stray pick-ups, VSD converter powerful pulse intervention and residual after filter harmonics: random, casual or incidental noise, etc. Analyzed were several typical for VSD with PWM converters harmonics spectra for the steady state regime of ESP and the typical share and fraction of harmonics values after FSA filters are presented for this case [3] as noted in the below Figure 3. Usually used to assess the impact of the various harmonics, the coefficient representing harmonics taken with certain weights will be used here. The most common are two factors: photometric weighting and C-transfer.

## 3. Short Overview to Adaptive Filter Algorithms

The least mean squares (LMS) and recursive least squares (RLS) methods are two different approaches in adaptive filter algorithms. The LMS method is used to adjust the filter coefficients in order to get the minimum cost function. Apart from RLS methods the LMS is relatively simple and requires less computational time because no matrix operations are used. The LMS algorithms which are less complicated due to the fewer computational and memory requirements surpass the RLS algorithms [6-8].

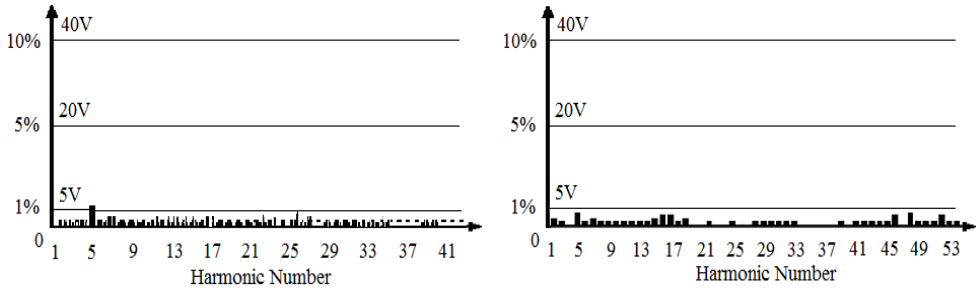



**Figure 3.** The typical share of current (a - 10-20th and 40-50th current harmonics of ~ 0.5-1.0% (~ 2-4 A) of $I_{Peak}$ value (374 A) of the motor current) and voltage (b - 10-20th and 40-50th voltage harmonics of ~ 0.5-1.0% (~ 3-5 V) of the motor voltage $V_{Peak}$ value (496 V)) harmonics values of the VSD PWM converter after build-in FSA.

For the standard LMS method the filter coefficients vector is given as:

$$\vec{w}(n+1) = \vec{w}(n) + \mu \cdot e(n) \cdot \vec{u}(n) \quad (1)$$

where $\mu$ is the step size of the adaptive filter, $\vec{w}(n)$ is the weight vector, $e(n)$ is the error signal and $\vec{u}(n)$ is the filter input vector. The value of the step size plays an important role in determining the filter convergence rate.

A modified version of standard LMS which is called as the normalized NLMS uses a time-varying step size $\mu(n)$:

$$\vec{w}(n+1) = \vec{w}(n) + \mu(n) \cdot e(n) \cdot \vec{u}(n) \quad (2)$$

The time varying step size which is reducing the calculation burden of adaptive filter is given as:

$$\mu(n) = \mu / \|\vec{u}(n)\|^2 \quad (3)$$

For digital signal processing (DSP) algorithms, it is required to exploit a simplified version of LMS method which is named as the sign LMS. The fundamental idea underlying the sign LMS algorithm is to use the sign function for implementing the adaptive filter algorithms to digital systems such as DSPs, FPGAs. In these algorithms the multiplication operation is a single operation for a given value of the selected parameter.

Using the sign function for adaptive filters can be yielded as sign-error, sign-data and sign-sign algorithms. The sign-error LMS algorithm is based on using the sign function of error.

$$\vec{w}(n+1) = \vec{w}(n) + \mu \cdot \mathrm{sgn}(e(n)) \cdot \vec{u}(n) \quad (4)$$

The sign-data LMS algorithm uses the sign function of the input signal vector while the sign-sign LMS algorithm uses both the sign function of both the input and the error as following:

$$\vec{w}(n+1) = \vec{w}(n) + \mu \cdot e(n) \cdot \operatorname{sgn}(\vec{u}(n)) \quad (5)$$

$$\vec{w}(n+1) = \vec{w}(n) + \mu \cdot \operatorname{sgn}(e(n)) \cdot \operatorname{sgn}(\vec{u}(n)) \quad (6)$$

Beside of the simplifying the calculation effort, the sign LMS algorithms have slower convergence speed and higher values of steady-state error [17-21].

## 4. An Objective and Comparative Evaluation of MATLAB Noise-Cancelling Adaptive Methods

For the comparative evaluation of MATLAB noise-filtering adaptive methods, here applies a twice heavier case for the study – voltages of the jam of accompanying harmonics $V_H$ =10 V each (not 3–5 V) and random noise $V_{RN}$ = 1 V (not 0.1–0.5 V). As mentioned above, the controlled and very useful exponential variable – the parameter of the ESP motor cooling oil R–resistance signal can decay from 10 V to 0.6 V. All of the MATLAB Adaptive Filtering Methods presented below were one-by-one tested under the above-mentioned conditions for R–resistance decreasing exponentially the signal, corrupted by the jam of the accompanying useful signal harmonics and noises. Afterwards, the R–resistance signal is recognized and analyzed [22–28].

### 4.1 The Mathematical Model of the Process Change of Insulation Oil Resistance

The insulation resistance controlling signal voltage can be presented as the following time variable function:

$$U_{R-insul} = 10e^{-t/a(t)} \text{ (Volt),} \quad (7)$$

Where *a(t)* is the insulation change coefficient, t = 0, 1.., 12 are the months, $U_{R\text{-}insul}$ is the motor winding isolation resistance sensor signal voltage, which is varies via time usually decay very slowly during months from initial 10V (~350kΩ) to risky 0.6V (~30kΩ) in a very knotty exponential manner (as a rule under irregular constant *a(t)*). This is why the very knotty exponential variable was presented as a solution of the differential equation Eq. (8): where the definition of *y* is the mentioned sensor signal voltage (*y(0)=10V*). And *f(t)* is right part (non-homogeneous, equals zero) of differential Eq. (8).

$$\frac{dy}{dt} + a(t)y = f(t), \quad y(0) = Y_0 \qquad a(t) = \frac{1}{500}e^{t/\tau} \quad (8)$$

### 4.2 Analysis of Results

The adaptive filter algorithms were tested under a mixture of random noises and dominated harmonics for f = 1; 2.5; 5 kHz. There are two main zones in the below result curves: The first zone - filter output signal at the beginning of the filter adaptation time $t_{AD}$ = 3–30 days (not very important) which converge towards the desired exponential R(t)- signal

and then continuously controls it until the critical point; The second zone - filter output signal at the end of the observing time – three and more months (up to a year), - till the very important critical point, when signal $R(t) \leq 0.6$ V, which means that the ESP motor cooling oil resistance less than 30 kΩ – it is an extremely risky moment and the ESP must be switched off. It should be reminded here once again that any error in the interpretation of the R(t)-signal critical point may bring a wrong and unreasonably early expensive lift of the ESP for the motor cooling oil removal and renewal service, or in the worst case of the R(t)-signal critical point's late prediction - to short a circuit inside the motor, causing an emergency lift of the ESP for service and a very expensive restoration of the motor.

Unfortunately, the characteristics of some of the adaptive filter algorithms have shown very low frequency ripple oscillation (like Sign-error, Sign-data and Sign-sign filters), which may bring additional errors in the late or early interpretations of the critical point issue. The less amplitude of this oscillation means the better the adaptive filter algorithm. The Normalized LMS FIR filter algorithm adaptfilt.nlms was recognized as the best one.

**Table 1** MATLAB Adaptive Filtering Methods and their results' analyses.

| **Type of Adaptive Filter Methods** | **Frequency (kHz)** | **Stabilization Time (days)** | **Remarks (take into account stabilization time and critical time)** |
|---|---|---|---|
| 1. The Normalized LMS FIR filter algorithm (adaptfilt.nlms) | 5 | 16 | The Best |
| | 2.5 | 28 | |
| | 1 | 32 | |
| 2. The Sign-sign LMS FIR filter algorithm (adaptfilt.ss) | 2.5 | 51 | Good |
| 3. The Sign-error LMS FIR filter algorithm (adaptfilt.se) | 2.5 | 55 | Good |
| 4. The Sign-data LMS FIR filter algorithm (adaptfilt.sd) | 2.5 | 63 | Good |
| 5. The Traditional LMS FIR filter algorithm (adaptfilt.lms) | 2.5 | 78 | Fair |
| 6. The Delayed LMS FIR filter algorithm (adaptfilt.dlms) | 2.5 | 83 | Fair |
| 7. The Adjoint LMS FIR filter algorithm (adaptfilt.adjlms) | 2.5 | No Stabilization | Very Bad |
| 8. The FFT-based Block LMS FIR algorithm (adaptfilt.blmsfft) | 2.5 | No Stabilization | Very Bad |
| 9. The Filtered-x LMS FIR filter algorithm (adaptfilt.filtxlms) | 2.5 | No Stabilization | Very Bad |
| 10. The Block LMS FIR adaptive algorithm (adaptfilt.blms) | 2.5 | No Stabilization | Very Bad |

### 4.2 MATLAB Adaptive Filtering Methods and Their Results' Analyses

The complete analyses and some simulation results are given in Table 1 and Figures 4–12. As it can be seen in Table 1 the best result is accomplished by using the NLMS algorithm. The NLMS algorithm is a more robust method which is independent from the environmental effects. Taking the advantages of the time-varying step size and normalizing the input power the NLMS algorithm outperforms the other LMS methods as depicted in the following figures. The superior behavior of NLMS algorithm is seen in Fig. 4, Fig. 5 and Fig. 6 for different frequencies. The incremental behavior which is useful for digital signal processing and the slower convergence speed of the sign LMS algorithms can easily be seen in Fig. 7 and Fig. 8. The traditional LMS algorithm shows more poorly performance due to its sensibility to the scaling of the input as expected.

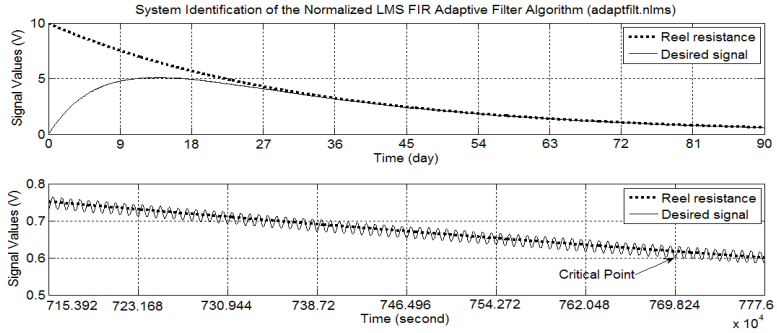



**Figure 4.** The Normalized LMS FIR adaptive filter algorithm adaptfilt.nlms (1kHz) (The Best).

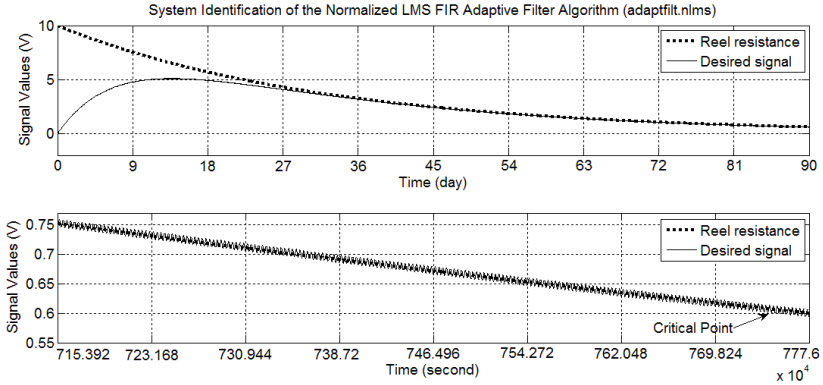



**Figure 5.** The Normalized LMS FIR adaptive filter algorithm adaptfilt.nlms (2.5kHz) (The Best).

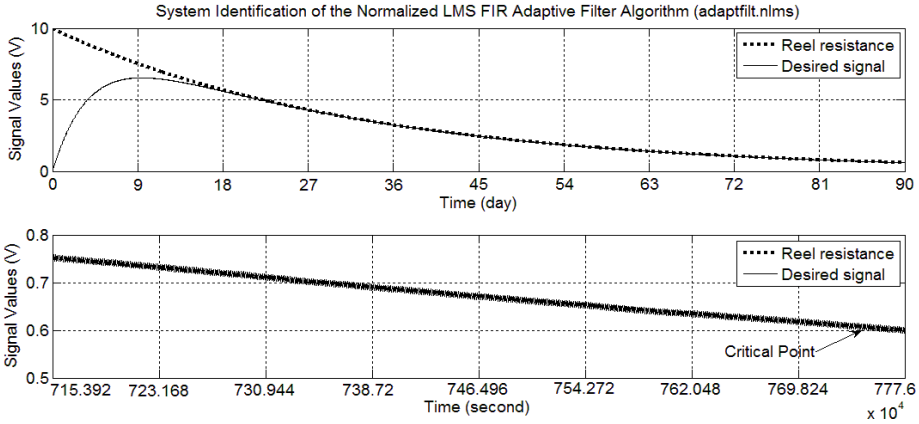



**Figure 6.** The Normalized LMS FIR adaptive filter algorithm adaptfilt.nlms (5kHz) (The Best).

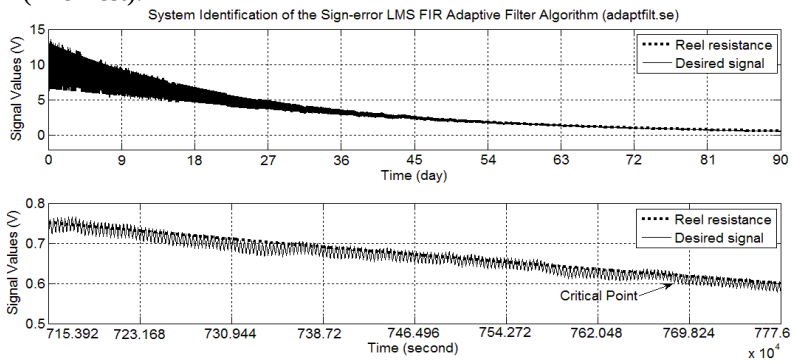



**Figure 7.** The Sign-error LMS FIR adaptive filter algorithm adaptfilt.se (2.5kHz) (Very good).

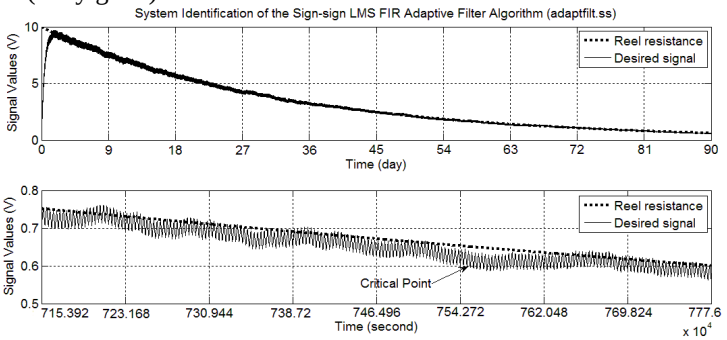



**Figure 8.** The Sign-sign LMS FIR adaptive filter algorithm adaptfilt.ss (2.5kHz) (Very good).

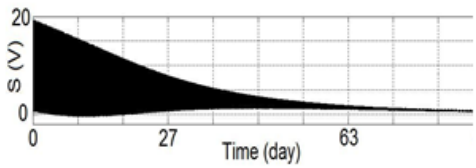



**Figure 9.** .The Traditional LMS FIR adaptive filter algorithm adaptfilt.lms (2.5 kHz) (Fair).

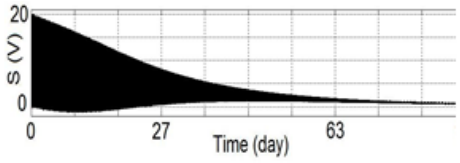



**Figure 10.** The Delayed LMS FIR adaptive filter algorithm adaptfilt.dlms (2.5 kHz) (Fair).

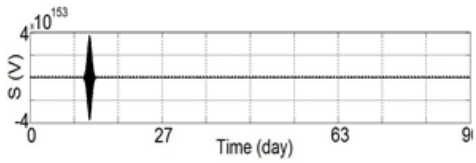



Figure 11. The Adjoint LMS FIR adaptive filter algorithm adaptfilt.adjlms (2.5 kHz) (Very Bad).

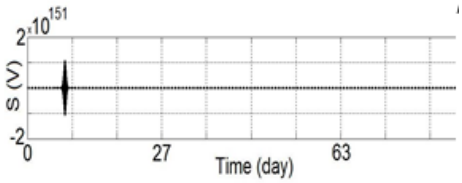



Figure 12. The Block LMS FIR adaptive filter algorithm adaptfilt.blms (2.5 kHz) (Very Bad).

## 5.Conclusion

The Electro Submersible Pump (ESP) electric motor is an extremely expensive part of any ESP system. It is the transformer cooling oil filled motor, separated by strong gasket from pumping liquid that is row petroleum mixed with aggressive salty underground water. But sand and clay which usually presents in the underground liquid very slowly destroys the gasket. This aggressive water may gradually enter into the motor cooling transformer oil. As a result, it can very slowly spoil its electrical insulation resistance from normal – more than 500 kΩ, to risky one less than 30 kΩ. The ESP electric motor is a very sensitive to this regime parameter and to run the ESP under risky conditions is hazardous. Therefore down-hole submersible telemetry system must control its electrical resistance in real-time under the extremely high-level noise.

In this paper, several Adaptive Filter algorithms from MATLAB have been applied to solve in real time the problem of early prediction of disruptions in the oil industry ESP electric motor. From the analysis of the results, it is possible to claim that the start of trouble is predictable within a very long time interval of practical interest. Unfortunately, some of the adaptive filter algorithms have shown bad (4) and fair (2) results, which may bring additional errors in the late or false early interpretation of the critical point of the ESP electric motor insulation disruption issue. Some of the adaptive filter algorithms have shown successful and very good results of the early prediction of the ESP motor real insulation disruption (like Sign-error, Sign-data and Sign-sign filters). The best among the ten analyzed adaptive filter algorithms (methods) for application in ESP telemetry was recognized as - The Normalized LMS FIR filter algorithm - adaptfilt.nlms.

## References

[1] Mohn, F. *Electric Submersible Pumps*. Bergen, Norway: AS Oil & Gas Division, Frank Mohn AS, 1994.

[2] Lyons, W. C., and Plisga, G. J. *Standard handbook of petroleum and natural gas engineering*. Oxford, UK: Gulf Professional Publishing, 2005.

[3] *Smart tools for oil industry*. The leading manufacturer of the oil surface equipment industry in Russia, www.electon.ru.

[4] Bashirov, M.G., Kuznetsov, A.S., Sablin, S.A. "Parameters and characteristics analyses in choosing high-speed automatic transfer switch (HSATS) devices for oil and gas plants power systems," *Oil and Gas Business: Electronic Scientific Journal*. Vol. 4, pp. 11 - 19, 2013.

[5] Chen, C., Zhang B., George, Vachtsevanos, G., and Orchard, M. "Machine Condition Prediction Based on Adaptive Neuro-Fuzzy and High-Order Particle Filtering," *IEEE Transactions on Industrial Electronics,* Vol. 58, No. 9, pp. 4353 – 4364, Sept. 2011.

[6] Stearns, S., and David, R. *Signal Processing Algorithms in Matlab*. NJ, United States: Prentice Hall PTR, 1996.

[7] Sayed, A. H. *Adaptive Filters*, NJ, United States: John Wiley & Sons. Inc., 2008.

[8] *Least Mean Squares Algorithms,* LabVIEW Adaptive Filter Toolkit 1.0 Help, 2008. www.zone.ni.com/reference/en-XX/help/372357A-01/lvaftconcepts/aft_lms_algorithms.

[9] Shynk, J. J. "Frequency-Domain and Multirate Adaptive Filtering," *Proc. IEEE Signal Magazine*, Vol. 9, No. 1, pp. 14-37, 1992.

[10] Hayes, H. *Statistical Digital Signal Processing and Modeling*. New York, USA: Wiley, 1996.

[11] Poularikas, A. D., and Ramadan, Z. M. *Adaptive Filtering Primer with Matlab*. Lahore, Pakistan: International Pvt. Ltd, January 2006.

[12] Ali-Zade P. G., and Kulizade P. K. *Method of electro submersible pump motor insulation checking*. Patent USSR: AC SU 271644, 1968/1970.

[13] IEEE Recommended Practice for Specifying Electric Submersible Pump Cable-Ethylene-Propylene Rubber Insulation. IEEE Std 1018™, *IEEE Industry Applications Society*, New York, NY 10016-5997, USA, 13 April 2005.

[14] Dillard, S. M., and Greiner, T. D. "Transient Voltage Protection for Induction Motors Including Electrical Submersible Pumps," *IEEE Trans. Industry Applications*, Vol. IA-23, No. 2, pp. 365 – 370, 1987.

[15] Boldea, I., and Nasar, S. A. *Electric Drives. Boca Raton*. Florida, USA: CRC Press, 1999.

[16] Bose, B. K. *Modern Power Electronics and AC Drives*. Upper Saddle River NJ, USA: Prentice-Hall, 2002.

[17] Sridhar, B., Sheriff, I. A., Narayanan, K. A., and Kumar, S. S. "Comparison of Cascaded LMS-RLS, LMS and RLS Adaptive Filters in Non-Stationary Environments," *Springer, Part of Springer Science+Business Media, Novel Algorithms and Techniques in Telecommunications and Networking*, 2010, pp. 495-499.

[18] Eweda, E. "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, Vol. 4, No. 11, pp. 2937 – 2944, 1994.

[19] Bhotto, Z. A. "Andreas Antoniou A. A Family of Shrinkage Adaptive-Filtering Algorithms," *IEEE Trans. Signal Processing*, Vol. 61, No. 7, pp. 1689 – 1697, 2013.

[20] Kozacky, W. J., and Ogunfunmi, T. "An active noise control algorithm with gain and power constraints on the adaptive filter," *EURASIP Journal on Advances in Signal Processin*g, Vol. 17, pp. 1 – 12, 2013.

[21] Moravej, Z., Vishwakarma, D. N., and Singh, S. P. "Digital Filtering Algorithms for the Differential Relaying of Power Transformer: An Overview," *Taylor & Francis - Electric Machines & Power Systems*, Vol. 28, No. 6, pp. 485 – 500, 2000.

[22] Moschner, J. L. *Adaptive Filter with Clipped Input Data*. Ph.D, Stanford Univ., Stanford, CA, June 1970.

[23] Gersho, A. "Adaptive Filtering With Binary Reinforcement," *IEEE Trans. Inf. Theory*, Vol. IT-30, pp. 191-199, 1984.

[24] Cowan, C. N., Adams, P., and Grant, P. M. *Adaptive Filters*. Englewood Cliffs NJ, USA: Prentice Hall Signal Processing Series, 1985.

[25] Wan, E. "Adjoint LMS: An Alternative to Filtered-X LMS and Multiple Error LMS," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP);* pp. 1841-1845, Germany, 1997.

[26] Maragliano, G., Marchesoni, M. "Implementation of a Mixed Recursive Least-squares and Kalman Filter Approach in Parameters, Flux, and Speed Estimation for Vector-controlled

Induction Motor Drives," *Taylor & Francis - Electric Power Components and Systems*, Vol. 36, No. 4, pp. 359 – 386, 2008.

[27] Balci, M. E., Karaoglan, A. D. "Optimal Design of C-type Passive Filters Based on Response Surface Methodology for Typical Industrial Power Systems," *Taylor & Francis - Electric Power Components and Systems*, Vol. 41, No. 7, pp. 653 – 668, 2013.

[28] Arezki, M., Namane, A., Benallal, A., Meyrueis, P., and Berkani, D. "Fast Adaptive Filtering Algorithm for Acoustic Noise Cancellation," *Proc. the World Congress on Engineering*, Vol. II, London, 2012.

# Re-engineering of Supply Chain Process: A Literature Review

**Atour TAGHIPOUR**

Faculty of International Business
University of Le Havre
Le Havre, France
atour.taghipour@univ-lehavre.fr

**Abstract.** Supply chains are include independent business. The main question of supply chain management in an independent environment is the coordination of supply chain members' operations with minimum exchange of information. In practice, supply chain operations are generally coordinated and planned hierarchically through the central control unit, which requires a high degree of information exchanges. High degree of information exchanges lead to difficulties when independent members do not want to share information. In order to address these difficulties, decentralized approaches of coordination of operations planning decisions based on some minimal information sharing have been proposed in many academic disciplines. This paper first proposes a literature review of these decentralized approaches.

**Keywords.** Coordination, supply chain, operations planning.

## 1.Introduction

Supply chains are networks of organizations that create and deliver value through upstream and downstream linkages (Christopher 1998, Halal 1994, de Kok and Fransoo 2004). These chains are characterized by distinct, yet mutually interdependent decision domains with independent business objectives (Simchi-Levi et al. 2000). In this context, the lack of decisional, organizational and informational integration leads to inefficiencies related to poor coordination of production and distribution decisions, such as the bullwhip effect (Lee et al. 1997). In order to improve supply chain coordination, companies have developed collaborative practices across different functions of the supply chain (Simatupang and Sridharan 2002). This paper analyses the literature on the distributed coordination of partners in supply chains. A classification scheme of the different methodological tools used in the literature is proposed as guide line for future research opportunities.

The remainder of the paper is organized as follows. In the section 2 we define the problem. Section 3 outlines the research methodology used in this review. Sections 4 is our proposed technological classification of the literature. Finally, section 5 concludes and proposes some research perceptive.

## 2.Definition of problem

Supply chain operations planning as a component of a supply chain management can be define as the coordination of the release of materials and resources in the supply network *(*De Kok and Fransoo, 2004). In the review proposed in this paper, the problem that is specifically investigated focuses on methods that address the coordination of the operations planning function of supply chain partners. This involves making decisions such as what,

how much and when to produce and deliver, and to whom. In other words, this problem consists in synchronizing the supply chain partners' usage of their resources in order to avoid shortage and make sure materials, components and final products flow continuously whenever needed by downstream partners at minimal cost and subject to information asymmetry.

## 3.Methodology of research

In this research, we focus on the supply chain management literature dedicated to the decentralized coordination of operations planning. We first analyze selected contributions of the literature using a systematic analysis based on several dimensions. Next, using this analysis, we propose a classification scheme of these coordination approaches based on their underlying methodological tools.

## 4.Classification scheme for decentralized coordination

This section proposes a classification of different methods of coordination, according to their underlying methodological tools and techniques. This classification scheme includes five main classes of general techniques and different sub-classes of operations planning coordination methods, which include the following:

Exact decomposition and constraint-based techniques- These approaches exploit some form of mathematical decomposition methods to decompose a large supply chain planning problem into several sub-problems that are solved in a distributed manner. In brief, these methods propose mechanisms according to which decisions integration and coordination is carried out with the use of a rigorous mathematical technique that solves to optimality, or near optimality, the coordination problem in a distributed manner. The main characteristics of these approaches are the use of exact local planning process as well as exact coordination process. These techniques include Lagrange decomposition, Bender's decomposition, Dantzig-Wolfe decomposition. A recently proposed method is also added to this group of techniques as it proposes a comprehensive search of the coordination space. It is referred to as distributed search with constraint propagation.

Hierarchical planning and information sharing techniques- Initiated by Hax and Meal (1975), one of the first attempts to address the coordination of operations planning of supply chains involves simplifying complex decision problems into a hierarchy of mutually inter-dependent decision problems. Here, coordination is carried out in a cascade process from long term to short term decisions, or from customer to suppliers. This decomposition leads to simpler and interdependent planning functions. The most significant characteristic of these approaches is the use of greedy/one-way information exchange. Several sub techniques have extended this principle in various ways, which include greedy coordination; information sharing and anticipation model; and partial aggregation of decision domains.

Heuristic search techniques-This literature involve more or less advanced forms of iterative information exchange, during which supply chain partners progressively adjust their local initial plan through local search procedures (i.e., small incremental deviations). Such procedures allow the partners to mutually adjust their plans according to the constraints or capabilities of their partners. This form of coordination techniques requires the design of a convergence mechanism in order to guarantee the improvement and the feasibility of the collective plan, as well as termination conditions in order to stop the incremental process of mutual adjustment. The main characteristic of these approaches is the use of heuristic search

in the coordination process. Because of the heuristic nature of such coordination, these techniques are termed heuristic search techniques. Three sub-techniques have been identified in the literature: distributed heuristic search with local optimization, meta-heuristic search, and interaction based coordination. Some of these coordination heuristics only deal with the coordination of two supply chain partners (Dudek and Stadtler 2005), while others encompass more partners, or even the entire supply chain (Silva et al. 2006). Furthermore, some of these heuristics methodologically embed their coordination mechanism with their local decision processes (Silva et al. 2006), while other heuristics propose coordination mechanisms that can support virtually any local decision process.

Intelligent and adaptive techniques- Another class of supply chain operations planning coordination relies on intelligent software agents that have the ability to plan their own course of activities in order to achieve some goals or maximize their utility. This class is clearly different from the previous sub-class of interaction-based planning coordination because, here, agents are not limited to a heuristic sequence of activities to perform. These agents are called deliberative agents because they have advanced mechanisms to locally determine what activities to perform. Although they can use explicit description of interaction mechanisms, these agents may, or may not, use them according to the situation, their past experience, as well as some representation of their internal state and goals. Some literature are based on such advanced form of agents capable of coordinating their actions with other agents. More specifically, instead of focusing on the interaction as the main mode of coordination, these approaches of coordination focus on how agents should behave in order to contribute to the coordination problem. The coordination is achieved through some form of adaptive heuristic manner.

Bidding-based techniques-Another group of authors (Ahn and Lee 2004, Ito and Salleh 2000, Jiao et al. 2006, Lau et al. 2006, Lee et al. 2008, Lou et al. 2004, Luh et al. 2003, Maturana and Norrie 1997, Neubert et al. 2004, Osman and Demirli 2010, Qinghe et al. 2001, Sun and Wu 2009, Tian et al. 2006, Calosso et al. 2003, Chen et al. 2005, D'Amours et al.1997, Ertogral and Wu 2000, Fan et al. 2003, Kutanoglu and Wu 1999, Lee and Kumara 2007, Zhang and Luo 2009, Zhou and Liu 2007) exploit one form or another of bidding technique. These approaches are rooted in economics and based on so-called negotiations. Supply chain partner selection is the coordination problem addressed by the use of these techniques. These approaches can also be termed contract-net and auctions.

## 5.Conclusion

We analyzed the contributions to the general problem of supply chain operations planning coordination. Based on our analysis of the different techniques and tools used by these contributions, we propose a five-class classification of the coordination and local planning tools of these methods. This analysis highlights several important aspects of the development of such approaches. First, there is an opportunity to develop more dynamic and stochastic mechanisms of planning coordination of supply chains. Second, our analysis highlights a research direction that proposes to integrate exact approaches with other technique such as intelligent agents in order to exploit their complementarities. Finally, intelligent and adaptive techniques, which are a rather new concept in supply chain operations planning coordination. There is also an opportunity to develop collaborative planning standard, on which advanced planning tools could be based in order to guaranty both the complementarity of these tools as well as the freedom of companies to adopt any planning tools.

## References


1. Ahn, H.J., and Lee, H., 2004. An agent-based dynamic information network for supply chain management. *BT Tech. J.*, 22(2), 18-27.
2. Calosso, T., Cantamessa, M., Vu, D., and Villa, A., 2003. Production planning and order acceptance in business to business electronic commerce. *Int. J. Prod. Econ.*, 85(2), 233-249.
3. Chen, R.R., Roundy, R.O., Zhang, R.Q., and Janakiraman, G., 2005. Efficient auction mechanisms for supply chain procurement. *Management Science*, 51(3), 467-482.
4. Christopher, M., 1998. Logistics and supply chain management: Strategies for reducing cst and improving service. *Second ed., London*.
5. D'Amours, S., Montreuil, B., and Soumis, F., 1997. Price-based planning and scheduling of multiproduct orders in symbiotic manufacturing networks. *Eur. J. Op. Res.*, 96(1), 148-166.
6. Dudek, G., and Stadtler, H., 2005. Negotiations-based collaborative planning between supply chain partners. *Eur. J. Oper. Res.*, 163, 668-687.
7. de Kok, T., and Fransoo, J.C., 2004. Planning supply chain operations: Definition and comparison of planning concepts. *Handbooks in Operations Research and Management Science*, 11.
8. Ertogral K., and Wu, D.S., 2000. Auction-theoretic coordination of production planning in the supply chain. *IIE Trans.*, 32(10), 931-940.
9. Fan, M., Stallaert, J., and Whinston, A.B., 2003. Decentralized mechanism design for supply chain organizations using an auction market. *Information Sysyems Research*, 14(1), 1-22.
10. Halal, W.E., 1994. From hierarchy to enterprise: Internal markets are the new foundation of management. *Acad. Mngt. Exec.*, 8(4), 69-82.
11. Hax, W.E., and Meal, H.C., 1975. Hierarchical integration of production planning and scheduling. *M. A. Geisler, (ed.). Amsterdam : North-Holland*.
12. Ito, T., and Salleh, M.R., 2000. Blackboard-based negotiation for collaborative supply chain system. *J. Mat. Proc. Tech.*, 107(1-3), 398-403.
13. Jiao, J., You, X., and Kumar, A., 2006. An agent-based framework for collaborative negotiation in the global manufacturing supply chain network. *Rob. Comp.-Int. Manuf.*, 22(3), 239-255.
14. Kutanoglu, E., and Wu, S.D., 1999. On combinatorial auction and lagrangean relaxation for distributed resource scheduling. *IIE Trans.*, 31(9), 813-826.
15. Lau, J.S.K., Huang, G.Q., Mak, H.K.L., and Liang, L., 2006. Agent-based modeling of supply chains for distributed scheduling. *IEEE Trans. Sys. Man Cyb.*, 36(5), 847-861.
16. Lee, J., Lee, S.J., and Chen, H.M., 2008. Dynamic role binding with agent-centric contract net protocol in agent organizations. *IEEE*, 636-644.
17. Lee, S., and Kumara, S., 2007. Decentralized supply chain coordination through auction markets: Dynamic lot-sizing in distribution networks. *Int. J. Prod. Res.*, 45(20), 4715-4733.
18. Lee, H. L., Padmanabhan, V., and Whang, S., 1997. Information distortion in a supply Chain: The bullwhip effect. *Man. Sc.*, 43(4), 546-558.
19. Lou, P., Zhou, Z.-D., Chen, Y.-P., and Ai, W., 2004. Study on multi-agent-based agile supply chain management. *Int. J. Adv. Manuf. Tech.*, 23(3-4), 197-203.

20. Luh, P.B., Ming, N., Chen, H., and Thakur, L.S., 2003. Price-based approach for activity coordination in a supply network. *IEEE Trans. Rob. Aut.*, 19(2), 335-346.

21. Maturana, F.P., and Norrie, D.H., 1997. Distributed decision-making using the contract net within a mediator architecture. *Decision Support Systems*, 20, 53-64.

22. Neubert, R., Görlitz, O., and Teich, T., 2004. Automated negotiations of supply contracts for flexible production networks. *Int. J. of Prod. Eco.*, 89(2), 175-187.

23. Osman, H., and Demirli, K., 2010. A bilinear goal programming model and a modified Benders decomposition algorithm for supply chain reconfiguration and supplier selection. *Int. J. Production Economics*, 124, 97-105.

24. Qinghe, H., Kumar, A., and Shuang, Z., 2001. A bidding decision model in multiagent supply chain planning. *International Journal of Production Research*, 39(15), 3291-3301.

25. Silva, C.A., Sousa, J.M.C., Runkler, T.A., and Da Costa, J.M.G.S.A., 2006. Distributed optimisation of a logistic system and its suppliers using ant colonies. *Int. J. Sys. Sc.*, 37(8), 503-512.

26. Simatupang, T. M. and R. Sridharan (2002). The collaborative supply chain. *Int. J. of Log. Man.* 13(1): 15-30.

27. Simchi-Levi, D., Kaminski, P., and Simchi-Levi, E., 2000. Designing and Managing the Supply Chain: Concepts, strategies, and cases. *USA, McGraw-Hill*.

28. Sun, D., and Wu, J., 2009. Multi-agent coordination based on contract net protocol. *International Symposium on Intelligent Ubiquitous Computing and Education, IUCE*, 353-357.

29. Tian, J., Foley, R., Yao, X., and Tianfield, H., 2006. An extended contract net mechanism for dynamic supply chain formation and its application in China petroleum supply chain management. *Multiagent and Grid Systems- An International Journal*, 2, 183-207.

30. Zhang, Q., and Luo, J., 2009. Study on the supply chain coordination of perishable items based on double auction. *Industrial Engineering and Management*, 14(3), 33-44.

31. Zhou, Y.H., and Liu, K.J., 2007. Supply chain coordination under e-procurement auction. *14th International Conference on management Science & Engineering, China*.

# Building and Exploring Web-based semantics and Ontology

**K. Musumbu**

LaBRI (UMR 5800 du CNRS),
351, cours de la Lib´eration, F-33.405 TALENCE Cedex
e-mail: musumbu@labri.fr

**Abstract.** Ontology gives system architects the ability to describe the high-level structure of their software in the most direct and expressive manner. The Web has changed the way we share knowledge, the way we design distributed services and applications, the way we access large volumes of data, and the way we position ourselves with our peers. Successful exploitation of web-based concepts by web communities lies on the integration of traditional data management techniques and semantic information into web-based frameworks and systems. In this paper, we continue focusing on identifying challenges for building web-based useful services and applications, and for effectively extracting and integrating knowledge from the Web.

**Keywords:** Knowledge based systems, Model Driven Architecture, reasoning, Semantic Web, ontology.

## 1. Introduction

Searching in web is a topic which gains importance for the daily life of everybody in today's information society. Web users have experienced the hype of Web 2.0, acquired knowledge about new innovative user interfaces like Flicker, Del.ico.us, Google Maps and YouTube and become User 2.0. It is interesting to analyze the conceptual basis of the Semantic Web from a methodological point of view. It is important to keep in mind that we are here talking about data that are meaningful not only for humans but in a first instance for computers. Human beings do not need any markup metadata tags for understanding current Web pages. For them, the existing Web is already the semantic one. But for machines it is meaningless, and nonsemantic. Consequently, a question arises about the point at which the Webbecomes meaningful for computers, in other words, becomes semantic and to what extent it can be possible. The Semantic Web promotes the vision of an extended Web of machine-understandable information and automated services that allows knowledge technologies to reach Web-scale. The explicit representation of the semantics of the data and of the services will enable a new Web that provides a qualitatively new level of service. Automated services will improve in their capacity to assist humans in achieving their goals by "understanding" more of the content on the Web, and thus providing accurate filtering, categorization, and searches of information sources. The idea of the "samess" of Web data will provide a new solution to many current problems, such as more meaningful searches on Web. The use of RDF, OWL, or other high-level metadata technologies can make Web search even more powerful and computers will be able to automativally convert complex expressions from one domain-specific into another in order to process them.

In this paper we propose an ontology matching paradigm based on the idea of harvesting the SemanticWeb, i.e., automatically finding and exploring multiple and heterogeneous online knowledge sources to derive mappings. We adopt an experimental approach in the context of matching two real life, large-scale ontologies to investigate the potential of this paradigm, its limitations, and its relation to other techniques

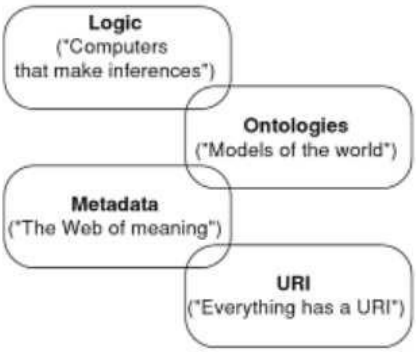



**Figure 1**: Block semantics

This paper will be divided in three parts. The first covers the basis of both main topics ontology, semantic web and standards. The second part, explain the Model Driven Architecture. The last is the central one, it starts with a review of several approaches and aim to bridge the gap between ontology development and software engineering methodologies. We will also discuss the possibilities and the benefits provided by mixing models and web reasoning.

## 2. Explore vs. Analyze

Another tension in user interface design is that between exploration and analysis. The WWW is something we explore, following links on a hunch that they will take us somewhere useful, reassessing the situation after each link. Exploring the semantic web in this way involves moving from node to node, finding more arcs, and reassessing our next move at each step. Most data applications, such as money management or calendar management, however, allow us to sort, process, and visualize data that has a very well-defined structure, such a set of bank transactions each with a date, payee, and amount. To deal with these needs the Tabulator operates in two interlocking modes – exploration and analysis. The user can start using the Tabulator by submitting a URI, or by adding ”?uri=URI” in the location bar of the browser. In the exploration mode, the user is unaware of what data is available, and can explore an RDF graph in a tree view, expanding nodes to get more information about them. As the user does this, the Tabulator implicitly follows links that may contain more RDF data about the relevant nodes. To move to the analysis mode, the user can select certain fields (arcs or predicates) to define a pattern, and ask the Tabulator to find all examples of that pattern. The Tabulator then performs this query, following links as it attempts to match the query pattern to the RDF graph. The results of a query can be displayed in a number of modular views, including tables, and projections of time and space axes onto a calendar and map. These allow the dense presentation of data that one expects from domain-specific applications. One can start a new exploration from any

table cell, or point in a calendar or map, and switch back to the exploration mode. If the subgraph pattern has temporal or spatial coordinates, these tables can be projected onto the calendar/timeline view or map view respectively. The user can double-click on any instance of a dense display of the selected cases to open a new exploration at that point. Exploration and analysis therefore can be interleaved as the user tackles the task at hand.

## 3. Exploring in outliner mode

Many RDF visualizers such as IsaViz and RDF Author represent data as circle- and-arrow diagrams. This gives one a feel for how a small number of things connect. It allows one to see clustering when a large number of things are related by the same properties, such as in Foafnaut or How/Why diagrams developed at W3C. Circles and arrows are very intuitive and useful when trying to understand the structure of data. However, it is not an appropriate way to look at data with many nodes and many different properties. It is not used in applications we think of as handling data, such as personal financial management, music management, calendar management programs, for example. In these cases tables or matrices are used. These are the densest way of comparing objects of the same class (strictly, which are likely to share common properties). mSpace is an example of a table based semantic web browser. These table-based systems, though, tend to operate on a restricted set of data, and do not naturally allow browsing outside it. The outliner mode of the Tabulator is an extremely common user interface metaphor and clearly natural for tree-oriented data. People are typically very comfortable in a tree-like environment. In fact, much data in the world has been organized into trees, and the web is largely composed of overlapping trees with local roots all over the place. This suggests that in fact a tree-oriented browser will feel natural even though the world is actually a web. In early hypertext, Peter Brown's Guide system was an outliner, and the Gopher system was a presented as tree (not in outline mode) though in fact it was a web. TreePlus, like outline mode, maps a graph to a tree, but it appears designed for graphs of homogeneous type, whereas Tabulator's outliner mode targets a heterogenous typed graph. The outliner view uses a denser tabular format designed to meet user demands for more information to be visible at a time. Another related approach is Ghoniem's work, which deals with visualizing large complex graphs with single arc types but their matrix form is different from the table view used in Tabulator. The outliner browser of the Tabulator is quite straightforward. Most of the design decisions involve how to represent options to load data, and whether to load it automatically - both of which have many variations.

## 4. The Model Driven Architecture (MDA)

The Model-Driven Architecture starts with the well-known and long established idea of separating the specification of the operation of a system from the details of the way that system uses the capabilities of its platform. The Figure 2 gives a general view of the MDA approach. We can see that a construction of a new Information System begins with the development of one or more requirements models (CIM). Then we may develop models independent from any platform (PIM). In theories, the latter models must be partially generated from the former. Platform independent models must be permanent, i.e. they do not contain any information about execution platform (is it a J2EE or .NET etc. application). For constructing the concrete application, we must have platform specific models (PSM). These models are obtained by transforming PIM and adding technical information relative to

platforms. PSM are not permanent models. All these models are for facilitating code generation. The MDA approach is widely used and advanced generators exist.

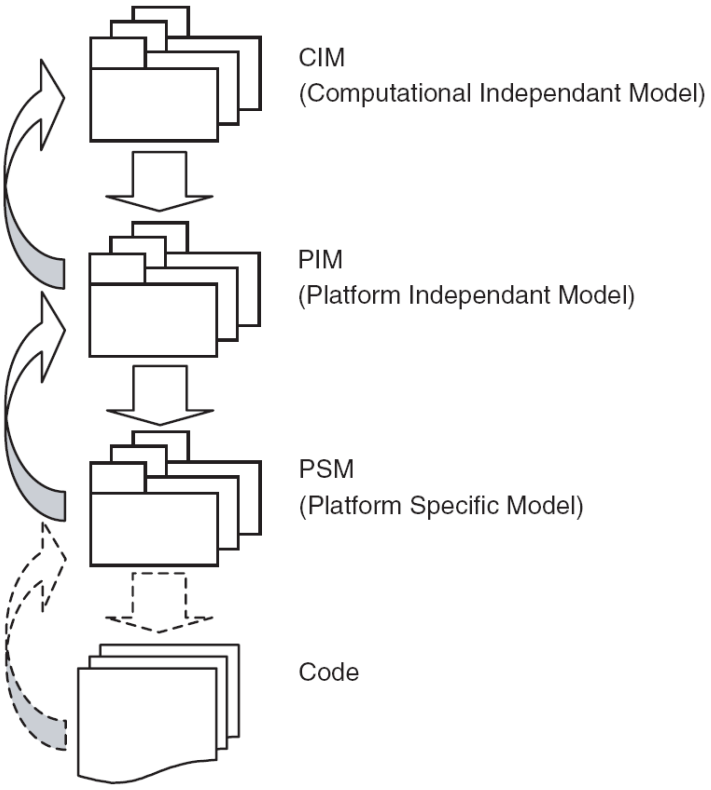



**Figure 2**: Global view of the Model Driven Architecture approach

**4.1 MDA models and semantics**

MDA principals are very interesting and allow economizing many times during application life cycle by code and model generation. However, MDA specification does not tell anything about semantics on models. MDA is only interested by content and not context then that using semantics will offers more interesting way in automatic generation.

**4.1.1 Why should MDA talk about semantics:**

Making transformations between CIM and PIM, between PIM and PSM, and between PSM and code are done by specifying transformations rules. Nowadays these rules are manuals and machines cannot generate it because there is no notion of semantics between the entities that are concerned by transformations.

Business rules are about meanings and act on models. Generating all business rules is impossible but it would be possible to generate a large set of generics among them. For example for the model in Figure 3, we should want to generate business rules like:

- Every *Human* must have a father and a mother.
- IF a *Human* is the mother of a *Human* then this *Human* is a *Woman*.
- IF a *Human* is the father of a *Human* then this *Human* is a *Man*.

For doing this automatically, it is clear that adding semantics in models is needed.

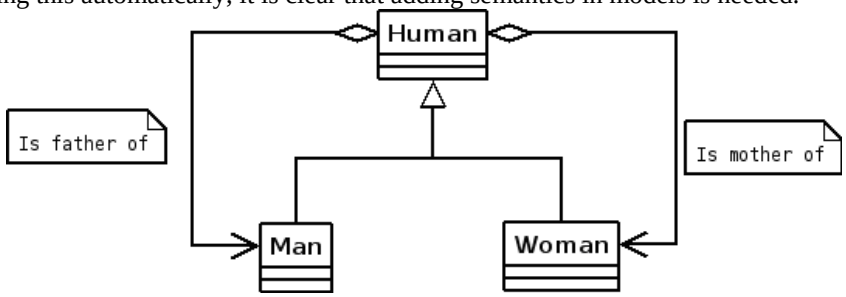



**Figure 3**: A little human model

**4.1.2 Eventual solutions for adding semantics in models:**

In MDA, an instance of MOF is use for representing models but our works are only concerned by UML models. For adding semantics in UML models we can use:

- *UML profile*: UML can be used for modeling many domains. The problem with this is that UML models are so generic that it is impossible to know either it is object application, a meta-model, a model, a database structures or anything else only by looking at it. For adding precision, the OMG has standardized the concept of UML profile. A UML profile is a set of techniques and mechanisms allowing to adapt UML in a particular and specific domain. UML profile can be used in any UML model and do not modify the structure of the meta-model. UML profiles are stereotypes or labels which can be injected into models. After having stuck labels on models, we can make inference using it. As we can see, doing this can solve our problem of semantics lack on model in a low level, but this is not exploitable by machines because there is no notion of logic and taxonomy and semantic is not formally defined.
- *Object Constraint Language*: In UML it was not possible to define the body of an operation (or a method) so the OCL was standardized by OMG for doing it. OCL allows expressing any kind of constraints on UML models. For example, we can express constraints like: "before renting a car you must be sure that it is OK". Well, OCL seems to be a good solution for our problem but it is not the case. The first problem with OCL is that he does not support side effect operations and the second is that he does not offer automatic inference for machines.
- *Action Semantic*: remember that the main constraint with OCL was that he only supports no side effects operations. To solve this constraint, the OMG standardize Action Semantic. Well, now we have a formalism being able to express any king of operations and constraints but it is not enough. This formalism is complicated to use, was not created while thinking to machine comprehension and self-use, and do not have a textual formalism.
- As we can see, none of the UML "techniques" is suitable for adding semantics in models. In other side a new domain of computer is growing more and more: semantic web. The aim of the semantic web is to make the web both comprehensible by humans and machines. A part of semantic web is about ontology and reasoning. Modeling concept defined by ontologies can be used to model the concepts in a domain, the relationships between them, and the properties that can be used to describe instances of those concepts. In addition, the Web Ontology Language

(OWL)[24] supports the inclusion of certain types of constraint in ontology, allowing new information to be deduced when combining instance data with these logic's description. At this point our dilemma was how can we use MDA models and Semantic Web? Ontology Definition Meta Model was the response to our need.

## 4.2 The Ontology Definition Meta-model

The MDA and its four-layer architecture provide a solid basis for defining the Meridel's of any modeling language, and thus a language for modeling ontology based on the MOF. The ODM is a proposal for an OMG's RFP (Request For Proposal) resulting of an extensive previous research in the fields of the MDA and ontology. The main objective of the ODM is to bridge the gap between traditional software tools for modeling (like UML) and artificial intelligence techniques (Logic's description) for making ontology. The principle of ODM is to merge two big domains of research which are Model Driven Architecture and Semantic Web. ODM is still in standardization process at the OMG when this paper is being written. Basically the ODM allows making ontology using UML (by using an UML refile with existing tools like Rational Rose or Poseidon) and transforming it to OWL/RDF, Topic Map or Common Logic (Figure 4).

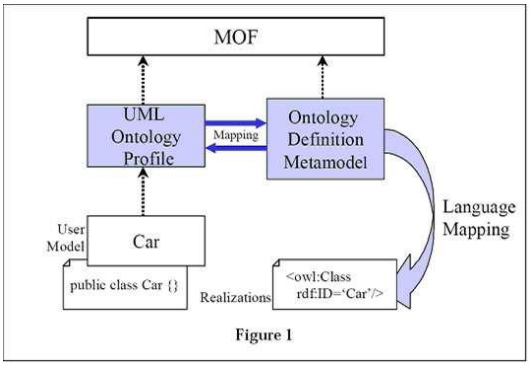



**Figure 4**: ODM principle

## 5. Ontology reasoning

Ontology is an area of great importance for the semantic web. An ontology establishes the things that a system can talk about and makes reasoning. Describing concepts and relationships between them formally, offers to machines the ability of making some varieties of logic, formally or not. Ontology supplies the concepts and terms; logic provides ways to make statements that define and use them, and to reason about collections of statements that use the concepts and terms. In the semantic web, logic plays many different roles:

1. Firing rules: having a set of facts, take a decision.
2. Making inference on facts: for example if we know that Adam's wife is Eve, we can infer that Eve is a woman.
3. Explaining why a particular decision has been reached.
4. Detecting contradictory statements and claims.

OWL exploits results of more than 15 years of Description Logics (DL) research. Indeed, for OWL a semantic was defined such that very large fragments of the language can

be directly expressed using so called Description Logics. Description Logics is a family of logic based Knowledge Representation formalisms descendants of semantic networks. It describes domains in terms of concepts (classes), roles (properties, relationships) and individuals. In description logics terminology, a tuple of a T-box and an A-box is referred to as a knowledge base. An individual is a specific named object. With some restrictions, one can state that the logical basis of OWL can be characterized with the description logics SHIQ(Dn)−. This means, with some restrictions, OWL documents can be automatically translated to SHIQ(Dn)− T-boxes. The RDF-Part of OWL documents can be translated to SHIQ(Dn)− A-boxes.

The logic SHIQ(Dn)− is interesting for practical applications because highly optimized inference systems are available (e.g., Racer). In such systems, the following reasoning can be made with T-box:

1. Concept consistency: Is the set of objects described by a concept empty?
2. Concept subsumption: Is there a subset relationship between the set of objects described by two concepts?
3. Find all incoherences between the concepts mentioned in a T-box. Inconsistent concepts might be the result of modeling errors.
4. Determine the parents and children of a concept: The parents of a concept are the most specific concept names mentioned in a T-box which subsume the concept. The children of a concept are the most general concept names mentioned in a T-box that the concept subsumes.

With A-box we can answer the following questions :

1. Check the consistency of an A-box: Are the restrictions given in an A-box w.r.t. a T-box too strong, i.e. do they contradict each other? Other queries are only possible w.r.t. consistent A-boxes.
2. Instance testing: Is an individual instance of a concept? The individual is then called an instance of the query concept.
3. Instance retrieval: Find all individuals from an A-box such that the objects they stand for can be proven to be members of a set of objects described by a certain query concept.
4. Computation of the direct types of an individual: Find the most specific concept names from a T-box of which a given individual is an instance.
5. Computation of the roles which make reference to an individual.

Given the background of Description Logic, these inference services can be used to solve actual problems with OWL knowledge bases.

In next section, we will see how ontology reasoning can be used to solve the lack of semantics in models.

### 5.1 Our Approach for business rules automatic generation

Our principle is to use the advanced researches in Semantic Web, to combine it with Model Driven Architecture in order to make automatic business rules generation.

For generating business rules automatically, we will use principally the semantics in OWL format.

In OWL reasoning, we can make automatic reasoning both with structures (TBox) and assertions on individuals and properties (ABox). In our case for example, if we have:

*Predicate : Domain1 → Domain2*

This declaration means that we have a property *Predicate* going from the domain *Domain1* to the range *Domain2*.

So we want to generate that:

*IF*
*Object1 Predicate Object2*
*THEN*
*Object1 is of type Domain1*
*AND*
*Object2 is of type Domain2*

Reasonings are done using both domain and range restrictions, individuals and also properties's characteristics (functional, transitivity, symmetric, inverse, etc.)

The Figure 5 describes our approach: using ODM, our model is generated in OWL/RDF model and this last one is enriched with semantics. With this semantically rich model two solutions are possible for generating rules: serialize the rich model in XMI and use e.g JMI for parsing it manually. Another solution is making inference directly with the OWL model using an OWL reasoner. We have adopted the last solution because there exist good OWL Reasoners and this solution uses less intermediary steps.

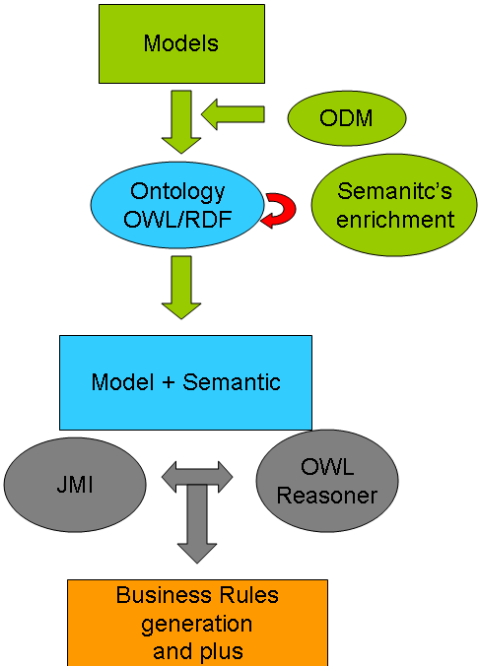



**Figure 5**: Our approach

Recall that on goal is not to generate all kinds of business rules. Indeed, this is infeasible. However, the part of them that able to generate will save time for business experts. Figure 6 summarizes our approach throughout MDA layers. As we can see the first step will be a generation according to the Computation Independent Model (CIM) in an

OMG SBVR like syntax (in strict natural language), the next step will be to generate executive rules according to the Platform Independent Model using our rule language and models based on XMI like standard. At the PIM level either our business rules language ERML, the RIF W3C standard, the PRR OMG proposal or RuleML may be used. At this step we use our "translators" for generating rules at the PSM level for a specific rule engine. If in the future, a standard business rules language is adopted, we'll either make a "translator" from our language towards the new standard or either store directly rules in the new formalism.

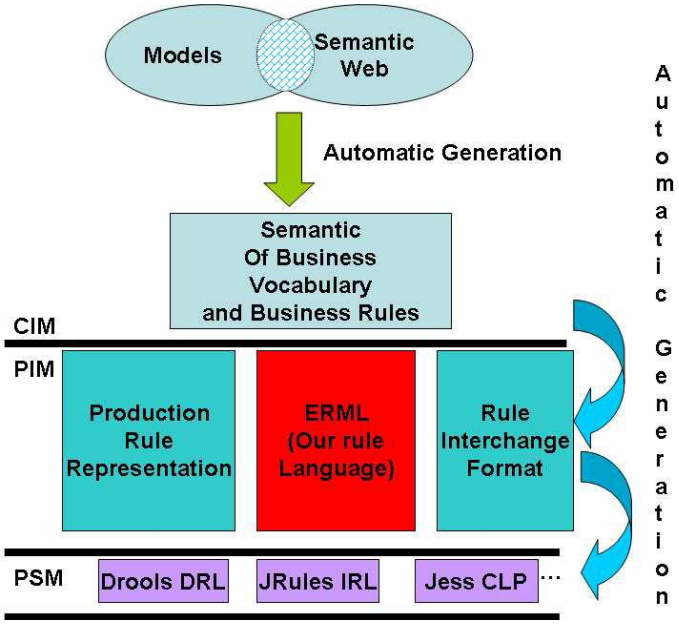



**Figure 6**: Our approach throughout the MDA layers

## 5.2 Implementation

Our approach is being implemented in a system of E-Government web application's generator from specifications. Figure 7 describes its architecture. A studio is used for allowing business experts to specify the future web application. During specification, business experts can use an integrated rule editor in natural language for writing business rules. First, we have worked on creating a rule language independently of any rule engine after having studied the previous attempts of rule's standardization. From the beginning we have chosen to implement our own rule language together with translators from this language to some other ones. Few attempts for standardizing rule languages were proposed in the past without success. CommonRules and RuleML were the most promising.

It's important to note that our goal was not to add a new rule language to the standardization process but, as we work in an industrial project, it was simpler to develop our own rule language than using one which may be accepted or not in this standardization process which is only in its beginning phase. ERML is also used in another project for automating ergonomic rules guideline inspection for web sites. The process of standardizing business rules formalism is in serious progress with OMG and W3C workgroups.

After this step of getting business rule formalism the next step was to implement our approach on business rules automatic generation. Figure ?? shows a first prototype of our approach in an eclipse environment. Our model is an ECore model, our semantics is in OWL. For us, the Ontology Definition Metamodel (ODM) arrives just in time, while we were thinking about how to use directly our semantics in OWL with MOF models (here ECore). We use an eclipse implementation of ODM (EODM). The process is the following: use EODM for transforming our ECore model to OWL model. This OWL model is a simple ontology (taxonomy) model. After this, using the UML profile of ODM, we enrich the OWL model with Abox and Tbox assertions, this is done using any UML tool (in our case MagicDraw) supporting profiles. The next step was to use an OWL reasoner like Racer or EODM reasoner for making inferences according to the domains, ranges and properties for generating business rules. At this state of the implementation, business rules are generated in a Semantics of Business Vocabulary and Business Rules (SBVR) like format.

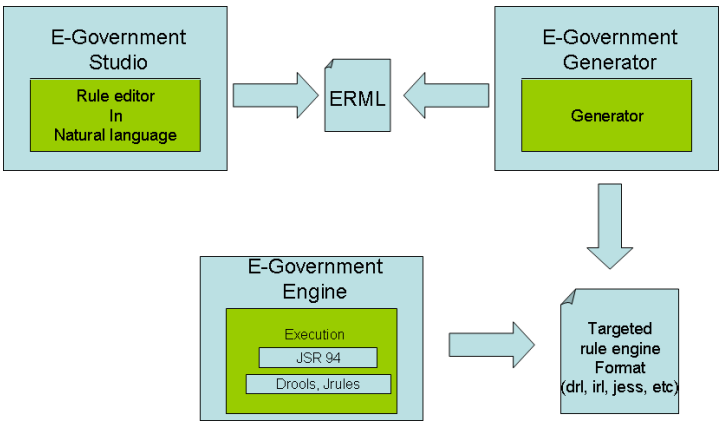



**Figure 7**: Architecture of our E-Government web application

## 6. Conclusions

A business rules application is intentionally built to accommodate continuous changes in business rules. The ability to change them effectively is fundamental for improving business adaptability. The platform on which the application runs should support such continuous changes. Offering to knowledgeable business people (experts) the possibilities to formulate, validate, and manage rules in a "zero-development" environment brings more value-added to this notion of "computer sciences in humanity's service". Allowing an automatic generation part of this business rules will be of valuable help. In this paper, we have seen that, by combining the two domains, Model Driven Architecture and Semantic Web, a solution is possible.

Right now we can only make generation according to the Computation Independent Model (CIM) in a OMG SBVR like syntax (in natural language). Due to the fact that the standardization of SBVR is recent, no implementation does exist. The best to our knowledge, the only free implementation we know is SBeaVER which is not at an advanced level and, at this moment, can only express Business vocabulary and not Business Rules. The next step will be to generate executive rules according to the Platform Independent Model using our

rule language and models based on XMI like format. The last step will be to have an editor allowing to edit both models and semantics.

Making simple generic business rules generation possible from models facilitates the use of the business rules approach which allows easier systems maintenance. It's clear that generating all kinds of business rules is an utopia and we must delimit the degree of generation we want to obtain. Adding semantics to conceptual models open exciting and interesting domains of applications like information merge.

The merits of our solution were twofold: the rooting in the tradition of Software Engineering and the use of sophisticated, up-to-date Web Engineering technology. We have given the maximum attention to "the method", consisting of gathering requirements, producing formal specifications, and then adapting them to the new context of the Semantic Web.

## References


1. Alani, H.: *Ontology Construction from Online Ontologies. In: Proc. of WWW (2006)*
2. Aleksovski, Z., Klein, M., ten Katen, W., van Harmelen, F.: *Matching Unstructured Vocabularies using a Background Ontology. In: Staab, S., Sv´atek, V. (eds.) EKAW 2006. LNCS, vol. 4248. Springer, Heidelberg (2006) CrossRef*
3. Arens, Y., Chee, C.Y., Hsu, C.N., Knoblock, C.A.: *Retrieving and Integrating Data From Multiple Information Sources. International Journal of Cooperative Information Systems 2(2), 127–158 (1993) CrossRef*
4. Berners-Lee, T., Hendler, J., Lassila, O.: *The Semantic Web. Scientific American 284(5), 34–43 (2001) CrossRef*
5. Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., Stuckenschmidt, H.: *Contextualizing ontologies. Journal of Web Semantics 1(4), 24 (2004)*
6. Bouquet, P., Serafini, L., Zanobini, S.: *Peer-to-Peer Semantic Coordination. Journal of Web Semantics 2(1) (2005)*
7. Castano, S., Ferrara, A., Messa, G.: *Results of the HMatch Ontology Matchmaker in OAEI 2006. In: Shvaiko, et al.*
8. Cimiano, P., Handschuh, S., Staab, S.: *Towards the Self-Annotating Web. In: Proc. Of WWW (2004)*
9. Collet, C., Huhns, M.N., Shen, W.: *Resource Integration Using a Large Knowledge Base in Carnot. IEEE Computer 24(12), 55–62 (1991)*
10. [d'Aquin, M., Sabou, M., Dzbor, M., Baldassarre, C., Gridinoc, L., Angeletou, S., Motta, E.: *A Gateway for the Semantic Web. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519. Springer, Heidelberg (2007)*
11. Ding, L., Pan, R., Finin, T., Joshi, A., Peng, Y., Kolari, P.: *Finding and Ranking Knowledge on the Semantic Web. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729. Springer, Heidelberg (2005)*
12. Dzbor, M., Domingue, J., Motta, E.: *Magpie - towards a semantic web browser. In: Fensel, D., Sycara, K.P., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870. Springer, Heidelberg (2003)*
13. Ehrig, M., Sure, Y.: *FOAM - Framework for Ontology Alignment and Mapping; Results of the Ontology Alignment Initiative. In: Proc. of the Workshop on Integrating Ontologies (2005)*
14. Giunchiglia, F., Shvaiko, P., Yatskevich, M.: *Semantic Schema Matching. In: Meersman, R., Tari, Z. (eds.) OTM 2005. LNCS, vol. 3760, pp. 347–365. Springer, Heidelberg (2005)*
15. Giunchiglia, F., Shvaiko, P., Yatskevich, M.: *Discovering Missing Background Knowledge in Ontology Matching. In: Proc. of ECAI (2006)*
16. Haase, P., Qi, G.: *An analysis of approaches to resolving inconsistencies in DL-based ontologies. In: Proc. of Int. WS on Ontology Dynamics, IWOD (2007)*
17. Kalfoglou, Y., Schorlemmer, M.: *Ontology Mapping: The State of the Art. The Knowledge Engineering Review 18(1), 1–31 (2003)*

*18.* Magnini, B., Serafini, L., Speranza, M.: *Making explicit the semantics hiden in schema models. In: Proc. of the Human Language Technology for the Semantic Web WS (2003)*

*19.* Mika, P.: Flink: *Semantic Web Technology for the Extraction and Analysis of Social Networks. Journal of Web Semantics 3(2) (2005)*

20. Schlobach, S., Huang, Z., Cornet, R., van Harmelen, F.: *Debugging Incoherent Terminologies. Journal of Automated Reasoning (2007)*

*21.* Stephens, L.M., Gangam, A.K., Huhns, M.N.: *Constructing Consensus Ontologies for theSemantic Web: A Conceptual Approach. World Wide Web Journal 7(4), 421–442 (2004)CrossRef*

*22.* Stuckenschmidt, H., Serafini, L., Wache, H.: *Reasoning about Ontology Mappings. In: Proc. of the ECAI WS on Context Representation and Reasoning (2006)*

23. Volker, J., Vrandecic, D., Sure, Y.: *Automatic Evaluation of Ontologies (AEON). In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729. Springer, Heidelberg (2005)*

# What's Up with WhatsApp? Connecting Classmates, Clubbers and Colleagues

**Saraswathy Thurairaj[1], Swagata Sinha Roy[2]**

[1] Universiti Tunku Abdul Rahman
tsaraswathy@utar.edu.my

[2] Universiti Tunku Abdul Rahman
swagata@utar.edu.my

Faculty of Creative Industries
Lot PT 21144, Jln Sungai Long Bandar Sg Long, Cheras 43000 Kajang Selangor

**Abstract.** The long and rather far reaching strides in technology have changed the means and modes of communication world-wide. A mobile messaging application WhatsApp, which is available for Android phones, the iPhone, Blackberry allows the exchange of messages via SMS for free. This has resulted in users creating groups as the group members can be literally 'unlimited.' Even more amazing is that images and audio messages can be sent and forwarded at length. In Malaysia, this has caught on as an alternative to Face book especially since it is less 'public' and it can be done as a click of the fingers on your mobile device. Some popular WhatsApp groups include work-place /department groups, lecturer-student groups, book club circles, similar interest circles, old school/institution/coursemate/classmate groups. Some of the participants are as young as 11years and as old as 75. This trend is predicted to stay for a long time. This paper presents a conceptual study of this communication trend by looking at 200 respondents and what they have to say about the pros and cons of being a part of a WhatsApp group.

**Keywords:** WhatsApp, Communication Trend, SMS.

## 1.Introduction

### What's H(APP)ening?

In Malaysia, having a mobile phone has become one of the top necessities in product ownership. Almost every man, woman and child (of appropriate age) owns a cellular device. With the launch of the WhatsApp application, everyone owning a smart phone has it installed, the reason being it is convenient, easy to use and incurs no cost. WhatsApp has become the new mode of communication used among work mates, classmates, similar interests groups and so on. In Malaysia, school going children as young as in the third standard, have a WhatsApp account. Some alumni have jumped on the bandwagon and now there is a never seen before proliferation of group accounts. What is truly interesting is we have users who range from nine to 75 years in age. Some work places (units and departments) have made it compulsory to join a WhatsApp group basically to transmit information (relevant and urgent) to all concerned at one go.

### Creating WhatsApp Groups

WhatsApp groups are being created at a rate unimaginable at any particular given times and in any given place. As long as the user has a smart device with wireless or data access, he or she can download WhatsApp from Play Store and Moborobo in android and iPhone devices. It is fairly easy to connect to others who have a WhatsApp account as the WhatsApp icon will automatically appear in the contact profiles once the contact number is typed in. Group accounts are also easy to start. One or a few in the group can volunteer to be the administrator thereby giving the group a decided-upon name and profile picture and keying in the numbers and names or transferring them from the contact list. You can add and delete contacts accordingly. In fact, instructions are spelt out clearly so anyone can follow them. The group profile picture can be changed at random by the administrators; in many instances, the profiles follow a theme, like Happy New Year or XYZ's wedding or Graduation or whatever is the current fad. For clubbers icons, there can be pictures of particular activities such as a hike, a ride and so on.

## 2.REVIEW OF SOURCES

### Influences of social media technology

Bouhnik and Deshen (2014) pose that WhatsApp has gained ground in schools by becoming the major source of communication between instructors and learners. The researchers studied why the educators started setting up WhatsApp groups with the students stating that motivation, communication, dialogue creation and the establishing of a positive social environment are the main reasons for such a set up. A major advantage of using What's App is its practically 'no –cost' charges, the simplicity of use and the more private nature of such an application. The study proved the following important outcome :the teachers became more aware of who each student was, what their motivations and their setbacks were.; there was a more open, positive social environment in class; the conversation was deliberate due to the teacher's presence, thereby making the students improve in their language ; the students who were not usually too participative in class, were more socially interactive on WhatsApp; it was also easy to search for materials pertaining to their subject through search engines; the learning experience could be had beyond the class room enclosure. Over the course of using the WhatsApp the groups went beyond the initial aims of the use. Communication between instructor and learner became more convenient and challenges could also be addressed and overcome. Being skilled at team work is one of the major needs of the century and one way of acquiring this skill is WhatsApp group communication. It was concluded that if WhatsApp were to become a common tool in the educational environment, research would need to be conducted to find out the best form of integration in pedagogical goals.

Still on the topic of learning influences, Yeobah and Ewur (2014), believe that it goes without saying that the use of social networking applications among the tertiary students in Ghana will rise alarmingly. The study carried out sought to examine how the students felt the use of the WhatsApp Messenger impacted their learning process and the learning environment. For the purpose of the study, 50 students from 50

institutions of higher studies were interviewed and 500 questionnaires were distributed after which the results revealed that the students felt negatively impacted by the use of this application in the following ways : (i) the focus was more on the messages, in responding to them which was time consuming; (ii) the grammatical construction kind of went haywire and the spellings became atrociously erroneous; (iii) it led to putting off what needed to be done then and there for a later time; (iv) time management between academic and online activities suffered an imbalance and (v) study schedules were not being kept to , among other difficulties. This easy-to-use application became cumbersome in the long run due to the reasons stated above, thereby causing an overall negative effect on the Ghanaian students.

According to Maniar and Modi (2014) applications such as WhatsApp helped connect friends and relatives in spite of the busy schedule of those who work and study away from home and familiar contacts. The cost of calls can be cut down dramatically by switching to WhatsApp messaging. College students from a college in western India, who were the sample of the study conceded that they used WhatsApp most of the time and that besides effective communication with family, there was much they gained through its use in their education.

Smart phones, the devices into which the applications are featured are causing a revolution even in the field of medicine. Johnston, King, Behar and Sevdalis (2015) looked at the use of WhatsApp in a medical environment, where it was found that interns made use of this application for initiating communication in which clinical questions were asked and responded to. The communication between team members of a clinical group proved to be beneficial and seemed to be a novel way for senior physicians to monitor their junior colleagues without increasing patient risk.

In the field of journalism, social media technologies are giving rise to the new journalist, where more and more readers are following the accounts of preferred individual journalists over the general news published. Thomas (2014) says that many of the websites such as Hufftington Post began as blogs but have risen to become major websites that receive many clicks and comments daily. The new journals are themselves sourcing materials from the internet and various social networking websites to tell their stories. This has given rise to news agencies that help to verify the content that these journalists churn out. The new breed of journalists are becoming more self-sustaining and fitting into their own moulds, creating a fan following, to put it aptly. This development apparently encourages healthy competition … or does it? Thomas also says that the fact that WhatsApp processes billions of messages from its ever-increasing number of users only makes it a major presence in the news scenario and this further helps journalists to disseminate information with a click of a button.

Lodder (2014) claims that within a year or so, smart phone users will total more than five million and that more and more apps will be designed to contain address books, music videos, audio, phone logs (history) and so on. Lodder is concerned about privacy laws which need to be considered especially with the proliferation of smart devices. Terms of contract, for the main purpose of privacy must be re written and laws need enforcement. In the case of WhatsApp, the Dutch Privacy Authority states that this application had violated globally recognized privacy rules such as retaining

the numbers of non-users in the hash form rather than deleting them. In this research, the author also discusses the copyright issue of the Pandora app in the European Union. The internet has always caused problems for the legal system since its establishment. Further, the widespread mobility of the population crossing over fluid boundaries of regions only increases legal problems if found that certain information is carried over the borders that should not have crossed over in the first place.

The sources have looked at and linked the use of and impact of WhatsApp in the fields of education, journalism, law and medicine.

## 3.METHODOLOGY

A qualitative methodology has been adopted in this paper to extract the participants' opinion (Creswell, 2007) of the *whatsapp* phenomenon. The informants were chosen on a purposeful sampling basis (Patton, 2002) due to the breadth of knowledge provided for the data required in this research. Voluntary participation of 15 WhatsApp groups comprising 200 group members *in toto* ranging from work-groups to interest groups in the Klang Valley and across the globe helped in providing the relevant information for this study. The participants are mostly urbanized and globalized individuals.

An in-depth interview approach via WhatsApp, Skype and face–to-face was adopted for the various groups thereby enriching the corpus. The face-to-face interviews were video recorded on a Samsung Galaxy Note 4 smart phone with the consent of the participants, a few in groups (four of them at one sitting). The open-ended questions that were posed to get the participants' opinions enabled the authors to understand their perspectives of the group-based applications. Pseudonyms were provided to the participants and the interview later transcribed verbatim. It is note-worthy that the group names were the real ones.

Three groups of five members in each group, including the group administrator participated in face-to-face interviews at three different eateries at three malls in. Malaysia's capital city Kuala Lumpur, namely Wangsa Walk, Mall, KL Festival City and Nu Sentral. One group was interviewed at one of the government-run secondary schools after school hours. Twelve members were met at the workplace to conduct the interviews. The younger participants (aged 13-15) were excited at being recorded while being interviewed and were very enthusiastic is answering the questions. Prior permission was obtained from the parents to interview the school going groups. The two groups were represented by four members, including the group administrator from each of the larger groups.

Peer checking was conducted on the interview transcription in order to ensure the accuracy of information. A few of the extracts will be highlighted in the process of the discussion.

## TAKING THE WHATSAPP TRAIL

There were certain interesting findings about how users go about creating names for the groups, and about who joins the groups, the frequency of access, what group administrators have to say about the members and what users think are the pros and cons of WhatsApp messaging.

Table 2-Group Names, types and member count and frequency of messages

| Name of group | Group Type | Number of members |
|---|---|---|
| Breaking News | School(class) | 21 |
| Life in School | School (class) | 8 |
| Staff | School (teaching staff) | 12 |
| Badminton Buddies | School Extra Curricular | 9 |
| Hermonites | Alumni (school) | 28 |
| D'Angels | Alumni (school) | 16 |
| OLA | Alumni (college) | 20 and counting |
| Gen X Moms | Alumni (college) | 8 |
| Eagles (fall out group) | Alumni (college) | 7 |
| English for Engineering | Work group | 4 |
| English for Management | Work group | 4 |
| Department | Work group | 16 |
| Revision | Study group | 4 |
| The Paperback Book Club | Similar interest/s group | 28 and counting |
| Hot Trekkers | Similar interest/s group | 15 and counting |

As is seen in the above table, some are very generic names whereas others are personal and creative, such as the choice of 'Breaking News' by a secondary two class group. Some are easier identifiers of the group type for example the departments, the profession or the interests. The fall out group is something like a sub-group, breaking away from a larger one as the people decided the others either did not serve their purpose and/or it paid to have a sub group due to similar objectives. The numbers varied from a mere 4 to almost 50.

The participants were also asked about how frequently they messaged and most replied that it was usually a thread or a stream; one kick-started a greeting or a particular topic drawing in members who were interested to contribute to the conversation thread. The content comprised jokes, pictures/photographs, video clips, comments and since the past year or so, voice recording which included birthday greetings and songs. A few members of the group just followed the thread without commenting or contributing. Many were aggressive, active members, giving comments at every possible juncture, or responding to whatever was asked. In some cases, the members just wanted to share information or opinions on issues without expecting feedback. There were a few passive members who were literally 'invisible' until they were drawn out of their 'hibernation' by the constant reminders and 'knock knock where are you' questions.

**Groups-in Sync**

The various group administrators were asked a few questions to which the responses are tabulated in the table below:

Table 1: Group Admin responses to being admin, reasons for starting the groups and the pros and cons of being on the group platform.

| Questions | Jay | Sam | Aliya | Kayla | Sunny |
|---|---|---|---|---|---|
| How did you become the group administrator? | By popular demand. I am the geek of the group so it fell upon me | As the unit leader of the subject, I took the initiative | I volunteered | I have good compute r skills so my circle asked if I could start the group | Asked by my Boss |
| Why did you start the group in the first place? | There was a need to keep people in the loop | To ensure that the informatio n, especially urgent ones reach the group | I just wanted us all to stay in touch and support each other | To keep everyon e updated on relevant activities or any decision needed to be made | Asked by my Boss |
| What are the advantages of being in a WhatsApp group? | To get informati on on time; to exchange opinions | To be able to consult with one another; it helps in decision making regarding tests, exam questions, assignment assessment s and so on | To know what is going on with one another, to share interesting jokes and pictures, to plan trips and make proper decisions, just to chill as it is more personal and intimate than | Share informat ion on subjects, ask about homewo rk, to ask for help regardin g class work, to confirm about timings for co- | Connecti ng with the unit, getting updates regularly, sometime s on an hourly basis |

| | | | | | |
|---|---|---|---|---|---|
| | | | Facebook | curricular activities | |
| What are the disadvantages, if any? | Interference as you have to make sure you check at intervals and respond | Other elements such as 'forwards' are shared and two members share responses whereas others are silent | Some 'forwards' are repeatedly sent; some very long extracts are sent which are difficult to read at one scroll; too many emoticons are used. | Some classmates keep responding and others are very quiet | Have to be on our toes and check messages all the time; we feel restrained as our Boss is part of the group; opinions given are done with careful consideration; and be asked to change the icon or add on group members |
| How are the responses of the members? | Most are active; for example, if one person falls ill, the rest ask about who will visit the colleague, etc | Everyone participates when important questions are asked | Participation is generally very active. Interesting to know how different people respond. There are at times a deluge of forwards or pictures ; then complete silence | If no wi-fi, no response then everybody starts responding – difficult to control the 'flood' of exchanges | Everyone responds accordingly; |
| How many members do you have as of now? | 12 | 4 | 20 | 8 | 16 |

| | | | | | |
|---|---|---|---|---|---|
| (December 12, 2014) | | | | | |

The authors tabulated responses of five group administrators via face-to-face interviews and voice exchange on WhatsApp. Their responses helped to study how and why the groups are formed in the first place. Some group admin mentioned that despite benefits of being the group admin, they did experience some negative feelings.

The members' responses to what they (ten considered here) thought of this social platform are seen in the table below. Pseudonyms were used for the participants.

Table 3: The responses of group members towards WhatsApp groups.

| **Names** | **What do you consider good about WhatsApp groups?** | **What about the not-so-good?** |
|---|---|---|
| Amin | I think it is cool – the fact that so many of us connect at the same time and are updated about each others' lives | Nothing. I enjoy the features and the fact that that we can share photos and video files |
| Bee Ling | The fact that we can get the message across to the concerned people all at once | Some group members are slow to respond or give feedback that may be urgently needed |
| Charlie | The video clips sharing; getting news fast | The fact that when you have nothing to say you still need to respond. I have begun to use emoticons to sum up what I want to express |
| Della | Some members update us on activities; we also get to know of important news that we may not read of in the papers | Too many messages come in so you may miss a few; tedious to keep scrolling back |
| Eugene | Sharing forwards (some 'corny' ones), info, photographs, takes on several issues, video clips and so on | Some forwards are too long-winded and a hassle to read at one go; some video clips |

| | | |
|---|---|---|
| | | are time-consuming |
| Francis | Sharing jokes with the group; also asking about homework; discussing assignments | Not always in a wifi area so cannot access the messages |
| Gayatri | To exchange information on activities and planning outings; getting everyone's contribution towards an issue | Not getting immediate responses from other members |
| Humera | Getting updated information; giving feedback making you have a sense of belonging to the company | Constraints. Cannot always give my honest opinion so as not to offend anyone or to be in the bad books of the boss |
| Jenny | Getting support from friends; can discuss issues that bother me and to get advice on how to go about things | I love my WhatsApp group …fullstop |
| Ken | Love the features; can express how and what I feel through emoticons; better than FB in many ways as I can communicate with a specific group | Nothing really.Just annoyed with repeated forwards |

Most of the members seem to like group messaging because of information sharing, in the form of news updates, photos and video clips. Some were just happy to stay connected and there were members who thought the features were amazing thereby making communication as a member of a WhatsApp group an interesting experience.

The authors also wanted to know about what kind of 'personalities' were seen in a WhatsApp group and the members interviewed were happy to share about their group member 'types'

These are some of the 'personality' types and what they do.

Table 4: WhatsApp Personalities and their characteristics

| **Personality Type** | **Characteristic(s)** |
|---|---|
| The Greeting Crazy | Sends greetings ranging from 'hello' to 'good morning', 'good night', 'happy Wednesday', 'blessed' weekend; occasion greetings (birthdays, anniversaries, etc) |
| The Forward | Sends all manner of forwards (including pictures and video |

| | |
|---|---|
| Fanatic | clips) often unchecked |
| The LOL Lover | Types in LOL as a response to funny forwards or comments |
| The Smiley Savvy | Always responses wordless and only with 'smileys' for all manner of emotions :sad, happy, apologetic, confused, annoyed, angry, sleepy and many others |
| The One Word Wiz | Responses with ONE word only - yes, no, hmmm, uff, good, great, terrible,<br>true, sure, can, goodness, ok, alright, fine, amazing |
| The Jet Set | Arouses group members' envy by posting pictures of holidays in exotic locales or at important meetings, expensive concerts and happening parties |
| The Plan Maker | Plans meetings etc .. Let's watch …; let's lunch at … |
| The Absent One | Almost no response so other members wonder if they exist at all |
| The Silent Follower | Follows the group progress and conversation threads silently; does not say anything |
| The Icon-ic Idiot | Plants icons all over the messages; messages are interspersed with emoticons –indicating food, books, night, day; changes group icons according to occasions or the weather |

For personality types, the authors were interested to know how members of a group responded to various texts and decided to come up with names for those who characteristically made use of certain expressions. Listed in the table are only ten of the 16 personality types the authors managed to categorize.

The findings based on simple interview questions helped the authors identify the way classmates, clubbers and colleagues communicate in an 'open' group, serving the crux of the research.

## 4.CONCLUSION

This paper shed a perspective on the pervasive social media WhatsApp phenomenon which is here to stay until a 'user-friendlier' model of social media connectivity is created. This study can pave the way forward to a communication-based theoretical model of research that incorporates technology such as seamless internet connections for a broader perspective of group communication APPlications. The sample can be larger and incorporate groups such as family, friends, professionals, sportspersons, media persons, researchers and so on.

## 5.ACKNOWLEDGEMENT

Our sincere gratitude goes out to all the participants who took part voluntarily in this study, making it possible.

## References


1. Bouhnik, D. & Deshen, M. 2014. WhatsApp goes to school: mobile Instant messaging between teachers and students. *Journal of Information Technology: Research, (13),* 217-231
2. Creswell, J. W. 2007. *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). London: Sage Publications.
3. Johnston, M.J., King, D., Behar, N. & Sevdalis, N. 2015. Smartphones let surgeons know whatsapp: an analysis of communication in emergency surgical teams. *The American Journal of Surgery, 209*, 45-51
4. Lodder, A.R. 2014. Cross-border use of whatsapp, pandora, and grindr: on global norms and how to enforce these when 'there' can be everywhere. *BILETA*
5. Maniar, A. & Modi, A. 2014 *Educating whatsapp generation through whatsapp. Zenith International Journal of Multidisciplinary Research, (4)* 8, 23-38
6. Patton, M. Q. 2002. *Qualitative research & evaluation methods* (3rd ed.). London: Sage Publications.
7. Thomas, G.W. (2014) The new journalism, the new journalist.*Journal of International Scientific Publications (3*), 1-5
8. Yeboah, J. & Ewur, G.D. 2014. The impact of whatsapp messenger on students performance in tertiary institutions in Ghana. *Journal of Education and Practice, 5(6),* 157-164

# The Role of Educational Institutions in Going Green : E-Learning Perspective

**Saraswathy Thurairaj[1], Farida Bhanu Mohamed Yousoof[2], Komathi Munusamy[3]**

[1] Universiti Tunku Abdul Rahman
tsaraswathy@utar.edu.my

[2] Universiti Tunku Abdul Rahman
faridabhanu@utar.edu.my

[3] Universiti Tunku Abdul Rahman
komathy@utar.edu.my
Faculty of Creative Industries & Faculty of Accountancy & Management (FAM)
Lot PT 21144, Jln Sungai Long Bandar Sg Long, Cheras 43000 Kajang Selangor

**Abstract.** Globalization has changed the world as well as how customer judges the value of a product. Environmental issues such as air and water pollution, ozone depletion, acid rains, overpopulation, global warming and hazardous waste had caused significant rise in environmental issues. Therefore, since 1970s, it has led to the rise of the green revolution to reduce further destruction to the environment. Green marketing is referred to all activities which are designed to facilitate and create the exchange intended in order to satisfy human needs, to cultivate the high level of satisfaction and therefore lesser destruction impacts on the natural resources in the world. As a result this study aims to identify the challenges of green management in Malaysia and how educational institutions in the country can address this issues and challenges with the use of information system such as e-learning. Online program in the form of e-learning will provide the opportunity to students and academicians to learn about green issues, global sustainability and associated cost cutting practices. Interest in environmentally friendly methods is booming and those trained to understand how to address sustainability issues will be able to market themselves better to prospective employers.

**Keywords:** Going Green, Green Marketing, Green Management, E-Learning, Educational Institutions, Sustainability

## 1.Introduction

According to United Nations Department of Public Information 1997, there is an issue for the global environmental concern in the world and it is getting more human attention since 1970's. In addition, the concern has pulled the green light of the academic researchers and has become a critical discussion topic in the world today. Environmental issues such as air and water pollution, ozone depletion, acid rains, overpopulation, global warming and hazardous waste had caused a significant rise in environmentalism in the Western countries. The consumers' behavior towards environmental friendly product can be affected by the rise of the environmental awareness in the public. Therefore, since 1970s', it led to the beginning of the green product revolution in order to further reduce the potential destruction to the environment (Alwitt and Pitts, 1993 and Elham and Nabsiah, 2011).

The consumers purchasing pattern has been growing over the past decades (Punitha Sinnapan, 2011) with a lot more focus on the environmentally friendly products. As such the manufacturing processing and the consumption of goods involving the handling and harming of natural resources such as air, timber, fossil fuels ore and water have come under lot more scrutiny by the consumer and the regulators in the much more developed nations. The excessive consumption and exploitation of the natural resources have caused continuous environmental destruction. Hence, the awareness of environmental problems and the right respond to these issues are absolutely necessary in order to ensure that the world and its natural resources are managed in more sustainable manner.

Environmental marketing or green marketing is referred to as the activity which is designed to facilitate and create the exchange intended in order to satisfy human needs and at the same time to cultivate the high level of satisfaction of the consumer, while having a minimal destruction impacts on the natural resources in the world (Polonsky, 1994). According to Shrikanth and Raju (2012), green marketing refer to the holistic marketing concept where the marketing and production process use disposable products and services which are less harmful to the environment in order to prevent the impacts of non-biodegradable solid waste, water pollution, air pollution and other dangerous impact of pollutants that leads to the increasing global warming.

In the past decade, more companies that embraces "eco friendly" and "sustainable development" in their marketing messages as well as their corporate mantra has seen the benefits of going green. As part of the efforts in going green, more corporate has embarked on a wide variety of measures such as planting tree, cleaning rivers, recycling as part of their social corporate responsibilities and at the same time investing in the innovation and the creation of affordable green products. With rising energy cost and threat of global warming, many businesses begin to realize the advantages of using green technology to reduce their carbon footprint and minimize waste. In addition, consumer awareness of climate change and protecting the environment are fast rising. Whether the consumers are willing to pay a premium for green products and services are still under continues discussion and debate, the consumers in general have higher regards for companies that have gone or attempting to green.

Consumer and marketers' are becoming more concern about the environment and there are many opportunities in the production and marketing of green product and services today and in the future. In addition, to establish a green product and service may be costly in the beginning, but the cost of product will be brought down in the future due to effective economy of scale as the consumer demand grows. However, while there are many research led influences' that will determine the purchasing patterns of the consumers', the demand and attitudes on green product is likely to be different across market segment and cultures (Ottman, 1992a, b).

While the scientists and researchers continue to highlight the adverse effects to the Mother Nature as a result of artificial products and interventions, the world is beginning to realize that the dangerous environmental impacts of irresponsible manufacturing and poor living habits practiced by the humans today. According to

Tehrani, Sinha (2011), there are increasing awareness towards green product campaigns and educational programmes' on the impact of becoming green for the sustainability of the environment.

Madu (2001) found that the influence of advertisement that able to encourage green efforts and green living to successfully draw consumers' attention to green products. In addition, his research shows the existing market for green product is still relatively new but with huge potential to grow in future. In order to fully implement and market green product to the consumers', there must be continuous and comprehensive studies of consumers behaviour and their decision on the green product. Morever, the factors influencing the green product purchase intention are positively correlated with the consumers' level of confidence toward the source from which the product is derived (Vazquez-Brust and Sarkis, 2012). According to Moreo (2008) and Madu (2001), to influence green product purchase intention, it is important to make sure that consumers are well-educated and understand the green product, its origin and the authenticity of its contents.

Malaysia is the one of the earliest nations in the world that has taken stern actions in regards to the environment by endorsing the Environment Quality Act way back in 1974. Apart from passing the acts, the government has also formed the Ministry of Energy, Green Technology and Water lately in order to provide the growing need and significance of green technology towards sustainable advancement (Nizam et al, 2014). The government has taken a general move which is known with the acronym of AFFIRM for the sustainability of the environment. AFFIRM stands for Awareness, Faculty, Finance, Infrastructure, Research Development, Commercialization and Marketing. This acronym is mainly proposed to acquire liability from all stakeholders in committing towards the environmental protection in Malaysia (Malaysia Green Forum, 2010). Datuk Seri Mohd Najib Razak in 2009, pledged to reduce the carbon intensity by 40% by 2020. The Malaysian Government has repetitively declared green technology to be the mainstay of its agenda and one of the drivers of the country's economy. In order to promote green technology the government during last year's budget presentation had announced the establishment of a RM1 billion fund to provide soft loans to companies that supply and utilize green technology. It also proposed to develop Putrajaya and Cyberjaya as pioneer townships in Green Technology as the showcase for development of other townships.

Green environment has become a main issue in many industry sectors globally, including higher education. Educational Institutions' (EIs) are repeatedly searching and discovering ways to conserve energy efficiently, deliver energy using alternative methods, resources and waste recycling technology (Prasad, 2012). EIs are being proactive in educating students, faculty staffs on the issues. E-learning is one of the strategic alternatives which can be used to deliver knowledge on green issues to students in higher education segment. However, how many EIs are really using the e-learning for promoting green issues is still questionable.

## 2.Problem Statement

The fast and rapid growth of the global economy has a strong positive relationship between consumers' consumption around the world. The environment deterioration caused by over utilization of natural resources and over consumption from the consumers is always a critical issue and getting attention from the public. If the environment deterioration continues to worsen, it will become a permanent common concern especially for developed countries. In addition, it is also alarming the developing countries toward green direction to protect the natural environment.

According to "The Star" (2008), Malaysia produced approximately in excess of 7.3 million tonnes of garbage which is equivalent to the forty two units of Petronas Twin Towers in the year 2007, as such it is paramount for Malaysians' to be reminded of the importance of waste management and benefit of going green with green products and to reduce the waste. According to Elham and Nabsiah (2011), there are a lot of green products such as eco-friendly products which have been introduced in Malaysia but have been unsuccessful in getting the attention from the public. Fortunately, the dynamics of the buyer-seller interaction will lead to further improvement of the green revolution and it will benefit the whole country in the future (Ottman, 1992a).

Although it has been widely accepted by researchers that marketers' plays an important role in convincing consumers towards purchasing green products, yet it cannot be denied that organizational efforts have been slow to change the attitude of consumers towards a favorable green purchasing behaviour. Therefore, EIs play a vital role in creating awareness to public through online program or e-learning by collaborating with government and the private sector. Online program in the form of e-learning can provide the opportunity to students, academicians and employees to learn about green products, global sustainability and associated cost cutting practices.

Hence, EIs must be much more effective in the development of the green on-line programs so that the awareness on green issues can be instilled and increased amongst the students and academicians. This will then allow green issues to be communicated in more efficient and effective way.

A review of the past literature shows that no study has been conducted in testing the role of e-learning in promoting green issues. According to Nizam, Mansor, Mukapit, and Yahaya (2011), based on studies conducted by the researchers' on the Malaysian perspective, the consumer behaviour shows the lack of intention and unwillingness to accept the purchasing of green products in the local market. Many studies also suggest that this could be due to various reasons such as lack of consumer education, awareness, and the lack of quality green products in market place. Apart from that, the lack of government efforts in promoting green issues in a meaningful and sustainable manner had also contributed to lackluster understanding and the importance of the green issues and green products.

Furthermore, previous studies on green marketing primarily emphasized on leadership, qualities, organizational structure, employee participation and incentive system as the success factors for green marketing. However, not many had purposefully looked into the effectiveness of EIs, and its participation and

contribution for a more successful and sustainable understanding of green issues among the consumers. The participation and collaboration between EIs and manufacturers' and the marketers' of green products in a meaningful way can lead to a much more successful marketing effort of green products and technologies.

## 3.Going Green

According to Lee, Ling, Yeow and Hasan (2012), green product is defined as a product that is designed or attributed, by using recycle resources, which will benefit the natural environment and reduce the toxic effect toward the natural environment on its entire life-cycle. Meanwhile Rezai (2007) stated different perspective for the definition of green product and referred it to the product that is not a threat to the natural environment, does not deplore any natural resources, and can be recycled and has re-usable packaging or non-toxic materials.

According to Dobson and Lucardie (2002) the term of green whereby do not have relationship with shade, however the name of green is the decision in fitting on the grounds that green is a colour that is basically related with nature. Green product item is viewed as a kinder to the earth compared with non green product. The term is widely used and it does not take much for an item to be called green, so it can blanket an extensive variety of customer product. An item is considered earth accommodating in the event that it is biodegradable, implying that it will represent no risk to the earth nature's turf, when it is discharged to the water, air or earth while being used or when discarded. This sort of item typically decays much faster in a landfill, compared with comparative things that are not biodegradable. Biodegradable families are from cleanser, dishwasher cleanser and clothing cleanser and so on. These are simply a couple of cases of this sort of green items.

According to Preston (2009), eco accommodating item defined as the item that do not destroy nature environment, whether in their utilization, generation or transfer. When using the developing environmental awareness item, it will help to monitor vitality, minimize carbon foot shaped impression, outflow of nursery gasses, and does not prompt contamination to the earth. Therefore, these do not destroy nature's turf or affect the environmental when they are discarded. There are some extra eco inviting items that are made from reused materials. This reuse item will help to decrease the requirement for new crude material and the measurement of waste will be sent to landfill where this waste could be redirected in making recycled item. Thus, interest for these reuse items will help to finish the reusing circle, and manage the reusing business. Thus, the eco cordial item could extend from eco friendly dress, naturally agreeable family cleaning item, and eco gifts that can be given to friends' and family.

Green marketing tools such as eco-label, eco-brand and environmental advertisement will make the perception and awareness of green products attributes and characteristics easier. This will in turn guide the customers to purchase environmentally-friendly products (Elham & Nabsiah, 2011).

## 3.1 Issues and Challenges in Going Green in Malaysia

The research relates to factors influencing green purchasing behaviors that have been carried out especially by developed nations. Nevertheless, dominant factor that affects the green purchasing behavior in one country can be different in another due to socio-economic and cultural differences of each country.

Green marketing activities are increasing in Malaysia, however marketers should stress the ecological knowledge in their organizations, their products and their advertising in order to achieve the goal of shifting consumer into purchasing green products (Mendleson, 1994).

Nevertheless in the context of green marketing, which has been aggressive on a global level, Malaysian businesses provide less likely example of this form of green marketing as compared to East Asia, North America and European countries. According to Azizan and Mohd Suki (2013), in Malaysia, consumers' acceptance on green product initiative has trended on a dramatically slow pace but still considered to be on an uptrend to achieve a significant positive change in green purchase intentions. Moreover, they also mention that a lack of availability on green product for the retail shelves could also be another factor to cause the low purchase intention of green product. They also stressed that although there are some levels of awareness from consumers towards the need for the environmental friendly products they are unwilling to spend money or pay the high prices for the green product. Although the intention for purchase may exist, such intention is deterred by other factors or elements that can create the positive purchase intention of consumers.

## 4.E-Learning System

E-Learning is rising as the new concept of modern education. This study considers e-learning as it applies to workplace or organizational environment. Due to its flexibility to access, just-in-time delivery and cost effectiveness, e-learning has been implemented by many organizations especially by small and medium–size enterprises for support of learning and training in the workplace (Sambrook, 2003; Driscoll, 1998). Furthermore, many countries are using e-learning technologies more aggressively to expand their educational market. E-learning is defined as "the use of computer network technology, principally over or through the Internet, to deliver information and instruction to individuals" (Welsh et al. 2003). Nevertheless, the challenge is to establish a right relation between academics and industry and convert industry as a strategic partner in all IT enabled and entrenched higher education initiatives (Prasad, 2012).

The rise of knowledge economy, in which economic productivity and growth is highly dependent on the development of new knowledge that creates a demand for highly educated workforce and life-long learning opportunities. The emergence of Web 2.0 technologies has been considered as a transformation of e-learning from a central controlled education system to an interactive and conversational learning network. However, there is a lack of overall strategies for the use of Web 2.0 technologies in e-learning applications. According to Rollett et al (2007), the Web 2.0 e-learning applications have been developed without profound business models. Most

applications are executed poorly in helping employees to improve their performance, or enable their social interactions integrated with their learning practice, which ultimately fail to achieve the organization's goals for success in the knowledge economy (Ran & Wang, 2008; Roy & Raymond, 2008). Web 2.0 e-learning, in particular have failed to align organizational goals and individual needs in a systematic way. While using e-learning in workplace, it is critical to consider how to shape an education system which fulfills the needs of learners and overall needs of the organization (Attwell, 2007)

Owing to lack of a general understanding of the workplace e-learning and appropriate conceptual and methodological tools for implementation, e-learning in the workplace remains a disintegrated, complex, and challenging area of research and practice (Servage 2005; Collin 2006). Incorporating green issues into e-learning is not only increases the understanding and provide the opportunity to the employees about the green issues but also will benefit the organizations by saving time and money by offering a 24/7 accessible online environment. In order to execute the online program efficiently, EIs can play a major role by designing a green technology program by using e-learning platform that can help the companies to improve environmental sustainability. Moreover, EIs can also extend the online green technology program to the faculty staff and students without incurring the time and expenses to build a separate curriculum. Students and academic staff trained in green and sustainability issues will able to market themselves better to prospective employers.

## 5.Conclusion

Green issue management strategies have become essential topic in many industry including educational institutions. In order to be successful, each organization must ensure that each and every employee has an in-depth knowledge pertaining to green issues. Technology will be able to play a vital role to improve the awareness and corporate sustainability which can then lead to increase in profits for organization while at the same time reducing the need for governments to spend lots money in managing green issues. The study has proposed EIs to use e-learning technology via collaboration with government and private sector to improve the public and employees' understanding on green issue management. The e-learning system should be able to align both individual learning needs and organizational goals in green issues. This will provide insights how EIs can plan ways to promote green management to organizations and in teaching.

## References

1. Alwitt, L.F. and Pitts, R.E. (1996).Predicting purchase intentions for an environmentally sensitive product.Journal of Consumer Psychology, Vol. 5 No. 1, pp. 49-64.
2. Attwell, G.(2007). Web 2.0 : A new wave of innovation for teaching and learning. Educause Review, 41, 32-44.
3. Azizan, S.A.M. and MohdSuki, N. (2013).Consumers' Intention To Purchase Green Product: Insights from Malaysia, pp. 1129 – 1141. Malaysia, IDOSI Publications.
4. Collin, K.(2006).Connecting work and learning: Design engineers' learning at work. Journal of Workplace Learning, 18(7/8), 4013-413.

5. Dobson, A., and Lucardie, P. (Eds.). (2002). The politics of nature: Explorations in green political theory. Routledge.
6. Driscoll, M.(1998). Web-based training: Using technology to design adult learning experience. San Francisco: Jossey-Bass, Pfeiffer.
7. Elham, R. and Nabsiah, A. W. (2011).Investigation of green marketing tools' effect on consumers' purchase behavior.Business Strategy Series, Vol. 12 Iss: 2, pp.73 – 83.
8. Lee, C. H., Heng, Y. L., Jian, A. Y. and Hasan, M. A. (2012*).Assessing Determinants of Green Purchase Intention*, pp. 25 – 30. Malaysia: Multimedia University.
9. Madu, C. (2001). *Handbook of Environmentally Conscious Manufacturing*, pp. 190 – 199. U.S.A.: Kluwer Publishers.
10. Mendleson, N. (1994), "Environmental marketing business partnering at AFC-developing strategic alliances", Environmental marketing: Critical Success Strategies for Reaching the Green Consumer, Proceedings of the 1994 Conference, Sdyney, February 22-23, pp 273-88.
11. Moreo, A. (2008). "*Green Consumption in the Hotel Industry: an examination of consumer attitudes"*, pp. 195 – 296. U.S.A.: ProQuest.
12. Nizam.,NurulZarirah.,Rajiani.,Ismi.,Mansor,Nusaibah.,Yahaya.,Siti Norbaya.,Hoshino.,&Yasuo (2014), Understanding Green Purchase Behaviour among Gen Y in Malaysia by Examining the Factors Influenced. Interdisciplinary Journal of Contemporary Research in Business 6.2, 181-194.
13. Ottman, J. (1992a). Sometimes consumers will pay more to go green.Marketing News, pp. 6-16.
14. Nizam, N.Z., Mansor, N., Mukapit, M. and Yahaya, S.N. (2011).Factors Influencing Customers' Decision To Buy Green Product Design in Malaysia, pp. 362 – 368. Malaysia:University Technical Malaysia Malacca.
15. Ottman, J. (1992b). Green Marketing: Challenges and Opportunities for the New Marketing Age, NTC
16. Polonsky, M. J. (1994). An Introduction To Green Marketing. *Electronic Green Journal*, 1(2)
17. Prasad, S.K (2012). Emerging Global E-Learning Trends. Asia Pacific Journal of Management & Entrepreneurship Research 1.3, pp 353-363.
18. Preston, M. (2009).Practically Green: Your Guide to Ecofriendly Decision-Making. F+ W Media, Inc.
19. Punitha Sinnappan and Azmawani Abd Rahman, 2011.Antecedents of Green Purchasing Behaviour among Malaysian Consumers. International Business Management, 5:129-139.
20. Ran,W.,& Wang,M.(2008).Develop adaptive workplace e-learning environments by using key performance indicator. Proceedings of International Conference on e-Learning in the Workplace (ICELw), New York.
21. Roy, A.,& Raymond,L(2008).Identifying and lowering the barriers to e-learning for SMEs, Proceedings of International Conference on e-Learning in the Workplace(ICELW), New York.
22. Rezai, G. (2007). Factors Influencing Public Intention towards Purchasing Green Foods In Malaysia, pp. 1- 20. Italy: University Putra Malaysia.
23. Rollet, H.,Lux, M., Strohmaier, M., & Dosinger, G(2007). The Web 2.0 way of learning with technologies . International Journal of Learning Technology, 3(1), 87-107.
24. Sambrook, S.(2003).E-learning in small organizations. Education+Trainning, 45(8/9), 506-516.
25. Servage, L. (2005). Strategizing for workplace e-learning: Some critical considerations. The Journal of Workplace Learning, 17(5/6), 304-317.
26. Shrikanth, R. and Raju, D.S. N. (2012) Contemporary Green Marketing – Brief Reference to Indian Scenario. *International Journal of Social Sciences and Interdisciplinary Research*, Vol.1 No. 1, January 2012, ISSN 2277 3630.
27. Tehrani, N. and Sinha, S. (2011). *Understanding Green Business*, pp. 320 – 331. U.S.A.: AuthorHouse.
28. Vazquez-Brust, D.A. and Sarkis, J. (2012).*Green Growth: managing the transition to a sustainable economy*, pp. 290 – 295. U.S.A.: Springer.
29. Welsh, E.,Wanberg,C.,Brown,K.,& Simmering,M (2003).E-learning: Emerging uses, empirical results and future directions. International Journal of Training and Development, 7(4), 245-258.

# Data Mining as a Tool for Teaching Lessons Organization

**Vesna Mufa[1], Violeta Manevska[2], Blagoj Ristevski[3]**

[1]vesna_mufa@hotmail.com, [2]violeta.manevska@uklo.edu.mk, [3]blagoj.ristevski@uklo.edu.mk

Faculty of Information and Communication Technologies – Bitola
St. Kliment Ohridski University in Bitola, Republic of Macedonia

**Abstract.** The good organization of teaching lessons contributes to ease understanding of teaching lessons by the students. In this paper, the results obtained by using data mining methods applied to the data set for students are presented. Results obtained from the models should serve for better organization of teaching lessons, whereby students can better understand teaching materials and teachers can attract and keep the attention of students.

**Key words:** data mining, data set, models, teaching lessons

## 1.Introduction

The task of each teacher is to make daily preparation for each teaching lesson. By making a daily preparation, he/she makes organization of teaching lesson. The lesson plan is a guide of what students need to learn and how it will be done effectively during the lecture. Prior to the teacher when plans its lesson is to identify the learning goals for the class meeting. Then, he/she can plan suitable activities. Every teacher should develop strategies to obtain feedback of student learning.

Pointing out concrete objectives for student learning should help the teacher to determine the kinds of teaching and learning methods and forms that will be used. These forms will define how the teacher will check whether the learning goals have been fulfilled or not. The relationship among student's objectivities is shown on Figure 1.

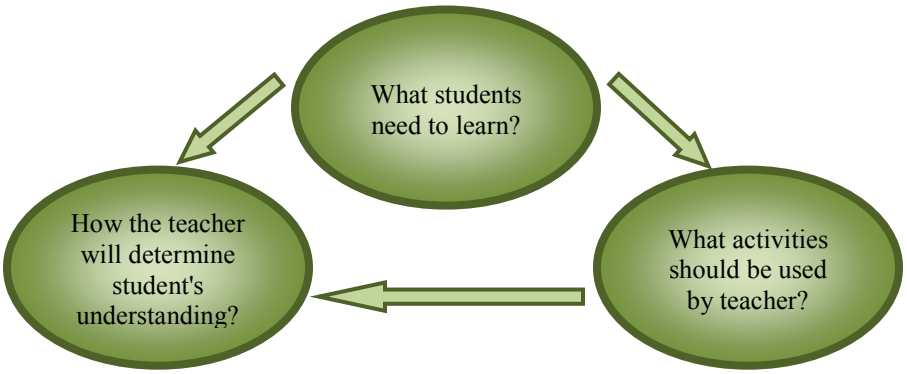



Figure 1. The relationship between student's objectives, teaching activities and student's understanding

### 1.1.Organization of Teaching Lesson through Lesson Plan

The well prepared lesson plan means better organization of teaching lesson because the organization of the teaching lesson is a reflection of the lesson plan. The primary thing that ought to be carried out is to determine what students need to learn and what they need to be able to do at the end of the class. Then, it should be follow these steps [1]:

1. **Introduction development**. When the goals are set, the introduction is planned. This part introduces students what will be done in the class. If new content is handled, the teachers have a different students with different individual experiences. They may already be familiar with the topic. That is the reason why the teachers might start with a question or activity to gauge student's knowledge of the subject. This information can help in shaping the introduction, learning activities etc. When teachers have an idea of the student's familiarity with the subject, they have a feeling of what to concentrate on.
2. **Plan the main part of the lesson**. The teacher needs to prepare several distinctive teaching methods and forms to get the attention of more students and appeal to diverse learning styles. The teacher should estimate how much time has been used on each method. Extended explanation or discussion should be considered.
3. **Conclusion development**. The teacher needs to go over the content covered in class by highlighting the primary purposes. This can be done in various ways: the teacher expresses the fundamental points itself ("Today we discussed about…"), or he can ask the students to summarize them. In this final part of the class, the teacher asks questions to students in order to check whether he achieves planned goals.

When planning the lesson, the teacher should decide what kinds of questions will be beneficial for discussion, not only in the final part, but throughout the whole class hour. He/she needs to make a balance between covering content (accomplishing learning goals) and ensuring that students understand them.

## 2.Data Mining as a Tool for better Teaching Lessons Organization

The good organization of teaching lesson contributes for easily understanding of teaching contents by the students. It is the purpose of the teacher: to attract and keep the attention of students and encourage them of thinking and draw conclusions, but the most important thing is to find a manner to make student learn continuously. But the best feedback about whether the lesson is well organized or not, theacher received by the students. Data mining is an excellent tool for analyzing the opinions of the students, in order to teachers to streamline the organization of lesson to achieve their goals.

The opinions are obtained using a survey conducted among students from the first year of a secondary municipal school. Each question corresponds to an attribute and each offered response amatch the value of the attribute. It tells us that the attributes have discrete values. Thus, a data set consisted of 12 attributes and 76 examples is obtained [2].

In Table 1, the questions from conducted survey, names of attributes that correspond to the questions and answers representing the values of the attributes are given.

Table 1. List of questions, attributes and answers of questions/values of attributes

| Question | Attribute | Answers of a question / Attribute value |
|---|---|---|
| Gender | gender | • male<br>• female |
| In processing new content, the most knowledge I gain through: | gained_knowledge | • discussion<br>• explanation<br>• demonstration<br>• presentation<br>• practical_work |
| Which teaching method is the most appropriate for you: | teaching_method | • frontal<br>• individual<br>• pairs<br>• groups |
| It is better when I work (at school): | work_in_class | • individual<br>• pairs<br>• groups |
| When I find difficult to overmaster teaching contents, I get information from: | getting_information | • teacher<br>• Internet |
| More reliable source is: | reliable_source | • teacher<br>• Internet |
| Do you think that the lesson is more interesting when it is supplemented with photos and video elements? | photos_videos | • yes<br>• no |
| Is teaching leaf helpful for facilitating your learning process and encourages you to think? | teaching_leaf | • yes<br>• no |
| Does research on the Internet contribute enhancement of your knowledge? | Internet_research | • yes<br>• no |
| Most of the teaching content I learn: | learn_teaching_lesson | • at_home<br>• at_school |
| Do you want when the teacher asks interesting questions during the class? | interesting_questions | • yes<br>• no |
| Teaching content I learn: | learn_teaching_lesson1 | • continually<br>• before_test |

### 2.1 Implementation of data mining methods on data set

For this research, predictive modeling methods were used, such as: classification rules and classification trees [6].

The purpose of predictive modeling is to predict the value of the class attribute based on the values of another attributes by using statistical analysis. In this case, the goal is through better organization of teaching lesson to mislead students of continually learning. The class attribute is *learning_teaching_lesson1*.

```
If photos_videos = no then continually
If gained_knowledge = explanation
  and teaching_leaf = yes then continually
If learn_teaching_lesson = at_school
  and gained_knowledge = explanation then continually
If gained_knowledge = practical_work
  and reliable_source = teacher then continually
If teaching_method = frontal
  and gained_knowledge = demonstration then continually
If gained_knowledge = explanation
  and teaching_method = individual then continually
If reliable_source = teacher
  and teaching_method = groups
  and gender = male then continually
If gained_knowledge = explanation
  and teaching_method = pairs
  and getting_information = teacher then continually
If teaching_method = individual
  and work_in_class = pairs then continually
If work_in_class = individual
  and gained_knowledge = discussion then continually
If work_in_class = individual
  and gained_knowledge = demonstration then continually
If teaching_method = frontal
  and work_in_class = groups then continually
If work_in_class = individual
  and interesting_questions = no then continually
If gained_knowledge = presentation
  and gender = female
  and work_in_class = individual then continually
If Internet_research = no
  and gender = female then continually
If teaching_method = groups
  and reliable_source = teacher
  and interesting_questions = no then continually
```

If gained_knowledge = presentation
and teaching_method = pairs
and gender = male then continually
If reliable_source = Internet
and work_in_class = groups then before_test
If reliable_source = Internet
and gained_knowledge = demonstration then before_test
If gained_knowledge = discussion
and teaching_method = pairs then before_test
If gained_knowledge = presentation
and Internet_research = no then before_test
If reliable_source = Internet
and gained_knowledge = practical_work then before_test
If gained_knowledge = presentation
and teaching_method = pairs
and gender = female then before_test
If interesting_questions = no
and gender = male then before_test
If gained_knowledge = presentation
and gender = male
and teaching_method = frontal then before_test
If interesting_questions = no
and teaching_method = frontal
and work_in_class = pairs then before_test
If Internet_research = no
and teaching_method = frontal then before_test
If gained_knowledge = presentation
and learn_teaching_lesson = at_school
and gender = male then before_test
If gained_knowledge = demonstration
and teaching_method = pairs
and gender = male then before_test
If interesting_questions = no
and teaching_method = pairs
and getting_information = Internet then before_test
If teaching_method = groups
and gained_knowledge = explanation
and work_in_class = groups
and teaching_leaf = no
and interesting_questions = yes then before_test

**Classification rules**

The classification rules are a good way to present extracted information or knowledge. The classifier based on rules used set of IF-THEN rules. Each rule specifies independent information or part of knowledge.

In order to obtain a model of classification rules, the data set was divided into two parts: 70% were used for learning, while the remaining 30% were used for testing. By using the algorithm **Prism** 31 rules are generated [3]. They are presented in Figure 2.

According to the generated rules, students who study continuously receive from the subject teacher as a trusted source additionally information about teaching contents. This is an indicator for additional teaching. Teaching leaf is motivation for students to be active and encourage their thinking, and provide teaching contents widely to overcome when they are at the school, but get up the curiosity contents to be learned at home.

The students, who study before the test, additional information about teaching contents obtain from the Internet, which they considered as a reliable source. The teacher should emphasize that it is advisable to explore the Internet in order to obtain additional information, but the truth of the information obtained from the Internet should be checked by the subject teacher. These students believe that teaching contents they learn at the school is sufficient to repeat them before the test, which is completely wrong. Teachers should give homework assignments to the students and they should respond if they read the content at home. Working in groups should be avoided as far as contributing to sharing ideas and building on existing knowledge, it only contributes to certain members of the group to be active, and the rest to be passive.

Teaching methods don't depend on the type of learning.

The accuracy of the algorithm is about 70%. From the confusion matrix we know how many of the examples are classified correctly, and how many are wrongly classified. The confusion matrix is presented in Figure 3.

```
=== Confusion Matrix ===

  a  b   <-- classified as
 14  2 |  a = continually
  3  2 |  b = before_test
```

Figure 3. Confusion matrix of Prism algorithm

From the confusion matrix can be noted that from 21 examples of testing, 16 are correctly classified, while 5 are incorrectly classified. For the class *continually,* 14 examples are correctly classified, while with the class *before_test* are incorrectly classified 3 examples. For the class *before_test,* 2 examples are correctly classified, while 2 examples are incorrectly classified.

**Classification trees**

Classification trees are a popular method of classification that result in standard decision trees where the root and internal nodes tested the values of attributes, and each branch represents the output of the test. The nodes - list contains the classes [5].

Trees represent relations between data found in the training data. The highest and most important node is called a root of the tree. Using the data for training, the method of classification trees generates the nodes such rules. Each node offers a choice of different number of alternatives, where the number of alternatives is equal to the values of the attribute representing the node. Each leaf is a classification or decision. The classification of new examples is done by testing the values of attributes through the tree, starting from the root and ending with the leaves. The method of classification trees is popular because of the intuitive representation and easily applicable rules for describing and understanding[4].

To obtain a classification tree, the data set was divided into two parts, with an identical proportion as a classification rule. Using the algorithm **J48, a** classification tree is generated, as shown in Figure 4.

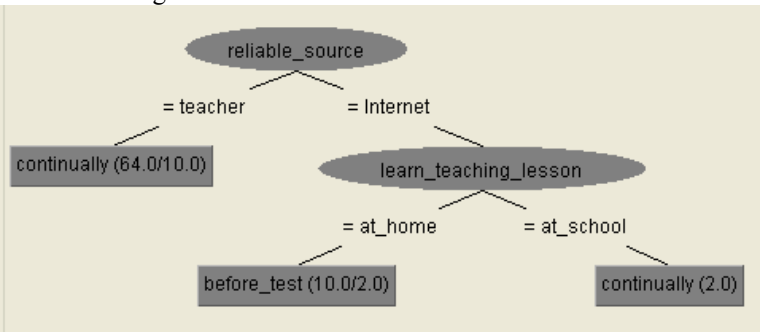



Figure 4. The classification tree generated by using the algorithm J48

From the resulting classification tree we can conclude that the type of learning depends on whether the teacher or the Internet is a reliable source of information. If the teacher is a trusted source of information, the students learn continually. If the Internet is a reliable source of information and if teaching contents are largely taught at the school, then learning is continually because most information is provided by the teacher. The students, who study before the test, additional information about teaching contents obtain from the Internet, which is completely true because the solution of all doubts they are browsing on the web due to the inability to ask the teacher because they learn just before the test.

The accuracy of the algorithm is about 87%, which indicates that the classification trees with greater accuracy predict examples as opposed to classification rules. The confusion matrix is presented in Figure 5.

```
=== Confusion Matrix ===

  a  b   <-- classified as
 18  0 |  a = continually
  3  2 |  b = before_test
```

Figure 5. Confusion matrix of the J48 algorithm

From the confusion matrix can be noted that from 23 examples of testing, 20 are correctly classified, while 3 are incorrectly classified. For the class *continually,* 18 examples

are correctly classified, while with the class *before_test* are incorrectly classified 3 examples. For the class *before_test*, 2 examples are correctly classified, while there aren't examples that are incorrectly classified.

## 3.Conclusion

The results obtained from the models (classification rules and trees) can be used for better organization of the teaching lesson. Good organization means that students might easier and with greater enthusiasm to study teaching contents, and it contributes to attracting and retaining their attention.

Teachers need to create an atmosphere in which students will feel free to ask questions or express their opinions, ideas and thoughts. It should also stimulate interest in lessons by using different methods, with the exception of group works, by setting up interesting questions and teaching leaves that will get them thinking and draw conclusions. Giving homework as often as possible makes teaching contents to be learned at home. This will provide continually learning by the growing number of students.

## References

1. Milkova S., *Strategies for Effective Lesson Planning, Center for Research on Learning and Teaching*, (http://www.crlt.umich.edu/gsis/p2_5).
2. Mufa V., Manevska V., *Identification of Clusters from Data Sets, 3rd International Conference on Application of Information and Communication Technology and Statistics in Economy and Education*, Sofia, Bulgaria, 2013.
3. Bramer M., *Principles of Data Mining, Springer*, London, UK, 2007.
4. Hand D., Mannila H. & Smyth P., *Principles of Data Mining*, Massachusetts Institute of Technology, London, UK., 2001.
5. Gupta G. K., *An Introduction to Data Mining with Case Studies*, Monash University, Clayton, Australia, 2006.
6. Mufa V., *Information extraction with implementation of data mining methods*, master thesis, FAMIS – Bitola, Macedonia, 2013.

# Application of Quality Tools and Methods in Food Processing Industry

**Pedja Milosavljevic[1], Ivana Mladenovic-Ranisavljevic[2], Dragan Pavlovic[1], Milena Todorovic[1]**

[1] Faculty of Mechanical Engineering, University of Nis, Aleksandra Medvedeva 14., 18000 Nis, Serbia
pedja@masfak.ni.ac.rs,

[2] Faculty of Technology, University of Nis, Bulevar Oslobodjenja 124, 16000 Leskovac, Serbia
ivanamladenovic@tf.ni.ac.rs

**Abstract.** The goal of this paper is to present a possibility of application of quality tools, such as Pareto, Ishikawa and 5S, and methods like Lean and Six Sigma in food processing industry. In order to present application of these quality tools and methods, this paper shows the example of improvement of olives and peppers production process. Problems that occur in the production process have been found by application of quality tools, and certain improvements have been proposed based on the results analysis.

**Keywords.** Quality tools, quality methods, food processing, Lean, Six Sigma.

## 1.Introduction

To improve the production process of olives and peppers, it is necessary to find new things or reorganize the production process in order to minimize losses and maximize profits.

The aim of this paper is to identify the problems, which cause low quality of final products in the company "Ecofood" doo Nis, Serbia, by using quality tools, such as Pareto, Ishikawa and 5S, and to propose measures to eliminate these problems and increase quality of the products.

## 2.Processes that Olives and Chilli Peppers undergo

Figure 1 represents the process flow chart mapping of the mega production process of olives and peppers that consists of four activities:

- Procurement of raw materials;
- Storage of raw materials;
- Production and packaging;
- Product storage in the warehouse

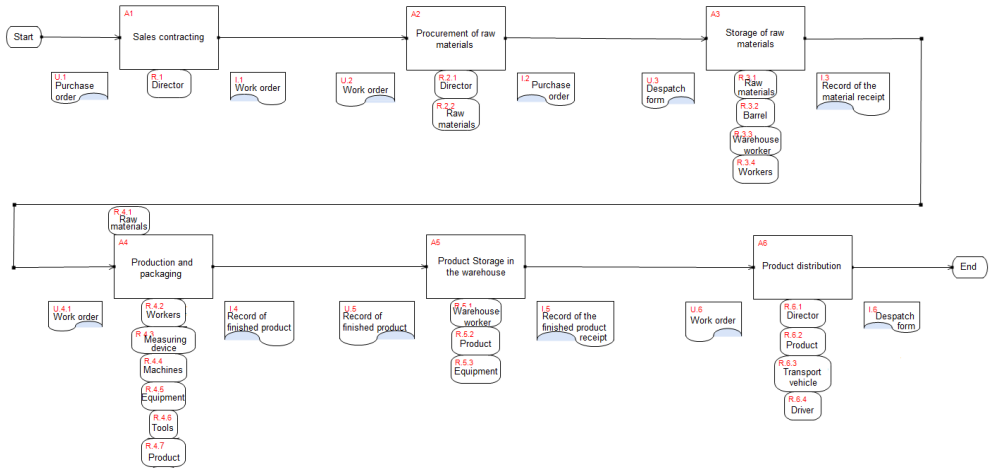



Figure 1. Process flow chart map of olives and peppers production

The process flow chart is a visual representation of the process that consists of all most important elements, parameters and their interactions. The process flow charts are used to easily identify the activities in the process [1, 2].

Chili peppers are grown in a field. After picking, they are stored in barrels and poured with liquid that consists of water, salt and acetic acid, after which they are transported to the plant. In the factory, workers put chili peppers into jars and pour the liquid again, whereupon closed jars are being sealed and sent to pasteurization. The process of pasteurization is done by having the jars put into the pasteurizer filled with water, after which the water is heated to the temperature of 86 ºC. After pasteurization, the product is sent to labeling, putting the production date on, and packing process. That way we obtain finished products, which are sent to the finished products warehouse, from where they are being transported to customers.

The production process of olives begins with the procurement of olives from the growers. The grower brings barrels of olives, which are stored in the raw material storage. There are a few different types of olives, such as: green olives with stone, green-pitted olives, black olives and olives stuffed with paprika. These olives are, after issuing the working order, carried to the manufacture department where they are put into jars. The jars are filled and closed and then sent to pasteurization. After pasteurization, they are labeled, dated and packed. The product, obtained this way, is sent to the finished products warehouse from where it undergoes further distribution.

## 3.Pareto Analysis

In order to identify the problems within the production process that affect the quality of chili papers and packed olives the most, the Pareto analysis is performed on a number of samples. Pareto analysis, named after Italian economist Vilfredo Pareto, was developed as a diagram method for grouping causes of problems according to their relative importance [3, 4]. It represents the process of selecting priority issues for solving, and it is used to focus on vital minority (20%), which leads to significant improvement (80%) [5].

Pareto analysis was performed by analyzing the problems that have arisen in the process of olive production from the beginning to the end of the process; from filling the jars

through pasteurization, labeling, packing and obtaining the final product. The analysis was performed on the specimen of five pasteurizers, which is the daily production norm.

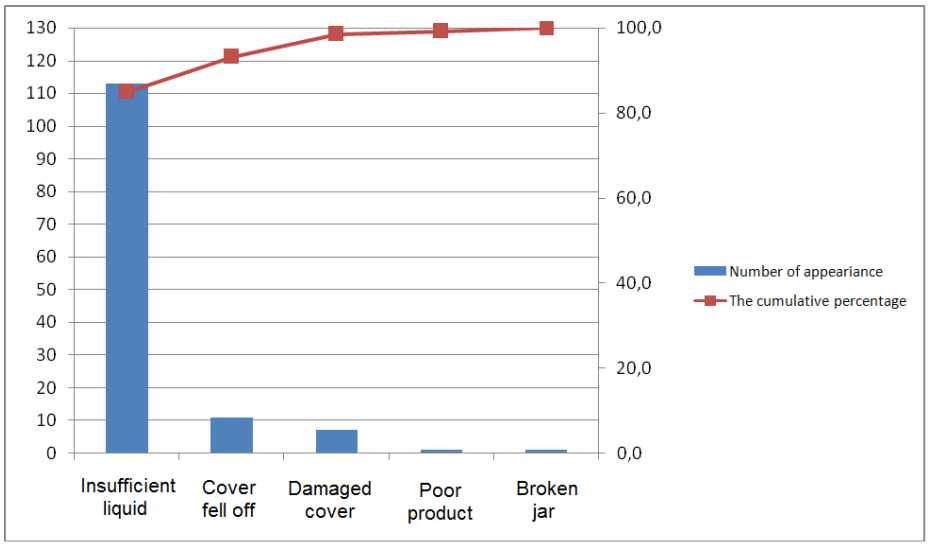



Figure 2. Pareto diagram of olives production

As it can be noticed, the vital minority is represented by jars with insufficient liquid, which represent even 85% of the problem. Furthermore, there are problems such as lids that have fallen off, that represent 8.2% of the problem, damaged lids, which represent 5.3% of the problem, and poor products as well as broken jars with a share of 0.8%.

Pareto analysis of problems that have arisen during the production of chili peppers was conducted in the same manner as in the case of olives.

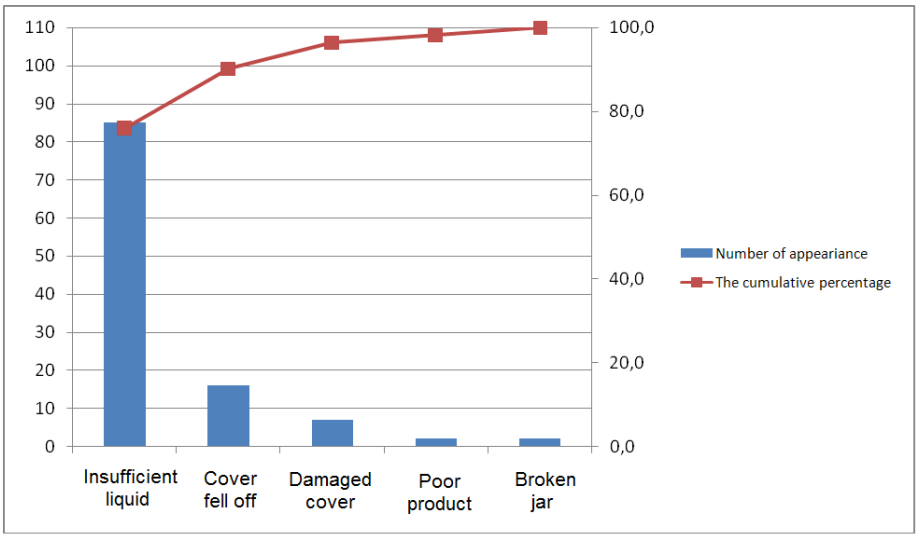



Figure 3. Pareto diagram of chilli pepper production

In the production of chili peppers, as well as in the production of olives, the vital minority is represented by jars with insufficient liquid, which in this case represent 75.9% of problems (figure 3). This is followed by damaged lids, which make 14.3% of problems, loose lids with 6.2% and in the end, poor products and broken jars with a share of 1.8%.

## 4.Ishikawa Analysis

Ishikawa diagram is a tool that helps identification, sorting, and displaying possible causes of a specific problem or quality characteristics. The diagram graphically shows the relation between specific consequence and all factors that influence the consequence [6].

The diagram is constructed by starting from the consequence, which is inserted in the right side of the diagram. Further step is drawing of the "fishbone" diagram and attaching categories to it. The diagram is developed by considering all categories and asking WHY a certain category affects the result. The same process of construction is repeated for every cause and the question WHY is asked for every level. It can be concluded that the method for constructing Ishikawa diagram consists of considering all possible causes and their organization into categories and subcategories.

The Ishikawa diagram analyzes the problems of insufficient liquid and scratched covers.

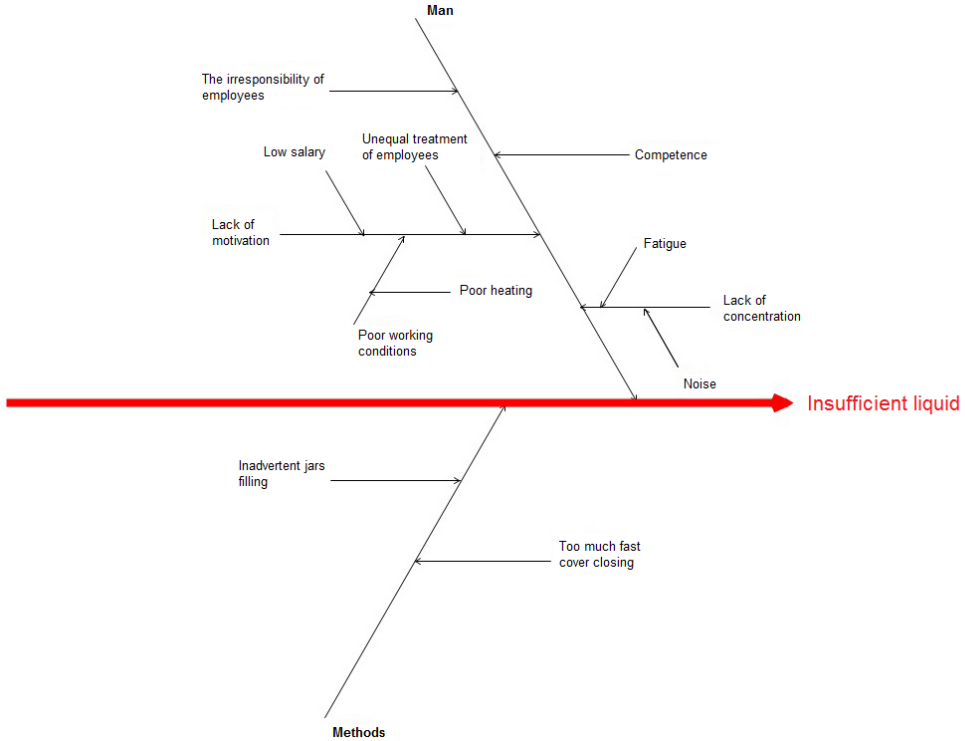



Figure 4. Ishikawa diagram of insufficient liquid occurrence

As it can be seen from the diagram (figure 4), the main cause of insufficient liquid is human factor or, more precisely, error of workers who fill the jars with liquid and close the lids. This can be reduced by influencing the workers to be more careful when they fill the

jars and close them, as well as by creating better working conditions for those workers so they would have better motivation and by reducing the causes that lead to lack of concentration.

This can be reduced by introducing a new method of pouring the olives over. The lids should not be put on the jars immediately after filling but a minute or two should be given for the olives to soak up and fill in the holes. Then, the liquid should be poured again and the jars should be closed. This would slow down the process of filling the jars for about 5 to 10 minutes per pasteurizer, which would in total slow down the whole process for 1 to 2 minutes, but it would significantly reduce the number of defects.

It is estimated that application of this method in the olive production would reduce the number of defects for up to 40-50 on a daily basis, which would, again on a daily basis, bring savings of about 40 €, and given that olive production roughly represents 50% of total production, this is about 800 € on monthly basis.

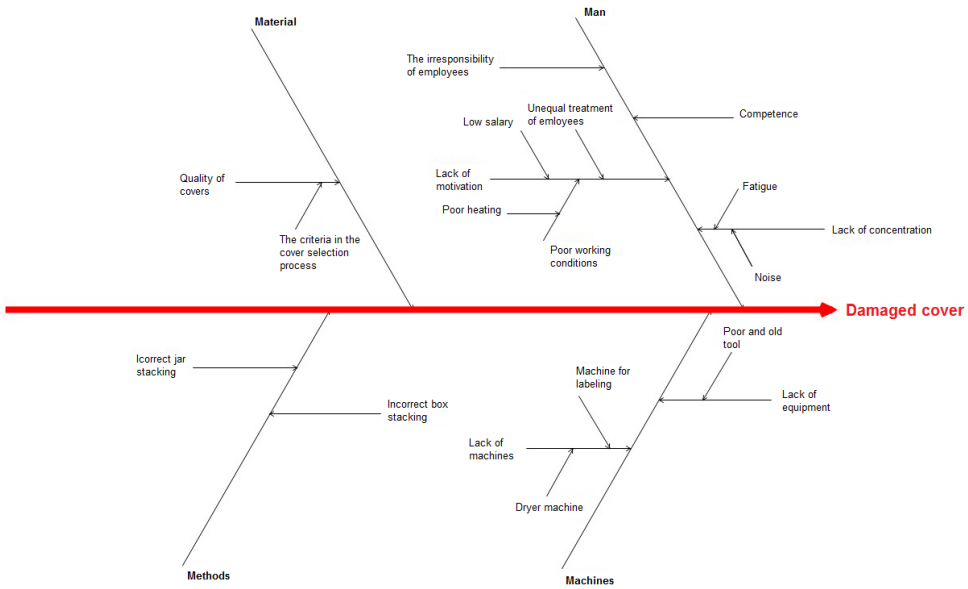



Figure 5. Ishikawa diagram of damaged lid occurrence

It can be seen that occurrence of damaged lid is influenced by several factors (figure 5). Human factor, as well as the occurrence of insufficient liquid, can be influenced by improvement of working conditions and increase in employee's motivation. Regarding material, this factor can be acted upon by choosing better quality lids. Regarding working methods, it is necessary to follow the rules when sorting the jars. They should not collate over each other and protective cards, which prevent scratching of the lids, should be used when putting crates one onto another.

Deeper analysis led us to the conclusion that most damage on the lids is done while discharging the Pasteurizer, sorting the jars onto the pallets to dry in a room for packaging and sorting during labeling. Therefore, the introduction of machines for drying and labeling would not only speed up production process, but would also have impact on drastic reduction in damaged lids, thus reducing scrap and making the process more efficient.

The same applies to broken jars, since this usually happens while sorting the jars onto the pallets in order to dry and while labelling.

Due to the reduction of these two cases, we can expect decrease of defects of 10-15 samples on a daily basis, which would lead to cost reduction of at least 8 € on a daily basis, which is over 200 € monthly.

## 5. 5S Method

5S represents a structural approach for development of working discipline. This is technique in 5 steps which is used to change the awareness of employees about the working environment. 5S helps to identify and eliminate useless things / waste in the working place. It helps to establish the productive, better working environment. This is a set of rules for the organization of workplace for each worker. The goal is to organize each workplace so as to be maximally effective and to expedite and facilitate work. 5S is certainly the most recognizable technique of Lean concept, because it is the easiest to implement and the results are visible almost immediately.

The term 5S represents an acronym of Japanese terms: Seiri – sorting, Seiton – neatness, Seiso – purity, Seiketsu – perfection and Shitsuke – discipline The term 5S has also become an acronym for expressions in English: Sort, Set in order, Shine, Standardize and Sustain.

In our case, application of 5S method has multiple significance. We get organized workplace, which increases the level of productivity.

The process of sorting in raw material storage house will significantly speed up the process of searching for raw materials, which can last for more than 15 minutes, thereby delaying the beginning of production until adequate raw material is found. After determining the raw material place allocation and barrels marking, whereby the properties of raw materials will be written on each that are therein, the process of searching for suitable raw material will be significantly shortened.

Items that are used should be given a permanent place that will be marked, and should be put back there after every usage. It is often the case to misplace items such as scissors, scalpels, tape, and glue. This will save time searching for them, and therefore speed up the production process.

Figures 6.1-6.4 show examples of improvements obtained by applying the 5S method.

Figure 6.1 – Raw materials' warehouse before improvements

Figure 6.2 – Raw materials' warehouse after improvements

Figure 6.3 – Packaging warehouse before improvements

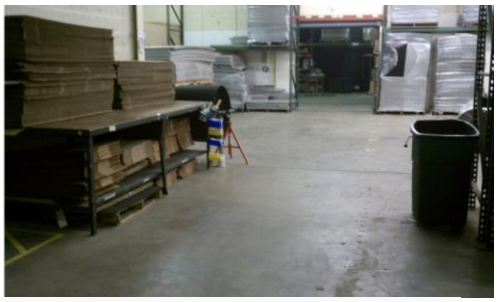


Figure 6.4 – Packaging warehouse after improvements

## 6.Suggested improvements

In this paper, many improvements are suggested by application of management tools. Enormous wastage – activities which last long and consume resources without adding value, were found in the production process. These activities are transportation of products after drying and pasteurization and manual labeling. To avoid waste and shorten the process time, improvements were introduced in the form of procurement of systems for drying and machines for labeling. This enables the increase of production by 9%, or savings of about 315 € on monthly basis, unless there is a need to increase production.

The vital minority of problems within production process is found by applying the Pareto analysis. The problem was insufficient liquid in jars after pasteurization. The problem was analyzed by application of Ishikawa method and the conclusion was that the main cause of this problem was human error, carelessness or haste of workers who were filling the jars and closing the lids. Improvement of working conditions and motivation of workers, as well as drawing attention to the problem, should reduce this occurrence. In addition, there was a new method of pouring liquid over olives, which would slow down the process of production for 1 to 2 minutes per pasteurizer, but would significantly reduce the occurrence of insufficient of liquid. With this improvement it is expected to save approximately 430 € on monthly basis.

Ishikawa method is also used for occurrence of damaged lids, but by detailed analysis it was concluded that this occurs mainly during sorting the jars for drying and during labelling and with the proposal of process improvements this activities are eliminated, thus reducing them to a negligible minimum. This should bring savings of at least 210 € on monthly basis.

Table 10.1 – Savings achieved by process improvement

| **Improvement** | **Savings on a monthly basis** |
|---|---|
| Speeding up the process by introducing machines for drying and labeling | 315 € |
| Elimination of damaged lids | 210 € |
| Implementation of a new method for pouring liquid over olives | 430 € |
| **Total** | **955 €** |

Application of 5S method provided a suggestion for improving the organization within the warehouse where raw materials would get their permanent place and each barrel would be marked. Thus, the process of searching for suitable raw materials would be faster as well as the entire production process. Application of this method also proposes determining and marking a permanent place for additional objects that are usually lost.

## 7.Conclusion

By applying the quality tools, it is possible to increase the efficiency and speed up the production process. This paper presents the application of quality tools to identify major problems in the process of production of olives and chilli peppers in the company "Ecofood" from Nis, Serbia, in order to increase quality of final products. Based on all above presented, it can be concluded that the production procedure is not good enough, and it has shortcomings which directly result in the reduced quality of finished product. Main problems that occur are insufficient fluid and damaged jar lids.

By eliminating these two problems, which make 80% of the problems, proposed measures for improvement of production process would increase the quality of products.

Based on the above, it can be concluded that the quality tools can be applied in food industry.

## References

1. Pavlovic, D., P. Milosavljevic, S. Mladenovic, *Application of Lean Six Sigma Method in education process*, Sixth International Working Conference – Total Quality Management-Advanced and Intelligent Approaches, Belgrade, Serbia, pp. 538-543, 2011.
2. Milosavljevic, P., M. Manojlovic, A. Berket Bakota, M. Markovic, *Improvement of Patient Admission Process of in Health Institution Health Care Center Nis*, International Journal Advanced Quality, vol. 40, no.1, pp. 17-22, 2012.
3. Stoiljkovic, V., Z. Jankovic, I. Markovic, *Application of Pareto Analysis in Pharmaceutical Institutions*, International Journal 'Total Quality Management & Excellence', vol. 38, no. 3, pp. 93-97, 2010.
4. Seferovis, E., D. Cengic, *Alati Za Upravljanje Kvalitetom u Primjeni*, Masinstvo, vol. 4, no. 2, pp. 193-204, 1998.
5. Hung, H. C., M. H. Sung, *Applying Six Sigma to Manufacturing Processes in the Food Industry to Reduce Quality Cost*, Scientific Research and Essays, vol. 6, no. 3, pp. 580-591, 2011.
6. K. Ishikawa, *Guide to Quality Control*, Asian Productivity Organization, Tokyo, 1982.

# Random and Exhaustive Generation of Combinatorial Sets with Special Properties

**Igor Grebennik[1], Oleksandr Lytvynenko[2]**

[1] Chair of Systems Engineering of the Kharkiv National University of Radio Electronics
grebennik@kture.kharkov.ua
[2] Kharkiv National University of Radio Electronics
litvinenko1706@gmail.com

**Abstract.** A general approach to solving the problem of random generation of *k*-sets has been proposed. This approach is based on a unified method of random generation of various basic combinatorial sets. The algorithms of generation of *k*-sets have been described; their complexity has been evaluated with examples analyzed.

**Keywords.** Combinatorial generation, k-set, basic combinatorial set, random generation, complexity.

## 1. Introduction

Generation of various combinatorial objects is frequently required in developing and implementing methods and algorithms for solving many scientific and applied problems [1-6]. Generation is usually understood as a construction of all combinatorial structures of a given type [3]. In these sources, the problem of generation of simple combinatorial objects like permutations, combinations, splits, trees, binary sequences is mainly solved. Solution of generating more complex combinatorial objects is hindered by lack of special structural tools and by significant computational costs caused by redundant results of using well-known methods and algorithms of generation.

Quite complex combinatorial configurations can be formally described and generated with the help of structural tools of description of compositional $k-$images of combinatorial sets ($k-$sets), proposed in [7]. A combinatorial set is understood as a variety of tuples constructed from a finite set of arbitrary elements (generating elements) in accordance with certain rules [7, 8]. Permutations, combinations, arrangements, binary sequences etc. can serve as examples of classical combinatorial sets. The apparatus of $k-$sets has been widely explored [7–9]: general concepts of their generation are considered in [8], the task of exhaustive generation of $k-$sets was solved in [9], and some special cases of this task were studied in [10]. However, the task of random generation of $k-$sets hasn't been solved yet.

Both random and exhaustive generation of $k-$sets require solving the problem of generation of basic combinatorial sets used to construct $k-$sets. Basic sets can be combinatorial ones with known descriptions and algorithms of generation: either classical combinatorial sets (permutations, combinations etc.) or non-classical, e.g., permutations of tuples, compositions of permutations, permutations with given number of cycles etc. [7-10]. Algorithms of generation of many basic combinatorial sets have been described [1-3, 5].

However, in most cases, each generation algorithm is based on specific properties of combinatorial sets.

In this paper, a general approach to solving the problem of random generation of *k*-sets has been proposed.

## 2. Composition *k*-images of combinatorial sets (k-sets)

Let $z^{\beta} = \{z_1^{\beta}, z_2^{\beta}, ..., z_{n_\beta}^{\beta}\} \subseteq \mathbf{Z}_{\boldsymbol{\beta}_i}$, where $\mathbf{Z}_{\boldsymbol{\beta}_i}$ are sets of arbitrary elements, $\beta \in \boldsymbol{\beta}_i$, $i \in J_k^0 = \{0,1,...,k\}$, where $\boldsymbol{\beta}_0 = \{0\}$,

$$\boldsymbol{\beta}_i = \{\beta_j, j = 1,2...,\eta_i\},\ \beta_j = (\alpha_1, \alpha_2, ..., \alpha_i),\ \alpha_1 \in J_n,\ \alpha_2 \in J_{n_{\alpha_1}}, ..., \alpha_i \in J_{n_{\alpha_1...\alpha_{i-1}}},$$

$$\eta_1 = n,\ \eta_2 = \sum_{j=1}^{n} n_j,\ \eta_i = \sum_{\alpha_1=1}^{n}\sum_{\alpha_2=1}^{n_1}...\sum_{\alpha_{i-1}=1}^{n_{1...i-2}} n_{\alpha_1...\alpha_{i-1}},\ i = 3,4,...,k, \tag{1}$$

Let us consider mappings [7, 8]:

$$\Gamma_{\boldsymbol{\beta}_0}: \mathbf{Z}_{\boldsymbol{\beta}_0} \to \mathbf{Y}^0,\ \Gamma_{\boldsymbol{\beta}_i}: \mathbf{Y}^{i-1} \times \mathbf{Z}_{\boldsymbol{\beta}_i} \to \mathbf{Y}^i,$$

where

$$\mathbf{Y}^0 = \{Y_{\boldsymbol{\beta}_0}(z^{\beta}), \beta \in \boldsymbol{\beta}_0\},\ \mathbf{Y}^i = \{Y_{\boldsymbol{\beta}_i}(Y_{\boldsymbol{\beta}_{i-1}}, z^{\beta}), \beta \in \boldsymbol{\beta}_i\},\ i \in J_k,\ J_t = \{1,2,...,t\},$$

$Y_{\boldsymbol{\beta}_i} = \Gamma_{\boldsymbol{\beta}_i}(Y_{\boldsymbol{\beta}_{i-1}}, \mathrm{z}^{\boldsymbol{\beta}_i}) = F(Y_{\boldsymbol{\beta}_{i-1}}, \tilde{\Gamma}_{\boldsymbol{\beta}_i}(\mathrm{z}^{\boldsymbol{\beta}_i}))$, $i \in J_k$, $F(Y_{\boldsymbol{\beta}_{i-1}}, \tilde{\Gamma}_{\boldsymbol{\beta}_i}(\mathrm{z}^{\boldsymbol{\beta}_i}))$ – a mapping that realizes $n$ - composition which consists in replacing each generative element of set $Y_{\boldsymbol{\beta}_{i-1}}$ with elements of primary combinatorial sets $Y_{\beta} = \tilde{\Gamma}_{\beta}(\mathrm{z}^{\beta}),\ \beta \in \boldsymbol{\beta}_i$, respectively,

$\tilde{\Gamma}_{\boldsymbol{\beta}_i}(\mathrm{z}^{\boldsymbol{\beta}_i}) = (\tilde{\Gamma}_{\beta}(\mathrm{z}^{\beta}),\ \beta \in \boldsymbol{\beta}_i)$, $\mathrm{z}^{\boldsymbol{\beta}_i} = (\mathrm{z}^{\beta},\ \beta \in \boldsymbol{\beta}_i)$, $\tilde{\Gamma}_{\beta}(\mathrm{z}^{\beta})$ are the primary mappings [7,8]. It means that

$$(z_{l_1}^{\beta}, z_{l_2}^{\beta}, ..., z_{l_\beta}^{\beta}) \in Y_{\beta},\ z_{l_t}^{\beta} \in Y_{\delta},\ l_t \in J_{l_\beta},\ \beta \in \boldsymbol{\beta}_{i-1},\ \delta \in \boldsymbol{\beta}_i,\ i \in J_k.$$

Let us denote

$$\boldsymbol{\Gamma}_i = \{\Gamma_{\boldsymbol{\beta}_i}\},\ i \in J_k. \tag{2}$$

*Definition*. Composition *k*-image of combinatorial sets $Y_0$, $Y_1, Y_2, \ldots, Y_n$, $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$, …, $Y_{\underbrace{1...1}_{k}}, \ldots, Y_{nn_1...n_{k-1}}$ ($k-$set) generated by sets $z^{\beta_k}$, $\beta_k \in \boldsymbol{\beta}_k$ is the combinatorial set [7,8]:

$$W_z = \Gamma_k \circ \Gamma_{k-1} \circ ... \circ \Gamma_0(z), \tag{3}$$

where mappings $\Gamma_i \in \boldsymbol{\Gamma}_i$, $i \in J_k$ are determined by (2).

Cardinality of the set (3) can be obtained as [7, 8]:

$$Card(W_z) = \sum_{\{\gamma_1,\gamma_2,...,\gamma_r\} \subset J_n} \alpha_{\gamma_1,\gamma_2,...,\gamma_r} \cdot \prod_{i=1}^{k} \prod_{\beta_i = (\alpha_1\alpha_2...\alpha_i)} Card\ Y_{\beta_i} \tag{4}$$

As the generation of $k-$sets is based on the generation of basic combinatorial sets, we need an algorithm for generating these sets.

## 3. Generation of basic combinatorial sets

Let us associate the set $p(T)=\{A,m,S\}$ with each basic combinatorial set $T$, where $A=\{a_1,a_2,...,a_n\}$, $a_1<a_2<...<a_n$ is a set of generating elements, $m$ is the length of tuple $t\in T$ (let us consider that all tuples in the set are of the same length), $S$ is a set of parameters describing the set $T$ (e.g., parameters $n_1,n_2,...,n_k$ for permutations with repetitions and other parameters specific for different classes of combinatorial sets). We understand a class of combinatorial sets as its belonging to permutations, combinations, etc.

Let the basic set $T$ and its parameters $p(T)$ be defined. We need to generate all elements $t\in T$, where each element is a tuple of the length $m$. Let us denote $t^i=(t_1,t_2,...,t_i)$, $\forall t_i\in A$, $A\in p(T)$, $i\in J_n$. It means that $t^0=(\ )$ is an empty tuple and $t^m=t\in T$.

Firstly, let us recall the main concept of the algorithm of random generation of basic sets [9]. This algorithm is of a recursive nature: at each recursion level $i\in J_{m-1}^0$ it expands current tuple $t^i=(t_1,t_2,...,t_i)$ by adding the following element $t_{i+1}$ and thus obtaining a tuple $t^{i+1}=(t_1,t_2,...,t_{i+1})$ at level $i+1$. At level $m\le n$ the algorithm adds tuple $t^m=t$ to the resulting set *T*.

The fact that set *T* belongs to a certain class of combinatorial sets imposes some restrictions on element $t_{i+1}\in A$. At each level $i\in J_{m-1}^0$ let $F^i=\{f_1,f_2,...,f_k\}\subseteq A$ denote a set of all generating elements that satisfy these restrictions. In this case, for each $j\in J_k$, the algorithm adds element $t_{i+1}=f_j$ to the "input" tuple $t^i=(t_1,t_2,...,t_i)$ and calls itself recursively with $t^{i+1}=(t_1,t_2,...,f_j)$ as an input.

The described algorithm can generate elements of set *T* randomly: for that we just need to do a recursive algorithm call at level $i$ not for all $j\in J_k$, but only for $q_i$ % of them (selected randomly). It should be noted that we can use various mechanisms of random selection (in this work we use the simplest way of selection: we just generate a uniformly distributed random value; though other distributions can also be used).

Let us consider specific features of the construction of set $F^i$ for some classes of combinatorial sets. In the case of arrangements with repetitions, set $F^i$ consists of all generating elements:

$$F^i=A.$$

For arrangements without repetitions and permutations as their special case, set $F^i$ consists of $n-i$ generating elements that have not appeared in $t^i$:

$$F^i = \{f_1, f_2, ..., f_{n-i}\} \subseteq A: f_l \neq t_j, \forall j \in J_i, \forall l \in J_{n-i}.$$

As for the combinations the order of the items is not important, let us generate them in the form of ordered sets where $t_1 < t_2 < ... < t_i$ is for combinations without repetitions and $t_1 \leq t_2 \leq ... \leq t_i$ is for combinations with repetitions.

For combinations without repetitions, set $F^i$ includes all generating elements that have not been included into $t^i$ and are greater than $t_i$:

$$F^i = \{f_1, f_2, ..., f_k\} \subseteq A: f_l \neq t_j, f_l > t_i, \forall l \in J_k, \forall j \in J_{i-1}.$$

For combinations with repetitions, set $F^i$ also includes generating element which is equal to $t_i$:

$$F^i = \{f_1, f_2, ..., f_k\} \subseteq A: f_l \neq t_j, f_l \geq t_i, \forall l \in J_k, \forall j \in J_{i-1}.$$

Let us describe the algorithm **GenBase** that implements the steps described. **GenBase** input data consist of $T$ - type of a basic set, a set of parameters $p(T)$ and tuple $t^i$. At each recursion level $i \in J_{m-1}^0$, the algorithm builds the set $F^i$, then randomly adds selected element $F^i$ to tuple $t^i$ as many as $v_i$ times and recursively calls itself. At each level, the number of recursive calls is determined by

$$v_i = \frac{|F^i| \cdot q_i}{100}.$$

To generate all the elements of set $T$, **GenBase** should be called with parameters $T$, ( ), $p(T)$.

Here $random(1, |F^i|)$ is a uniformly distributed random number between 1 and $|F^i|$. In order to generate combinatorial sets of other classes with this algorithm, it is sufficient to identify laws of constructing set $F^i$.

**function GenBase**($T$, $t^i$, $p(T)$);
**local** $F^i$;
**if** $i = m$ **then** $T := T \cup t^i$ ; **exit**;
**end if;**
**case** $T$ **of**
$\overline{A}_n^m : F^i = A;$
$A_n^m : F^i = \{f_1, f_2, ..., f_{n-i}\} \subseteq A: f_l \neq t_j, \forall j \in J_i, \forall l \in J_{n-i};$

$\overline{C}_n^m : F^i = \{f_1, f_2, ..., f_k\} \subseteq A: f_l \neq t_j, f_l \geq t_i, \forall l \in J_k, \forall j \in J_{i-1}$;

$C_n^m : F^i = \{f_1, f_2, ..., f_k\} \subseteq A: f_l \neq t_j, f_l > t_i, \forall l \in J_k, \forall j \in J_{i-1}$;

$T_n$ : **exit;**

**end case**;

$$v_i = \frac{|F^i| \cdot q_i}{100};$$

**for** $j = 1, 2, ..., v_i$ **do**

**GenBase**( $t^{i+1} = (t_1, t_2, ..., t_i, f_{random(1,|F^i|)})$ );

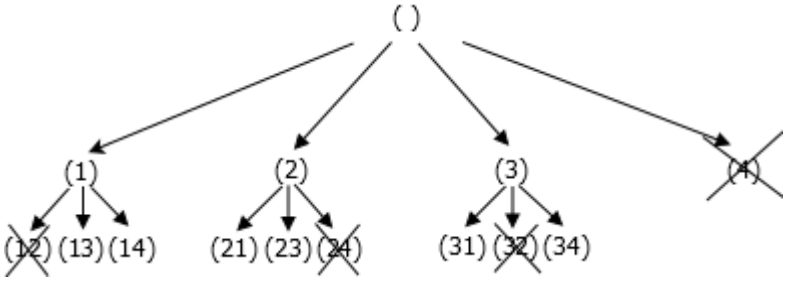
**end for;**

**end function;**

It should be noted that, to generate combinatorial sets of other classes with this algorithm, it is sufficient to identify laws of constructing set $F^i$.

**Example 1.** Let the task is to randomly generate arrangements of 4 to 2. Let $q_0 = 75$, $q_1 = 60$. At zero level, set $F^0$ includes all generating elements: $F^0 = \{1, 2, 3, 4\}$. Hence, $v_0 = \frac{4 \cdot 75}{100} = 3$. The algorithm randomly selects three elements $F^0$ (let them be 3, 2 and 4) and makes recursive calls of itself with $t^1 = (3)$, $t^1 = (2)$, $t^1 = (4)$ respectively.

Let us consider the situation at level 1 for $t^1 = (3)$. In this case, set $F^1$ includes all elements that have not been not included into $t^1$, i.e. $F^1 = \{1, 2, 4\}$. Hence, $v_1 = \frac{3 \cdot 60}{100} \approx 2$. The algorithm randomly selects two elements $F^1$ (let them be 1 and 4); it means that, at level 2, full arrangements (31) and (34) are obtained. A possible recursion tree of the algorithm operation is provided below.



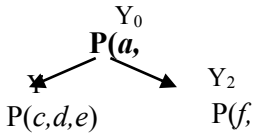
## 4. Generation of *k*-sets

First, let us briefly describe the algorithm of exhaustive generating *k*-sets [9]. At the beginning, it generates the elements of each basic set using **GenBase** algorithm (values of are individual for each basic set). After that, the algorithm sequentially implements mappings $\Gamma_{i+1} \circ \Gamma_i \circ .. \circ \Gamma_0(z)$, $z = A_0 \in p(Y_0)$ for each $i \in J_{k-1}^0$. In other words, it implements $n$-composition, where generating elements of a parent set are replaced by the elements-tuples of its child sets.

Algorithm Get_k-set in its original version in [9], while implementing an operation of *n*-composition, performs a sequential substitution of the element of a parent set by each element of a child set. The proposed algorithm Get_Random_k-set does it not for all, but for some elements of a child set. By analogy with the generation of basic sets, we can put the parameter *Q(Y)* for each basic set *Y* at level $i \in J_k$, which defines the fraction of elements participating in the operation of $n$-composition. The number of the elements is determined by

$$V(Y) = \frac{|Y| \cdot Q(Y)}{100}.$$

**Example 2**. Let us describe the generation of the composition of permutations by means of **Get_Random_k-set**. The structure of the corresponding *k*-set can be presented as follows




The elements "*a*" and "*b*" of set $Y_0 = \{(ab),(ba)\}$ are replaced by the elements of child sets $Y_1 = \{(cde),(ced),(dce),(dec),(ecd),(edc)\}$ and $Y_2 = \{(fg),(gf)\}$ respectively. Let $Q(Y_1) = Q(Y_2) = 50$. Hence, $V(Y_1) = 3$; $V(Y_2) = 1$, i.e. element "*a*" is sequentially replaced by three random elements of $Y_1$, and "*b*" – by one random element of $Y_2$.

A possible results of the generation is $W_z = \{\underbrace{(cdefg),(dcefg),(ecdfg)}_{result\ of\ replacing\ (ab)}, \underbrace{(gfced),(gfdec),(gfedc)}_{result\ of\ replacing\ (ba)}\}$.

## 5. The evaluation of the complexity of algorithm Get_Random_k-set

The complexity of the algorithm is determined by the complexity of generation of basic sets, as well as the complexity of operations of *n*-substitution and a number of levels of a certain *k*-set. In [9], the evaluation of the complexity of the «full» algorithm **Gen_k-set** is described. This formula can also be used in this case. It is sufficient to replace cardinality *Card(Y)* of each basic set by *V(Y)*. Then, the formula to evaluate **Get_Random_k-set** complexity is

$$\sum_{i=0}^{k}\sum_{j=1}^{\eta_i} O(Y_{ij}) + \sum_{i=0}^{k-1}(Card(P^i)\prod_{j=1}^{\eta_{i+1}} V(Y_{(i+1)j})),$$

where $O(Y_{ij})$ is the complexity of generating a basic set $Y_{ij}$ (*i* is the level of the basic set in *k*-set, *j* is the sequence number of the set at level *i*);

$P^i = \Gamma_i \circ \Gamma_{i-1} \circ ... \circ \Gamma_0(z)$ – "intermediate" *k*-set, which is the parent set at level *i*.

Details about designations and obtaining this formula are given in [9].

## 6. Conclusion

We described the general approach for solving the problem of random generation of $k-$sets based on the single approach to random generation of various basic combinatorial sets.

Algorithm for random generation of basic sets allows generating various combinatorial sets, and laws of constructing sets $F^i$ can be pre-set. If identification of the laws fails, the algorithm allows using other algorithms to generate basic sets.

The described approach to the random generation is very flexible since it allows obtaining various results by varying parameters $q_i$ and *Q(Y)*. In its turn, this allows adjusting the number of elements for both basic sets and *k*-sets.

The developed software allows solving the described problems of generation of *k* -sets and basic sets.

# References


1. Donald Knuth The Art of Computer Programming, Volume 4, Fascicle 2: Generating All Tuples and Permutations, Addison-Wesley, 2005. – 144 p.
2. Donald Knuth The Art of Computer Programming, Volume 4, Fascicle 3: Generating All Combinations and Partitions, Addison-Wesley, 2005.–160 p.
3. Donald L.Kreher, Douglas R.Stinson, Combinatorial Algorithms: Generation Enumeration and Search, CRC Press, 1999.– 329 p.
4. M. Bona Combinatorics of Permutations, Chapman Hall-CRC, 2004.– 383 p.
5. F. Ruskey (2003) Combinatorial generation, Dept. of Computer Science Univ. of Victoria, Canada, 1j-CSC 425/520.– 289 p.
6. Korsh J.F., LaFollette P. S. Loopless array generation of multiset permutations // The Computer Journal. – Vol. 47, № 5. – 2004. – P. 612 –621.
7. Yu.G. Stoyan, I.V. Grebennik Description of Classes of Combinatorial Configurations by Mappings // Proc. of NAS of Ukraine, 2008.– 10: p. 28-31 (in Russian)
8. Yu.G. Stoyan, I.V. Grebennik Description and Generation of Combinatorial Sets Having Special Characteristics // International Journal of Biomedical Soft Computing and Human Sciences, Special Volume "Bilevel Programming, Optimization Methods, and Applications to Economics" Vol. 18, №1, 2011. – P. 85-90.
9. I. V. Grebennik, O. S. Lytvynenko Generating combinatorial sets with given properties // Cybernetics and Systems Analysis, 48(6), 2012.– P. 890–898.
10. I. V. Grebennik Description and generation of permutations containing cycles // Cybernetics and Systems Analysis, 46(6), 2010.– P. 945-952.

# Elements Transpositions and Their Impact on the Cyclic Structure of Permutations

**Igor Grebennik[1], Olga Chorna[2]**

Kharkiv National University of Radioelectronics, 14 Lenin ave, 61166 Kharkiv Ukraine

[1] igorgrebennik@gmail.com

[2] titovaolga90@gmail.com

**Abstract.** In this paper we consider transpositions of elements in a permutation, the transpositions corresponding to the adjacency criterion in a permutation polyhedron. Six permutation types are considered in correspondence with the location of arbitrary components. We consider the impact of the corresponding components on the cyclic structure of permutations depending on the type of a permutation.

**Keywords.** Permutations set, permutation polyhedron, adjacency criterion, permutation properties, transposition.

## 1.Introduction

Permutations sets are very often considered in theoretical and applied research in the field of combinatorics and combinatorial optimization [1–10]. By now many properties of permutations have been investigated, in particular those associated with the cyclic structure of permutations. Some methods and algorithms allowing the representation of permutations as the product of cycles and the generation of permutations having a predefined cyclic structure are known [1, 4, 6-9].

A well-known way for investigating combinatorial sets is their immersion into the Euclidian space, which allows using tools of continuous mathematics when analyzing combinatorial problems [2, 5]. The convex hull of permutations set immersed into the Euclidian space is a permutation polyhedron [3]. One of the basic properties of this polyhedron is the fact that the adjacency criterion for its vertexes is satisfied [1-3].

The geometric and analytic interpretations of one transposition corresponding to the permutation adjacency criterion are well known and were investigated earlier [2, 3, 10, 11].

This research is devoted to the investigation of two consecutive transpositions of neighboring by value generative elements of a permutation and their impact on the cyclic structure of the permutation being considered, and also their relative location on the permutation polyhedron.

In this paper we introduce a classification of permutations in dependence of some components relative location and the impact of these components transpositions on the permutation structure.

**The aim** of this paper is the investigation of the cyclic structure and permutation properties based on neighbor elements transposition properties and the properties of the permutation polyhedron.

## 2.Basic definitions

Let $P_n^C$ be the set of cyclic permutations without repetition from $n$ real numbers $a_1 < a_2 < ... < a_n$: $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n) \in P_n^C$, where $c_i \in R$, $i \in J_n = \{1, 2, ..., n\}$ [4, 9]. Consider the notion of permutation in detail [9].

*Definition*. A linear ordering of the elements from a certain generating set $A = \{a_1, a_2, ..., a_n\}$ is called a permutation $\pi = \pi(a_1, a_2, ..., a_n) = (\pi(a_1), \pi(a_2), ..., \pi(a_n)) = (a_{i_1}, a_{i_2}, ..., a_{i_n})$ or, if it is necessary to stress the fact that it contains $n$ elements, $n$-permutation.

We denote as $P_n$ the set of permutations generated by the elements $a_1 < a_2 < ... < a_n$.

Consider a certain permutation $\pi = (\pi(a_1), \pi(a_2), ..., \pi(a_n)) \in P_n$ and its element $\pi(a_i) = a_j$, $\forall i, j \in J_n$. Then we can write down: $\pi(a_j) = \pi(\pi(a_i)) = \pi^2(a_i)$. Generally this formula can be represented in the following form: $\pi^{k-1}(a_j) = \pi(\pi^{k-1}(a_i)) = \pi^k(a_i)$, $\forall i, j \in J_n$, $k \le n$.

Thus [2] if for some $l \ge 1$ we have $\pi^l(a_i) = a_i$, $i \in J_n$ and all the elements $a_i, \pi(a_i), \pi^2(a_i), ..., \pi^{l-1}(a_i)$ are different, the sequence $(a_i, \pi(a_i), \pi^2(a_i), ..., \pi^{l-1}(a_i))$ is called an $l$ length cycle.

*Definition* [2]. A cyclic permutation is such a permutation $\pi$ from $n$ elements that contains a single $n$ length cycle, i. e. $\pi^n(a_i) = a_i$, $\forall i \in J_n$. We denote such permutations as $\pi_C$.

Note that $Card\ P_n^C = (n-1)!$ [4].

Let us consider a cyclic permutation example. In this work we will use the following notation way. We record two rows so that the generating elements are written not in the increasing order but in the order of their appearance in the cycle:

$$\begin{pmatrix} 1 & 4 & 6 & 2 & 5 & 8 & 3 & 7 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & 6 & 2 & 5 & 8 & 3 & 7 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 6 & 2 & 5 & 8 & 3 & 7 \\ 4 & 6 & 2 & 5 & 8 & 3 & 7 & 1 \end{pmatrix}.$$

One of the widespread directions of combinatorial research is the immersion (enclosure mapping) of a combinatorial set into the Euclidian space [5].

The immersion of combinatorial sets in the Euclidian space allows constructing combinatorial polyhedrons [1] with the help of which it is possible to investigate the properties of permutations sets classes in the Euclidian space.

Let us fulfill the enclosure mapping of the permutations set $P_n$ and cyclic permutations $P_n^C$ to the arithmetic Euclidian space $R^n$. According to [2, 5] the given mapping (which is called immersion) can be represented in the following form:

$$f : P \to R^n,\ \forall p = (p_1, p_2, ..., p_n) \in P,$$

$$x = f(p) = (x_1, x_2, ..., x_n) \in E \subset R^n,\ x_i = p_i,\ i \in J_n.$$

As a result of the immersion $f$ we have one-to-one correspondence between each set $P_n$, $P_n^C$ and $E \subset R^n$: $E_n = f(P_n)$, $E_n^C = f(P_n^C)$.

Because of the above operations it becomes possible to investigate properties of the elements of the subset $P_n^C \subset P_n$ with the help of the polyhedron $\Pi_n$.

## 3.Relative location of some permutation components

Consider a permutation polyhedron $\Pi_n$ generated by a set $a_1 < a_2 < ... < a_n$, $vert\ \Pi_n = E_n$ is the set of its vertexes.

Since any cyclic permutation belongs to the set of permutations $P_n$, $\pi_C = (\pi(a_1), \pi(a_2), ..., \pi(a_n)) \in P_n$ all cyclic permutations are vertexes of the permutation polyhedron $\Pi_n$.

Let us introduce $\Pi_n^C \subset vert\ \Pi_n$, which is the subset of permutation polyhedron vertexes corresponding to all possible cyclic permutations with the cycle of $n$ elements generated by the set $a_1 < a_2 < ... < a_n$, $\Pi_n^C = E_n^C$.

Let us denote $v^C$ the vertex $v \in vert\ \Pi_n$ corresponding to a certain cyclic permutation $p \in P_n^C$ i. e. $v^C \in \Pi_n^C \subset vert\ \Pi_n$.

The adjacency criterion for the vertexes of the permutation polyhedron deals with the elements of the generating set $A = \{a_1, a_2, ..., a_n\}$ and their location in the permutation, i. e.: a vertex adjacent to the vertex $v = (a_{v_1}, a_{v_2}, a_{v_3}, ..., a_{v_n})$ that corresponds to the permutation $p \in P_n^C$ is any vertex corresponding to the permutation $p_k$ obtained from $p$ by the transposition of components equal to $k$ and $k+1$, $\forall k \in J_{n-1}$. And the two permutations $p_1, p_2 \in P_n$, corresponding to the vertexes $v_1, v_2 \in \Pi_n$ are called adjacent permutations if the vertexes $v_1, v_2$ are adjacent vertexes of the polyhedron $\Pi_n$ [11].

Further in this paper, without losing generality we suppose $A = \{a_1, a_2, ..., a_n\} = \{1, 2, ..., n\}$.

Let us consider the location of the components equal to $i, i+1$ and $j, j+1$, $i, j \in J_{n-1}, j \neq i$ in an arbitrary cyclic permutation $p \in P_n^C$ and write them down in the form of chains. The components can be located in a chain in six ways that we call types. Let us fix the first two elements of the chain. This is a certain component $x \in A = \{a_1, a_2, ..., a_n\}$ the value of which is arbitrary and the component $i$ such that $\pi(x) = i$. Thus the beginning of a chain always looks as follows: $(x, i, \_, \_, \_)$, and further there are three positions in which the components $i+1, j, j+1$ can be located in a different order. It is this order that sets the type of a cyclic permutation for components $i, j$. The number of ordering ways for 3 components is equal to $3! = 6$, which is the number of permutations from 3 elements.

Let us write down all the six permutations from the elements $i+1, j, j+1$:

1. $j, i+1, j+1$;
2. $j+1, i+1, j$;
3. $i+1, j, j+1$;
4. $j+1, j, i+1$;
5. $i+1, j+1, j$;

6. $j, j+1, i+1$.

Consider the elements chains corresponding to the given sequences.

Since we consider the location of components in an arbitrary cyclic permutation $p \in P_n^C$ it is not important which component will be the beginning of a chain because in a cyclic permutation any component can be obtained from any other component by the number of steps $\leq n-1$. Further in this paper, without losing generality, we will always start chains with a fixed component $i$ and an arbitrary component $x$ the value of which is inessential but the component $i$ is a mapping of $x$.

Thus using the above considerations let us show all the six location types for the components $i, i+1$ and $j, j+1$, $i, j \in J_{n-1}, j \neq i$ in an arbitrary cyclic permutation $p \in P_n^C$:

Type I: $\begin{pmatrix} x & \pi^a(i) & \pi^b(j) & \pi^c(i+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j & i+1 & j+1 \end{pmatrix}$, $a, b, c \in \{0, ..., n-2\}$.

Type II: $\begin{pmatrix} x & \pi^a(i) & \pi^b(j+1) & \pi^c(i+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j+1 & i+1 & j \end{pmatrix}$, $a, b, c \in \{0, ..., n-2\}$.

Type III: $\begin{pmatrix} x & \pi^a(i) & \pi^b(i+1) & \pi^c(j) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & i+1 & j & j+1 \end{pmatrix}$, $a, b, c \in \{0, ..., n-2\}$.

Type IV: $\begin{pmatrix} x & \pi^a(i) & \pi^b(j+1) & \pi^c(j) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j+1 & j & i+1 \end{pmatrix}$, $a, b, c \in \{0, ..., n-2\}$.

Type V: $\begin{pmatrix} x & \pi^a(i) & \pi^b(i+1) & \pi^c(j+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & i+1 & j+1 & j \end{pmatrix}$, $a, b, c \in \{0, ..., n-2\}$.

Type VI: $\begin{pmatrix} x & \pi^a(i) & \pi^b(j) & \pi^c(j+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j & j+1 & i+1 \end{pmatrix}$, $a, b, c \in \{0, ..., n-2\}$.

We will use the introduced definitions and types of cyclic permutations for the formalization and investigation of the adjacency properties for elements from the set $\Pi_n^C$.

## 4.Cyclic properties of adjacent permutations in the permutation polyhedron $\Pi_n$.

In correspondence with the criterion of vertexes adjacency in the permutation polyhedron $\Pi_n$ [1] for any vertex $v \in vert\ \Pi_n$ there are $(n-1)$ adjacent vertexes obtained

from the $v$ transposition of the components equal to $i$ and $i+1$ correspondingly, $i \in J_{n-1}$. Note that this is true also for any vertex $v \in \Pi_n^C$.

For any vertex $v \in vert\ \Pi_n$ we will call the transposition of components equal to $i$ and $i+1, i \in J_{n-1}$ belonging to the same $k$ length cycle of the corresponding permutation $p \in P_n$ a "break" transposition, since as a result of this operation the vertex $v_1 \in vert\ \Pi_n$ adjacent to the original one will be obtained and the permutation $p_1 \in P_n$ corresponding to the obtained vertex contains at least two cycles of length $k_1$ and $k_2$, $k_1 + k_2 = k$.

Therefore for the vertexes $v \in vert\ \Pi_n$ one transposition of components equal to $i$ and $i+1$, $i \in J_{n-1}$ can be either "break" or "conjunction" of cycles to which these components belong. If the original vertex $v \in vert\ \Pi_n$ corresponds to the permutation $p \in P_n^C$ belonging to the set of cyclic permutations and has a single cycle of length $n$, any transposition of the components $i$ and $i+1$, $i \in J_{n-1}$ will be a "break" [10, 11].

Next let us consider if it is possible to keep the cyclicity of a permutation when we have two transpositions of components equal to $i, i+1$ and $j, j+1$, $i, j \in J_{n-1}, j \neq i$.

Consider the case where there are four components involved in two transpositions of components.

**Statement 1**. If in a certain permutation $p \in P_n^C$ 2 consecutive transpositions of the elements $i, i+1$ and $j, j+1$, $i, j \in J_{n-1}, j \neq i, i+1$ have been fulfilled then the obtained permutation $p_{i,j} \in P_n$ will be cyclic if the original permutation $p \in P_n^C$ for the given components $i, j$ belongs to the type I or II.

**Proof**. Let us consider the fulfillment of transpositions for all the six types of permutations for some components $i, i+1$ and $j, j+1$, $i, j \in J_{n-1}, j \neq i, i+1$ and demonstrate which types keep the property of cyclicity. For all the six types: $a, b, c \in \{0, \ldots, n-2\}$

Type I: $j, i+1, j+1$. The original chain looks as follows:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j) & \pi^c(i+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j & i+1 & j+1 \end{pmatrix}.$$

Fulfill consecutively two transpositions: $i \leftrightarrow i+1$, $j \leftrightarrow j+1$ and represent the chain elements without changing the links to facilitate visual perception. As a result, we get:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j) & \pi^c(i+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j & i+1 & j+1 \end{pmatrix} \underset{\substack{j \leftrightarrow j+1}}{\overset{i \leftrightarrow i+1}{\Longrightarrow}} \begin{pmatrix} x & \pi^c(i+1) & \pi^b(j) & \pi^a(i) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i+1 & j & i & j+1 \end{pmatrix}.$$

Type II: $j+1, i+1, j$. The original chain looks as follows:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j+1) & \pi^c(i+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j+1 & i+1 & j \end{pmatrix}.$$

Fulfill consecutively two transpositions: $i \leftrightarrow i+1$, $j \leftrightarrow j+1$ and change places of chain elements without changing the links. We will get:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j+1) & \pi^c(i+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j+1 & i+1 & j \end{pmatrix} \underset{\substack{i \leftrightarrow i+1 \\ j \leftrightarrow j+1}}{\Longrightarrow} \begin{pmatrix} x & \pi^c(i+1) & \pi^b(j+1) & \pi^a(i) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i+1 & j+1 & i & j \end{pmatrix}.$$

Type III: $i+1, j, j+1$. The original chain looks as follows:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(i+1) & \pi^c(j) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & i+1 & j & j+1 \end{pmatrix}.$$

Let us fulfill consecutively two transpositions: $i \leftrightarrow i+1$, $j \leftrightarrow j+1$ and change places for the chain elements without changing the links. We will get:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(i+1) & \pi^c(j) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & i+1 & j & j+1 \end{pmatrix} \underset{\substack{i \leftrightarrow i+1 \\ j \leftrightarrow j+1}}{\Longrightarrow} \begin{pmatrix} x & \pi^b(i+1) & \pi^a(i) & \pi^c(j) \\ \downarrow & \downarrow & \updownarrow & \updownarrow \\ i+1 & j+1 & i & j \end{pmatrix}$$

Type IV: $j+1, j, i+1$. The original chain looks as follows:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j+1) & \pi^c(j) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j+1 & j & i+1 \end{pmatrix}.$$

Fulfill consecutively two transpositions: $i \leftrightarrow i+1$, $j \leftrightarrow j+1$. We will get:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j+1) & \pi^c(j) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j+1 & j & i+1 \end{pmatrix} \underset{\substack{i \leftrightarrow i+1 \\ j \leftrightarrow j+1}}{\Longrightarrow} \begin{pmatrix} x & \pi^a(i) & \pi^c(j) & \pi^b(j+1) \\ \downarrow & \downarrow & \downarrow & \updownarrow \\ i+1 & j & i & j+1 \end{pmatrix}.$$

Type V: $i+1, j+1, j$. The original chain looks as follows:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(i+1) & \pi^c(j+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & i+1 & j+1 & j \end{pmatrix}.$$

Fulfill consecutively two transpositions: $i \leftrightarrow i+1$, $j \leftrightarrow j+1$. We will get:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(i+1) & \pi^c(j+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & i+1 & j+1 & j \end{pmatrix} \underset{\substack{i \leftrightarrow i+1 \\ j \leftrightarrow j+1}}{\Longrightarrow} \begin{pmatrix} x & \pi^b(i+1) & \pi^a(i) & \pi^c(j+1) \\ \downarrow & \downarrow & \updownarrow & \updownarrow \\ i+1 & j & i & j+1 \end{pmatrix}.$$

Type VI: $j, j+1, i+1$. The original chain looks as follows:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j) & \pi^c(j+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j & j+1 & i+1 \end{pmatrix}.$$

Fulfill consecutively two transpositions: $i \leftrightarrow i+1$, $j \leftrightarrow j+1$. We will get:

$$\begin{pmatrix} x & \pi^a(i) & \pi^b(j) & \pi^c(j+1) \\ \downarrow & \downarrow & \downarrow & \downarrow \\ i & j & j+1 & i+1 \end{pmatrix} \underset{\substack{i \leftrightarrow i+1 \\ j \leftrightarrow j+1}}{\Longrightarrow} \begin{pmatrix} x & \pi^a(i) & & \pi^c(j+1) & \pi^b(j) \\ \downarrow & \downarrow & & \downarrow & \updownarrow \\ i+1 & j+1 & & i & j \end{pmatrix}.$$

Thus, after fulfilling two transpositions of components $i, i+1$ and $j, j+1$, $i, j \in J_{n-1}, j \neq i$, in the permutation $p \in P_n^C$ only two types of the components original location in the chain correspond to keeping the cyclicity property. The other four types do not keep cyclicity after component transpositions and lead to appearing cycles of a length less than $n$: $\pi^k(z) = z$, $k < n$, $z \in \{i, i+1, j, j+1\}$.

The statement is proven.

## 5.Conclusion

The given work has been devoted to the investigation of adjacent permutations and their cyclic properties. We have investigated permutation properties with the help of the permutation polyhedron by using the immersing in the Euclidian space.

Based on the known adjacency criterion for the vertexes of the permutation polyhedron $\Pi_n$, similar transpositions of components in permutations have been investigated.

Depending on the location of arbitrary components $i, i+1$ and $j, j+1$, $i, j \in J_{n-1}, j \neq i, i+1$ six types of permutations have been introduced. For these types the changes in the cyclic structure of a permutation that appear after fulfilling two consecutive transpositions of the corresponding components have been investigated.

## References


1. Emelichev V.A., Kovalev M.M., Kravtsov M.K. Polyhedrons, graphs, optimization. – M.: Nauka, 1981.– pp. 344. (In Russian)
2. Stoyan Y. G., Yakovlev S. V. Mathematical models and optimization methods of geometric engineering. – K.: Naukova Dumka, 1986. – pp. 268. (In Russian)
3. Grebennik.I., Pankratov.A., Chugay.A. and Baranov.A. Packing *n*-dimensional parallelepipeds with the feasibility of changing their orthogonal orientation in an *n*-dimensional parallelepiped // Cybernetics and Systems Analysis, vol 46, № 5, 2010. – pp. 793-802.
4. Stanley, R. Enumerative combinatorics, vol 1, Wadsworth, Inc. California, 1986.
5. Stoyan, Yu., Yemets, O. Theory and methods of Euclidean combinatorial optimization, ISDO, Kyiv, 1993. (In Ukrainian)
6. Grebennik I. V. Description and generation of permutations containing cycles // Cybernetics and Systems Analysis, 46(6), 2010.– pp. 97-105.
7. Donald Knuth The Art of Computer Programming, Volume 4, Fascicle 2: Generating All Tuples and Permutations, Addison-Wesley, 2005.– pp. 144.
8. Donald L.Kreher, Douglas R.Stinson, Combinatorial Algorithms: Generation Enumeration and Search, CRC Press, 1999.– pp. 329.
9. M. Bona Combinatorics of Permutations, Chapman Hall-CRC, 2004.– pp. 383.
10. Isachenko Y.A. Application of polyhedral approach to the problem of cyclic permutations // Advanced computer information technology, YKSUG, vol 2, Grodno, 2008.– pp. 203-206. (In Russian)
11. Grebennik I. V., Chorna O.S. Cyclic properties of adjacent permutations of different elements // Bionics of intelligence, №1 (82), 2014.– pp. 7–11.

# Application of the Information Works Maturity Model for the Management of Corporate Information Resources

**Sergii Kavun[1], Kostyantyn Sibilev[2]**

[1]Kharkiv Institute of Banking of University of Banking of the National Bank of Ukraine
kavserg@gmail.com
[2]Simon Kuznets Kharkiv National University of Economics, 61166, Kharkiv, Ukraine
sibilev-ks@rambler.ru

**Abstract.** The paper considers the possibility of estimating the quality of information resources management at the enterprise by using the information work maturity model. The main components of the model are studied and given a detailed characteristic. A comparison of the key components of the model is presented both with the possible ways of its practical application.

**Keywords.** information resources, information work, maturity model, information systems

## 1.Introduction

The current stage of economic development is characterized by further increasing role of the information component in almost all spheres of corporate economic activities. The rapid development of computerized tools that support work with information has led to the emergence of the economy of a new type – the economy of information and knowledge, in which the operational effectiveness of various types of enterprises is directly dependent on the management efficiency of a specific economic resource – corporate information resource (IR). For most business institutions that operate in the conditions of the economy of a new type a successful solution of the IR management problem has become the key to ensuring sustainable development and competitiveness. In this regard, further development of the effective methodology for managing IR at the enterprise remains an urgent issue of the day.

## 2.Statement of the problem

There is a considerable number of researches devoted to the study of the specifics of information as an economic resource, for example [3-5], but up to this day there was proposed no comprehensive methodology for the organization of IR management at enterprises. Some works like [2, 6] and others make an attempt to look at the management of information resources (IRs) in terms of the classic management methodology, but most do not take into account all aspects of IR manifestation at the enterprise.

From the point of the authors of this article, the idea of corporate IR management is presented most fully in [7], where there is presented a hypothesis that the effective management of IRs can be achieved through the study and analysis of information works (IWs) done at enterprises as the core activities of personnel using IRs. In particular, it is noted that the nature and manner of doing IWs directly determines the efficiency of IRs use at the company. So actually, IR management problem becomes transformed into the problem

of managing IWs. However, no specific tools for neither IRs nor IWs management are proposed in any of the mentioned studies. In this context, the focus of this work is to present a methodical approach to the management of corporate IRs based on the information works maturity model.

## 3.Information works maturity model

Information work (IW) can be considered a primary personnel activity done with the use of IRs, and is the main way of involving corporate IRs into business processes. IWs result can be expressed explicitly in the form of a specific information product or be implicit in nature and serve as a call to action or be an administrative influence, which proves the relationship that exists between information and decision-making [1]. This approach to understanding results of IWs allows to make an assumption that the way and manner of doing IWs at the company directly affects the way how core and supporting business processes are run and generally determines corporate business efficiency.

As any IW is closely connected with the use of IRs and actually represents a separate business process that can be described using a model approach, such as in [7], so the study of its specifics and nature may be based on the existing methodology of business process analysis, including models based on process improvement in organizations. Detailed analysis of such models allowed building an integrated model of corporate IWs. The use of this model enables every company to evaluate itself in relation to the five levels of IWs maturity, each characterized by a certain state of the system of IR management.

The purpose of the model is to help business institutions to achieve the highest level of IWs maturity. The model offers a pathway that a company should follow to gradually improve its corporate IR management system so that it will be harmonized with specific business needs. In order to develop an effective IR management mechanism it is necessary to understand the current state of IWs management (how it is now and how it should be), the pros and cons of each level of IW maturity from the point of four key aspects: IR of the personnel (people), the process of obtaining IRs, corporate information culture and IT infrastructure. The model describes the levels that IWs passes while it develops and affects the use of IRs, which in turn improve in quality and become transformed into an intangible corporate asset. Thus, the workers become transformed from "owners" of useful information (IRs) into people, who freely use the public domain – corporate IRs - and become part of the corporate mind.

## 4.Information works maturity levels

Based on the four basic features of corporate IWs there can be presented the following detailed description of each of the five levels of IWs maturity.

Level 1: Operations. The cost of IRs at this level can be very high because of redundant information processes, duplicative procedures of data storage and extraction and uncoordinated processes of its collection. Accuracy and correctness of information is disputable, as it may be creatively redesigned by someone to achieve personal goals.

The increase of information that is controlled individually can endanger the implementation of tasks that a company may be facing, limiting opportunities for its improvement. Managers and leaders at upper managerial levels, being rather doubtful about the quality of IRs at lower levels of the company, may be taking most of the important decisions based on their subjective experience rather than objective facts. Companies that are

at the first level of IW maturity always miss the chance to benefit from their information assets, and individuals thrive at the expense of the company.

Since a success at this level depends mostly on the personal heroism of workers that do IWs, it becomes very difficult or sometimes impossible to repeat successful information processes after key workers leave the company. Such a way of organizing and doing IWs is impossible from the point of a long-term perspective, since the struggle for IRs and distortion of information have only negative effects on the company. The company runs the risk to become uncompetitive and get out of business.

Level 2: Consolidation. From the perspective of IWs that are organized within individual business units, this level looks quite acceptable. There are available basic analytical tools and skilled staff that works with information and knowledge and there is a corporate mission that is focused on the development of corporate approach to IWs management. However, from the point of the whole enterprise the situation with IWs management efficiency is far from perfect. The units and departments operate autonomously for their own benefit; their goals do not necessarily correspond to the strategic direction of the company.

The corporate staff is mostly not interested in how every separate decision will affect the company as a whole, and even if such a question arises then there is normally not enough information for the answer. Each business unit reports to a senior management on a limited set of indicators. This approach to IRs management and IWs organization at the company is worrying, because the decision that seems true in terms of the functional unit may be wrong, if you look at it from the point of the whole enterprise. The heads of business units, that include workers from different departments, can report divergent figures and numbers as they normally obtain them from different data sources, and each one usually would try to adjust the results for the favor of a specific business unit rather than a company as a whole.

Level 3: Integration. Progress at this stage adds value to existing information systems and at the same time lays the foundation for the transition to a previously unattainable level of corporate IRs and IWs management for the entire enterprise. Technological components that were in a separated state before, at this level become combined into one synergistic system. Information flows become able to go beyond isolated IWs functions, to cross boundaries of organizational units, computing platforms and specialized tools. It becomes possible to rapidly take decisions with a full account of the context of deep and hidden interdependencies.

At this level, business users can conduct their own research with the use of freely available expert systems that give the opportunity to perform analytical studies to those who do not know specific methods of objective analysis. Reaching this level is a considerable progress in organizing IWs at a company. Many industrial enterprises fail to achieve this level even in modern conditions of automated data processing facilities. Corporate management of companies that are at this level of IWs maturity, are able to cover the whole enterprise in one look, and quickly come to understanding of the strategic benefits of further progress in the development of methods of IWs organization and control and begin planning the transition to the next maturity level.

Level 4: Optimization. Enterprises of the fourth level can be considered as those that have reached competitive advantage and market leadership through effective organization of IWs. Using high-quality IRs, they continuously improve their processes and business models. The corporate information culture promotes interaction and collaboration of functional units, so these companies can quite easily adapt to changes. But, despite all the improvements, the company does not take big risks, even if they promise to be rewarded,

since the analysis clearly determines the opportunities that have to be accepted, and those that better should be neglected; automatic feedback ensures that from any possibility the company would get a least some benefit. However, at this level there comes time when optimization simply cannot significantly increase the final value that makes it necessary to generate innovations.

Level 5: Innovations. Enterprises that reach the fifth level of IWs maturity can be considered as those that have adopted direct progressive strategy to meet market volatility with continuous updating. The company of this level produces a steady stream of new information products, services and business models to work with IRs, always being ahead of competitors and maintaining a leading position at the market. Upgrading of the technology of performing corporate IWs is not a single event but a daily routine for the company of this highest IW maturity level. The change of tools and approaches to work with corporate IRs becomes a part of the overall corporate information culture, information processes and infrastructure. Being at this level, the company becomes able to model new propositions in a virtual environment before sending them to the real world. This feature allows it without taking high risks to feed an endless stream of new business ideas. This innovative company has the potential of prosperity even in times of economic crisis and instability.

## 5.Conclusion

The proposed descriptive IWs maturity model allows performing an analysis of the current state of the corporate IR management and the organization of the corporate information management in general. Application of the model at the industry level may provide the possibility for the classification of enterprises depending on the quality of their IWs. Practical application of the model for a particular company may allow assessing the quality of using its information for doing IWs for providing more effective management of corporate IRs. Further evolution of the presented approach should be focused on the development of tools for assessing the quality of corporate IWs and recommendations to promote IWs to higher maturity levels.

## References

1. Gates B. (1999). Business at the speed of thought, Warner Books, Inc.
2. Karminskij A. M., Karminskij S. A., Nesterov V. P., Chernikov B. V. (2004). Business informatization : concepts, technologies, systems, edited by A. M. Karminskij, 2nd edition, revised, Finances and Statistics, 624 p. [in Russian]
3. Castells M. (2001). The Internet Galaxy, Reflections on the Internet, Business and Society. Oxford, Oxford University Press.
4. Melnik L. G. (2010). Social and economic problems of the information society: monograph / L. G. Melnik, M. V. Brjuhanova. 2nd edition, University Press, 896 p.
5. Parinov S. I. (2002). The theory of network economy, IEOPP RAN, 168 p. [in Russian]
6. Pushkar A. I., Grabovskij E. N., Ponomarenko E.V. (2005). Strategic management of the development of electronic business and information resources (models, strategies, mechanisms): scientific edition, HNEU, 480 p. [in Ukraine]
7. Pushkar A. I., Nazarova S. A., Sibilev K. S. (2012). Management of corporate information activity in the economics of knowledge : monograph, Publishing house HNUE, 560 p. [in Ukraine]

# Statistical Analysis of Critical Infrastructure Protection

**Sergii Kavun[1], Vladimir Golubev[2], Olexandr Trydid[3], Revak Iryna[4]**

[1] Ph.D., Kharkiv Institute of Banking of the University of Banking of the National Bank of Ukraine, Department of Information Technologies, Victory Av., 55, 61174, Kharkiv, Ukraine, Tel: +38067 709-55-77; E-mail: kavserg@gmail.com

[2] Candidate of Legal Sciences, Director of Cybercrime Prevention Department, "Cyber Safety Unit", "Noosphere Ventures" Corporation; E-mail: vladimir@crime-research.org

[3] Dr. of Science, Prof., Kharkiv Institute of Banking of the University of Banking of the National Bank of Ukraine, Department of Information Technologies, Victory Av., 55, 61174, Kharkiv, Ukraine, Tel: +38057-337-99-96; E-mail: khibs@khibs.edu.ua

[4] Ph.D., Prof., Lviv State University of Internal Affairs, Department of Information Technology in the Work of the Interior and Economic Security, Horodotska Str. 26, 79007, Lviv, Ukraine, Tel: +38067 444-09-63; E-mail: irrevak@inbox.ru

**Abstract.** The article discloses the problems of an Internet-crime (cybercrime) prevention that is a purpose of this article. As one of an approach for presenting of these some results is a show of currency and further events and actions, which can create a discuss for next considerations in this area. At one point, the Internet enabled to commit previous traditional offences more effectively with no punishment. At another point, it produced new, recently unknown types of social assaults, complex and system of which reflected in such negative social phenomenon as Internet-crime. Under new fast changing country (on example of Ukrainian) realities it becomes necessary systematically and successively to study Internet-crime overall and most popular types and to develop effective measures or approaches to combat and prevent crimes in the global network. As a second approach, which the authors were wanted to show some time-frequency distributions of the main definitions in this research area, which are characterize the modern tendencies and their popularity for critical infrastructure protection in throughout world. As a result of this presentation can be more deep understanding of a situation in similar areas.

**Keywords.** Cybercrimes, computer crimes, Internet-crimes, Cyber-criminals, Cyber Security.

## 1.Introduction

Internet cyber criminals keep perfecting their fraud methods, leading to material losses up to tens of billion dollars and posing serious risks to many countries, including Ukraine. Therefore, specialized departments and structures are created to combat this type of crime. They constantly get more and more powers and better technical facilities. One of the recent examples is European center of Cybercrime Prevention that commenced its work in the beginning of 2013.

On March 19, Europol released a report “The EU Serious and Organized Crime Threat Assessment (SOCTA 2013) with an assessment of the growing globalized and organized crime rates by means of the Internet (Official site of Europol, 2013).

Unfortunately, Ukraine has ranked the fourth (following Russia, Taiwan, and Germany) among the world countries presenting the highest cyber threats. This data was

shown on the map of countries – cyber-attacks sources, visualized in the report of Deutsche Telekom, a leading German operator.

## 2.Time-frequency analysis for some definitions of the critical infrastructure protection

It should been mentioned that in nowadays we can see a strengthening of a popularization for some definitions or key aspects in this area (Kavun, 2009), for example, it covers the following definitions – Cybercrimes, Computer Crimes, Internet-crimes, Cyber-criminals, Cyber Security, etc. For those and some other similar definitions, the authors have carried out a research for all these definitions about their distribution on the time line (2000-2013). This research is based on the method of Internet-analysis (Kavun, Mykhalchuk, Kalashnykova, Zyma, 2012). At the same, this distribution was based on their popularization or demand; also, it can called as time-frequency analysis. Some results with their normalizing of this research (for better presentation) were shown in Table 1. All of these submitted data has been normalized for better visualization and to receive a possibility of comparing those results between themselves.

Table 1

Results of the demand distributions with their normalizing

| Year / Definition | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cybercrime (CbC) | 4,85 | 5,48 | 19,81 | 2,82 | 4,92 | 1,93 | 1,53 | 5,45 | 4,48 | 3,12 | 8,15 | 25,90 | 23,47 | 18,20 |
| Computer crime (CpC) | 8,28 | 6,52 | 2,32 | 1,50 | 6,19 | 2,75 | 31,91 | 16,06 | 6,04 | 5,35 | 11,62 | 8,78 | 5,75 | 4,04 |
| Internet crime (IC) | 0,05 | 0,03 | 0,02 | 3,09 | 3,06 | 1,20 | 0,05 | 16,73 | 0,07 | 28,48 | 15,28 | 7,47 | 23,03 | 1,55 |
| Cyber-criminals (CbCr) | 6,26 | 5,62 | 5,10 | 5,51 | 5,89 | 6,34 | 7,12 | 7,65 | 8,81 | 8,96 | 3,75 | 10,46 | 7,50 | 11,02 |
| Cyber Security (CS) | 1,16 | 1,44 | 1,30 | 2,92 | 2,61 | 6,02 | 17,80 | 11,13 | 22,00 | 5,11 | 2,65 | 3,69 | 9,28 | 16,79 |
| Crime analysis (CA) | 0,01 | 7,42 | 0,00 | 2,52 | 2,80 | 1,76 | 5,04 | 4,83 | 40,77 | 19,24 | 9,91 | 3,53 | 8,73 | 0,87 |
| Cyber-infrastructure protection (CIP) | 2,23 | 5,50 | 5,41 | 6,10 | 6,08 | 6,81 | 10,42 | 9,46 | 24,27 | 4,45 | 29,62 | 0,31 | 0,24 | 2,23 |
| Cyber-crime detection (CCD) | 2,85 | 34,27 | 33,65 | 37,57 | 34,48 | 2,50 | 2,46 | 3,10 | 3,03 | 3,16 | 3,36 | 3,70 | 3,55 | 3,10 |

As we can see from Fig. 1, which is showing very clearly and graphically. The most popular of these definitions were used since 2006. In addition, only for the definition “Cyber-crime detection (CCD)”, we can look a big growth since 2000 over 2004. For remaining definitions, we can see some fields of “activity” since 2006. Especially it concerns for the following definitions: “cyber analysis” in 2008 is received big growth, but after two years, this

area has become not interesting for the most specialists; "Computer crime" is an area, which had a variable success in 2006 and 2010; also as the definition "Internet crime" has a variable success only in 2009 and 2012. For the others definitions, which were studied in author's research, we can see a quite variable interest (ES INFECO, 2012-2013). At the same time, this interest is confirming a fact that notwithstanding that situation these areas (based on similar definitions) are remaining in the sphere of some interests for the most specialists and professionals. Based on those data from Table 1, we could built some graphics (map's frequency distribution and linear 3D-frequency distribution), which is shown in Fig. 1-2.

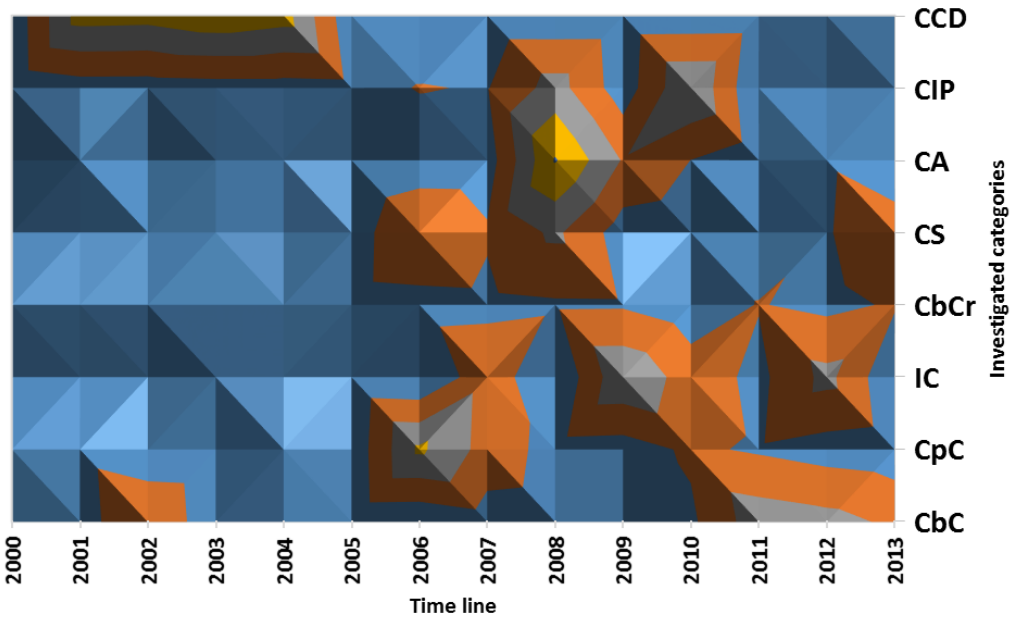
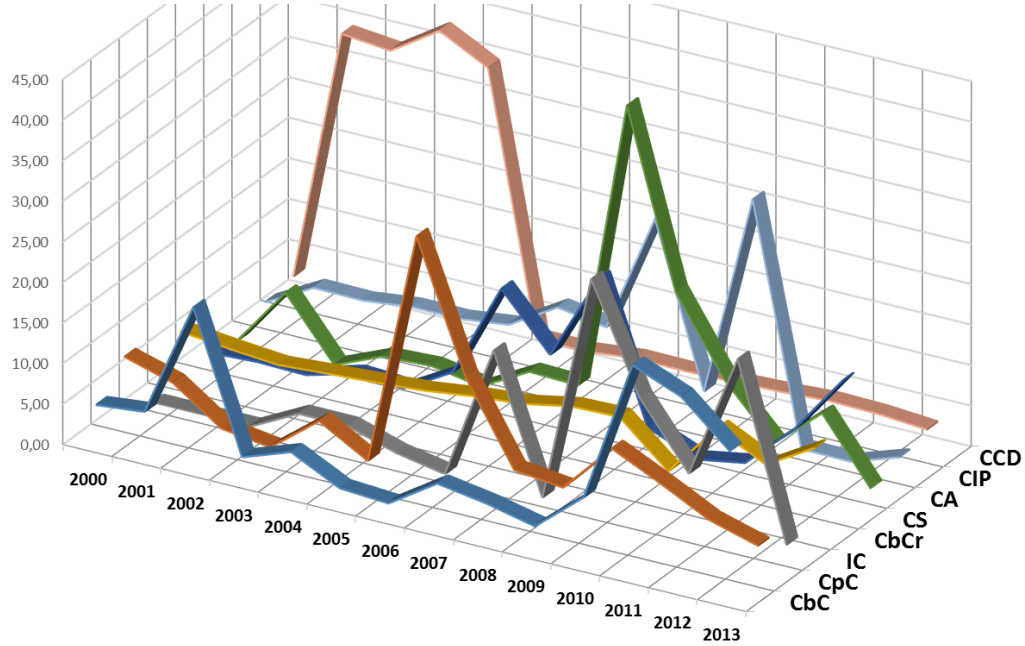



Fig. 1. Map's frequency distribution of the definitions

As a continue of our research for those definitions and similar areas of security, we could built some distributions for separated definitions with forecasting on the next period of time (for next several years) based on the well-known trend models (Kavun, Sorbat, Kalashnikov, 2012). For these trend models, we also are shown their approximation functions with the accuracy of the approximation ($R^2$), as it is shown in Fig. 3-11.

In Fig. 3 (also as in the next figures), the general distribution (distribution graph) is shown with help of a solid bold line and a trend line is shown with help of a dotted line. Based on the approximation function and its accuracy of the approximation ($R^2$, this accuracy is high enough) we can say that nowadays the popularity of cybercrime is decreasing, but it does not the fact that we have some reduction of some offenses in this area. For each point of all set of data in this Fig. 3 is shown some limits of standard accuracy (of error), which is determining an area of precision of our measuring (Kavun, 2009). In addition, you can see, in fact, this model is a model of 5th degree, because a coefficient at the variable $X_6$ is small enough, and it is equal about $10^{-3}$.

Based on the approximation function and its accuracy of the approximation ($R^2$), as is shown in Fig. 4, we can say that nowadays the popularity of computer crime also is decreasing, and besides, this distribution has this reduction since 2007. Therefore, it is a good tendency and, it will be able to acknowledge about some success in a fight with computer crimes of some different government and private organizations.




Fir. 2. Linear 3D-frequency distribution of the definitions

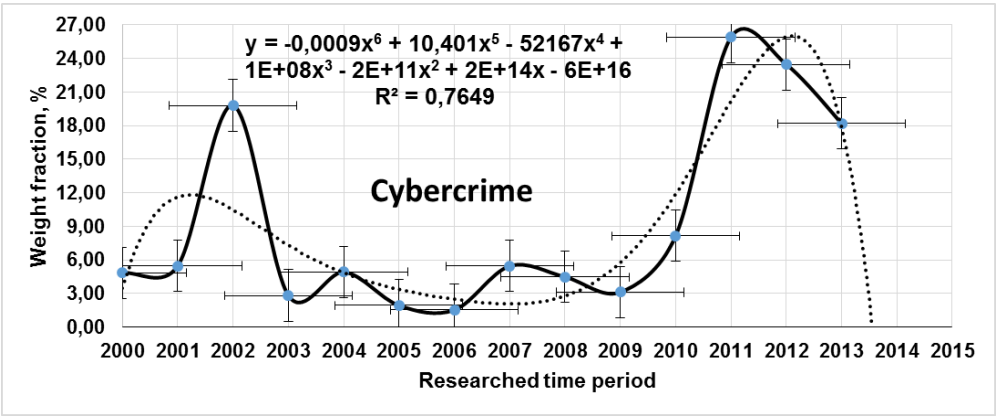



Fig. 3. Distribution graph of "popularity" or "demand" for the definition "Cybercrime" with forecasting based on a trend model (the polynomial model of 6th degree)

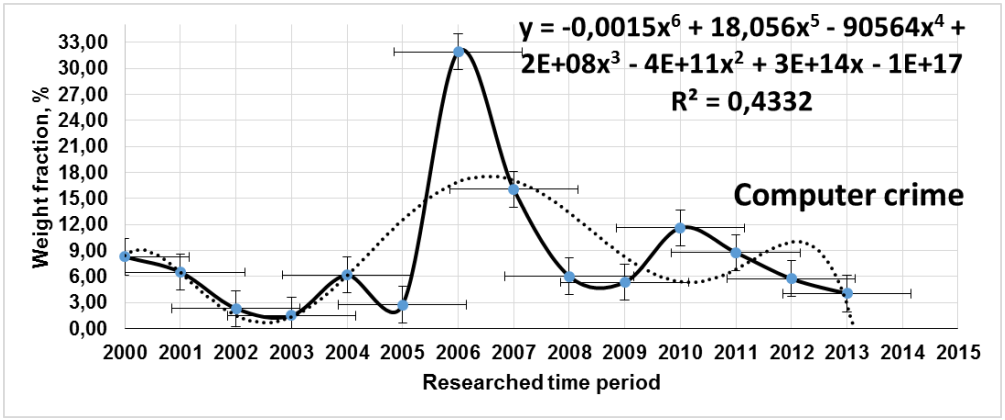



Fig. 4. Distribution graph of "popularity" or "demand" for the definition "Computer crime" with forecasting based on a trend model (the polynomial model of 6th degree)

However, we cannot trust enough of this distribution graph and its forecasting, because we have the accuracy of the approximation, which is small enough, so this forecasting we can consider only in approximate form. For the following definition "Internet crime", its distribution graph is shown in Fig. 5.

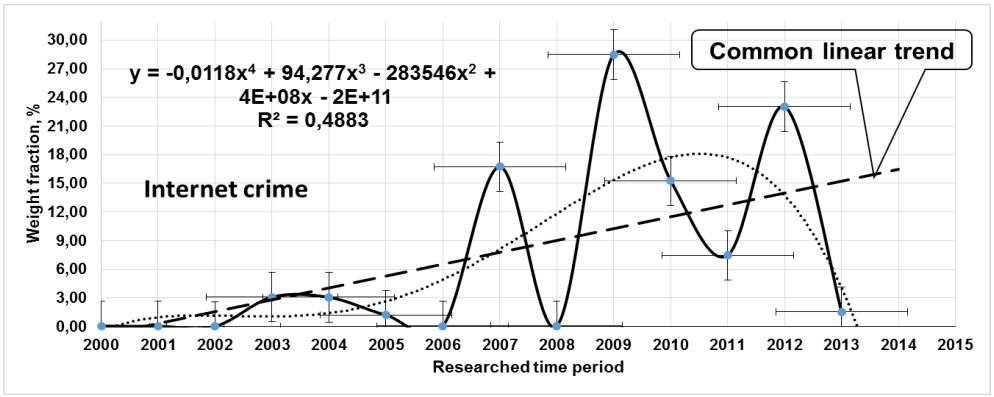



Fig. 5. Distribution graph of "popularity" or "demand" for the definition "Internet crime" with forecasting based on a trend model (the polynomial model of 4th degree)

Based on the approximation function and its accuracy of the approximation ($R^2$), as is shown in Fig. 5, we can say the time-frequency distribution for the definition "Internet crime", in general, has a common positive linear tendency (based on common linear trend, as it show with help of long dotted line in this figure) during all researched time period, but with the enough small accuracy of the approximation (< 0,5). This fact is not confirming our forecasting for this area, so we must use this forecasting with all caution.

In addition, we could be reduced a degree of this polynomial trend line to 4th degree in comparing with 6th degree without significant reducing of the accuracy of the approximation, as it showing below.

$y = -0{,}0001x^6 + 1{,}5418x^5 - 7732{,}4x^4 + 2E{+}07x^3 - 3E{+}10x^2 + 2E{+}13x - 8E{+}15$, $R^2 = 0{,}4893$.

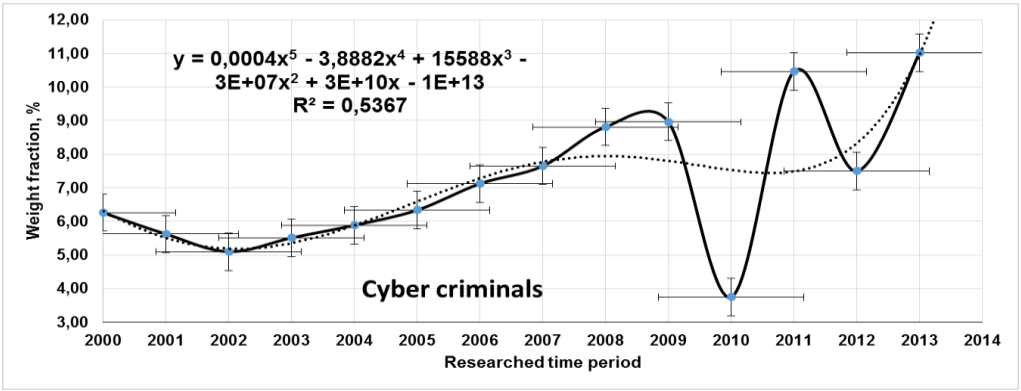



Fig. 6. Distribution graph of "popularity" or "demand" for the definition "Cyber criminals" with forecasting based on a trend model (the polynomial model of 5th degree)

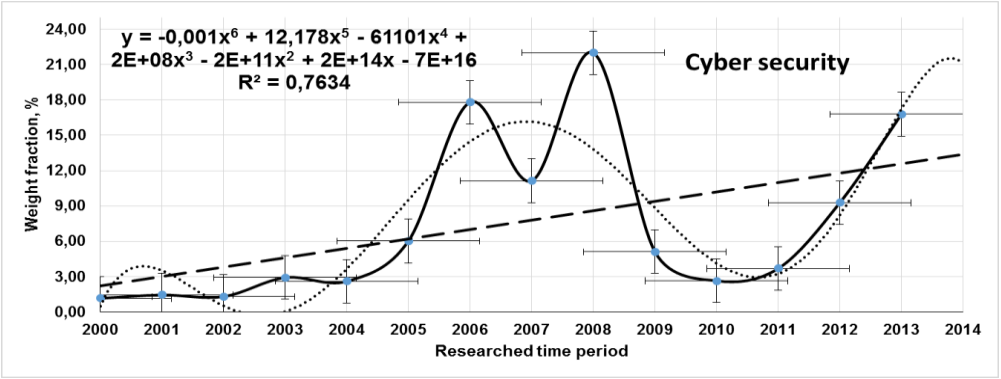



Fig. 7. Distribution graph of "popularity" or "demand" for the definition "Cyber security" with forecasting based on a trend model (the polynomial model of 6th degree)

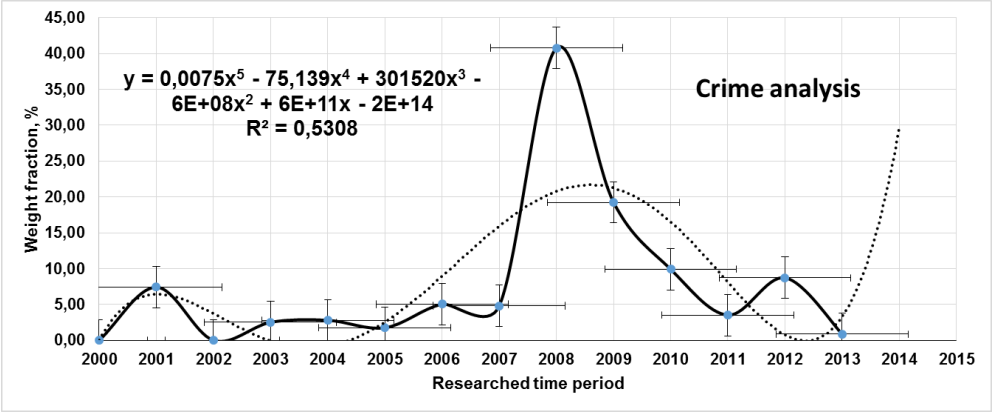



Fig. 8. Distribution graph of "popularity" or "demand" for the definition "Crime analysis" with forecasting based on a trend model (the polynomial model of 5th degree)

We can see very interested distribution in Fig. 6, which is presenting for the definition "Cyber criminals". From this figure, we can see a positive (with permanent

increase) trend for any kind of trend models, but with the enough small accuracy of the approximation, a little more 0,5. In addition, we can confirm that fact this distribution is sorrowful, because it is confirming a permanent increasing of different kinds of the cyber criminals in throughout the world. Therefore, we must have some interest for decreasing of this distribution, thus, this theme is actual. In fact, we could see this trend model based on the polynomial dependence has the polynomial model of 4th degree, because a coefficient at the variable $X_6$ is enough small (a little more $10^{-4}$).

We can see an increasing of a growth of popularity of the area "Cyber security" for the confirmation of a necessity for decreasing of a growth of popularity of the area "Cyber criminals", as opposed of this fact and as it is shown in Fig. 7. Moreover, this positive tendency is observed for any kind of the trend models, as for the linear, and for polynomial models. In addition, this growth is began since 2010-2011, when some well-known cyber-attacks on the different government and private infrastructures were done in throughout world. About this and other actions will describe below. The enough high accuracy of the approximation is increasing of our confidence and an independence level of forecasting, but only for determination of the general tendencies.

We can see a significant growth of popularity for this area in 2008 based on a distribution graph of the definition "Crime analysis", which is shown in Fig. 8. Nevertheless, after that period of time, we also can see a significant recession up to nowadays. Moreover, we can hope (because we have the enough small accuracy of the approximation, a little more 0,5) that in further period of time will be to rise some interest and our attention to this area. It can confirm the enough small positive tendency, which is a general tendency. The enough small confirmation of the significant growth of a popularity for this area in 2008 can be some normative and legislative acts and documents (about which will be describe below), which were accepted in some EU and other countries.

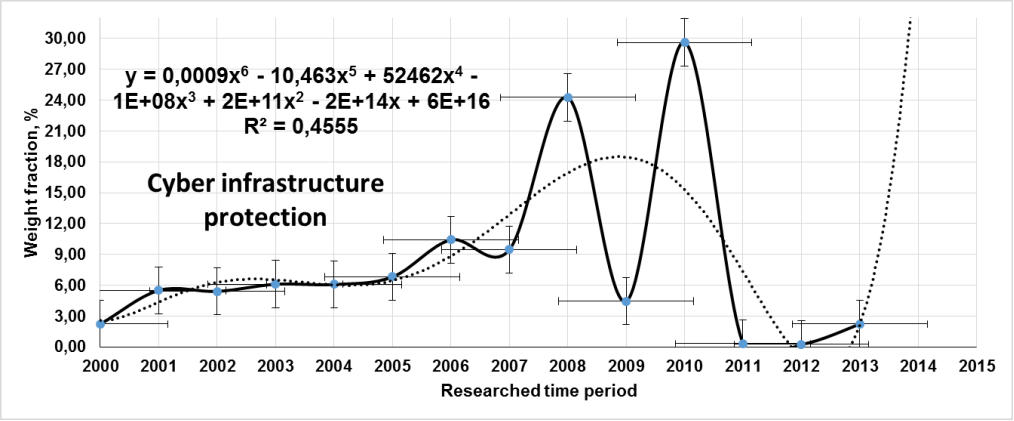



Fig. 9. Distribution graph of "popularity" or "demand" for the definition "Cyber infrastructure protection" with forecasting based on a trend model (the polynomial model of 6th degree)

For this definition "Cyber infrastructure protection", we can say only about some tendencies, because we have the enough small accuracy of the approximation (a little less 0,5). Based on the general tendency, we can see an enough small positive trend before 2010, but after 2011, we have a recession. It is also confirm some recession of an interest in this area, but we can hope after 2012 (it can see in Fig. 9) this interest and popularity will be have more growth, in compare with past actions.

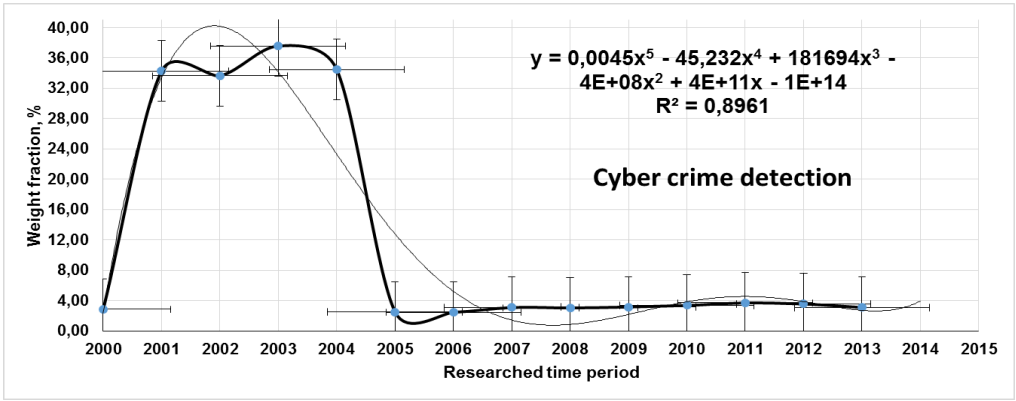



Fig. 10. Distribution graph of "popularity" or "demand" for the definition "Cyber-crime detection" with forecasting based on a trend model (the polynomial model of 5th degree)

It confirms that fact, in the beginning of the 2000s this theme or area was very interesting and popular. Besides, we can say with some confidence that the popularity and interest in this area will have quite stability in the next period, this assessment based on the enough high accuracy of the approximation (almost equal 0,9). Thus, we could described the common tendencies in this and other areas based on using of developed by one from authors method of Internet-analysis (Kavun et al., 2012). The method of Internet-analysis developed by the authors can be applied in any given area of an activity, regardless of its properties and features. The purpose of using this method is to get some assessment or collection of selected concepts (terms or definitions), which form the so-called categorical apparatus of a scientific research. According to estimates, authors can formulate a conclusion about the need for a further research in this area, show the urgency and demonstrated the importance of the calculations, to allocate a narrow major for a further research of young scientists. The proposed method can be used to study the activities of prominent scientists, professionals from designated areas of activity. Secondly, this method can be used for demonstrating the importance and the relevance of a research, and the conclusions and recommendations drawn, for example, to select consultants (leaders) in their scientific activities.

For a more complete openness, authors could show the possibilities of the research conducted by the authors on the example of a field of the information security, which was selected, because of its global relevance. Using of this method based on the specifics of the query language that is supported by all search engines and the shape of the query, the results on the set of selected search engines are averaged at a given time interval. Thereby, the dynamism of the study is achieved. The authors are showed as an example of some limited areas of interest an interpretation of the developed method with the help of graph theory to be able to use known techniques of optimization and further analysis (Kavun and Mikhalchuk, 2009). Thus, the scope of the developed method of Internet-analysis is multi-faceted in its specificity and tolerance in limited areas of an interest (Kavun, 2011).

## 3.Statistical and financial data in area of legislations (on example of Ukraine)

As a response to the difficult criminal situation in the Internet-area, Ministry of International Affairs of Ukraine has prepared a draft law "On cyber safety of Ukraine", handling hacker attacks of government websites and promotion of pornography, violence and separatism as a threat to national security (Site of Security Service if Ukraine, 2003).

The crime connected with the use of computer technologies presents a broader problem, which is at first conditioned by the time, precisely by the people's ability to engage electronics and computer technologies.

Just as there was a period of transport development and implementation into the everyday life, which required the protection of the process participants, similar period we face nowadays. If upon introduction of computer technologies there was no strictly regulating legal framework as a man had not then imagined where his invention would find its implementation, and therefore there was no defined circle of the process participants and no rhetorical question of legal protection of their interests. Presently, the inventions in computer technologies affect and involve more and more participants. Hence, the necessity of detailed and thorough regulation of this area of human achievements follows (Donetsk, 2008).

Since the beginning of 2013, a department of cybercrime prevention of the Ministry of International Affairs has detected 23 cases of illegal debiting from accounts of commercial enterprises amounting to 12.5 billion UAH, 9.2 billion UAH of which were successfully recovered. Financial frauds are gradually moving from open market to global network. The number of fraudulent payment cards, which could be accessed by criminals in the trading centers, is reducing in Ukraine. At the same time the crime rate, related to the Internet and "Client-bank" system, has boosted, and this trend continues.

It should be noted now new spheres and culture prompted by the progress has appeared, the pushing force for which was a human's laziness. The existing world of computer technologies offers great opportunities, it enables to communicate, search and save information, shop online and pay by electronic means, work and get remuneration. Thus, today a human is also able to lead a full and productive life in cyberspace (Darlington, 2013).

However, it is not just a mass computer distribution, which forms information sphere. A significant, even a key role in this process belongs to various constantly is improving computer technologies.

This way, the advancement of information technologies, growing production of supporting technical facilities and their application along with the development of electronic payment methods as a potential object of criminal offence and increasing availability of such facilities are the natural cause of upraise, existence and growth of property offence using computer technologies (Idov, 2013).

The independent poll undertaken by the Centre of cybercrime investigation among 100 people actively using the Internet for more than one year, shopping and paying online (using bankcards, e-wallets) showed the following results:

a) 87% of respondents have suffered from criminal offences involving computer technologies,

b) 2% of respondents have realized the property loss during the first hours,

c) 5% of respondents have realized the property loss during the first month,

d) 93% of respondents have realized the property loss after the first month.

Social danger of such crimes lies in their latency time, which in turn rises from the fact that the person often has no idea that he/she had become a victim of a crime. As the result the criminal feels his impunity and therefore above the society and law.

## 4.Aggregated classification of some definitions in the area of cyber security

Based on the authors research, was aggregated some volume of different information from more sources (Kavun and Brumnik, 2013). Therefore, we can form some aggregated classification of these similar definitions in this area for more helpful and comfortable perception. In addition, we want to present some main definitions in this area:

1. Computer crimes in a broad sense is a criminal act or some actions, which has been perfected with using and (or) with regard to computer information, a computer of any kind, computer systems, and their networks.

2. Crimes in a sphere of computer information is a criminal act or some social (publicly) dangerous actions, objects (or subjects) and (or) means of which are different computer information (data).

3. Computer crimes in the narrow sense is a criminal act or some actions, objects (or subjects) of which are different computer information (data).

Based on some lists for the EU countries, which were accepted and approved as the Recommendation No. R (89)9 of the Committee of Ministers to Member States on Computer-related Crime (adopted by the Committee of Ministers on 13 September 1989 at the 428th Meeting of the Ministers' Deputies), we aggregated a classification in visual type, as it shown in Fig. 11.

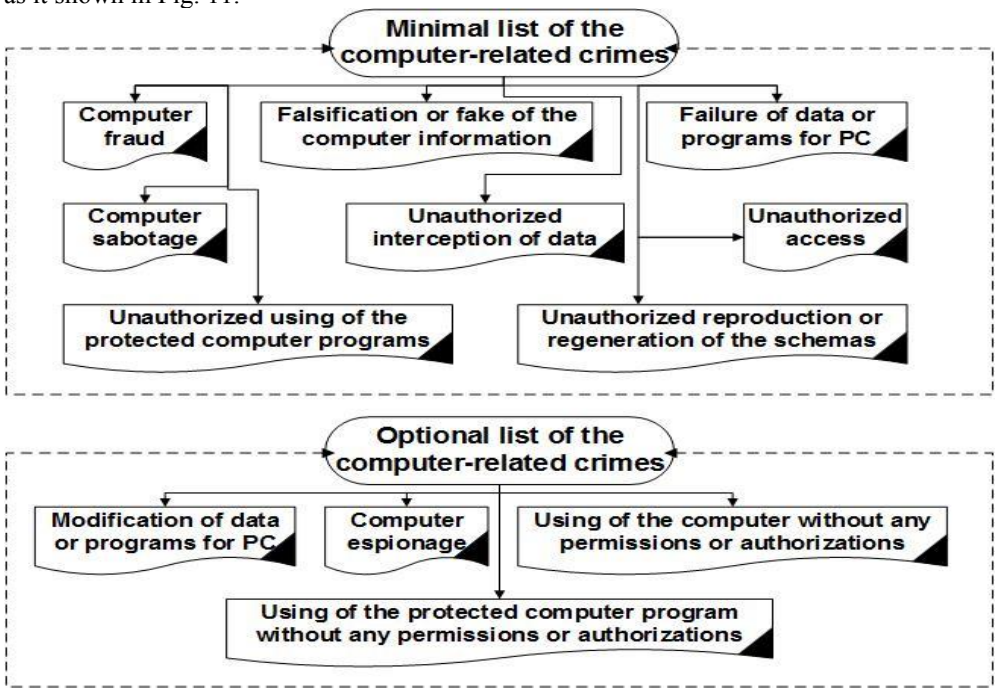



Fig. 11. Visualized classification (lists) of the computer-related crimes

The first complete classification of the computer crimes was offered in 1983, in Paris (France) from a group of some experts of the Organization for Economic Co-operation and Development, OECD. This classification based on some interests of owners and property owners. This classification has the following differentiations:

1. **Economic and computer crimes**: computer fraud, computer and economic espionage and a theft of some programs for a computer of any kind; computer sabotage; a theft of any kind services; unauthorized entry (an access) to automated and information system(s), and the traditional economic crimes, which are doing with help a computer;

2. **Computer crimes against individual people rights and the inviolability (privacy) of the private sphere**: input to computer system(s) an incorrect and invalid data about physical or legal person (or entity); illegally collecting the right (or personal) data; illegally misuse of some information, which is presenting on some machine carriers (computer devices); illegally disclosure of some information (for example, bank secrets, medical secrets, etc.);

3. **Computer crimes against some interests of state (country) and society**: some crimes against state or public (social) safety (security); violation of some rules for a transfer of information to abroad (illegally export of an information); disorganization of a work (or functioning) of defense systems; illegally misuse with automated systems for vote counting in elections and decision-making parliamentary.

In addition, we can create an aggregated classification of the all kind of the computer crimes, as it shown in Fig. 12.

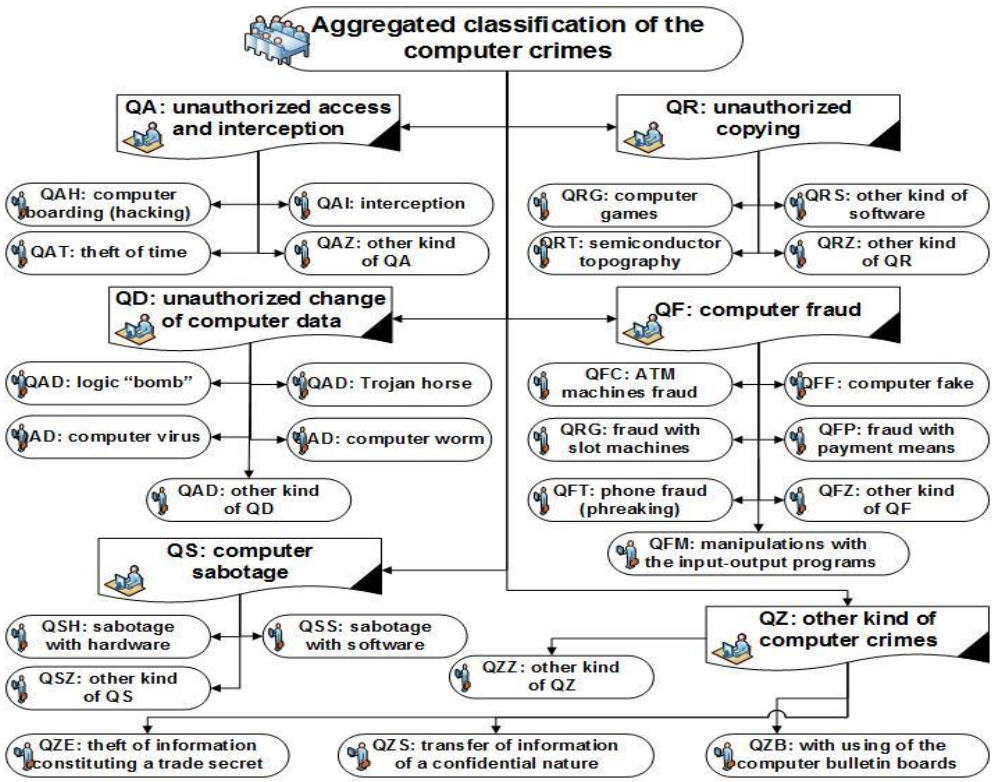



Fig. 12. Aggregated classification of any kinds of the computer crimes

For well-known methods of an unauthorized access and interception of information can pick out the following basic definitions. These definitions are also relevant in other different areas, such as information security, risk management on the enterprise, cyber-infrastructure protection, etc.:

1. **Bugs** is some kind of electronic devices, which involve of the microphone setup in a computer to intercept of some conversations between service personnel.

2. **Data leakage** is a possibility for the collecting of any kind of information, which needs for receiving of main data about some technologies, which are using in the system (Linde, 1975).

3. **Scavenging** is a data search, which were of some users after their working on a computer. It is dividing on the physical and electronic searching.

4. **Piggbacking** is a method, which characterized unauthorized penetration into space and electronic close zones (places).

5. "**Between the lines entry**" is a method, which based on an illegal connection to the communication link (or line) of a system user or computer network. Connection is doing in the time moment, when the user is finished its communication session, but he had not a time to disconnect from the network yet.

6. **Browsing** is a method of an unauthorized access to base data or some files of a legal user, is doing based on finding of some weaknesses in the protection system or computer network. If an offender were able to find at least one weakness, then he/his will be able to use this weakness each time for further reading and analyzing of any information into system. Hi/his also can copying and using of this information for their personal purposes.

7. **Trapdoor entry** is a method of an unauthorized access, which happening during detection of some errors or mistakes in a logic of building a program. These detected errors or mistakes can using repeatedly.

8. **Trapdoor** is some development of the previous method, which differ of a fact that some special control commands are including in the detected place of a program for further using for their own purposes. It also called as "program backing".

9. **Masquerading** is a method, at which the offender is penetrating into a computer system with using of necessary means pretending to be legal user.

10. **Spoofing or mystification** is a method, which is using at a random connection of an unfamiliar computer system to the system of a victim. The offender is forming some "credible" requests from the system of victims. After it, this offender can supported a delusion from that user (victim) during some period of time and received an information (confidential, secret, personal, etc.), which will be useful for this offender.

We and some other specialists and professionals in this area are considering that this classification has one significant weakness (United Nations Crime and Justice Information Network. Centre for International Crime Prevention, 2000). It is entering for using a literal "Z", which led to a chaotic (from a point of criminalistics) mixing of criminal and law principles and the technical features of automated data processing.

Some modern classifications are including the following directions: illegal actions with a computer information; illegal actions in area of communications; illegal actions with information devices; illegal actions with other kinds of an information.

Thus, as we can see in legal field has some different classifications, so you can use any kinds of these classifications based on your own purposes, tasks and personal experience.

The authors hope that these and other similar classifications (Council of Europe - Documents database, 1989) will be able to help some experts, specialists and professionals in

this area for their own researches and other kinds of activity. In our opinion, these aggregated classifications can use as some fundamentals in the researches in criminal and law fields.

## 5.Conclusion

High social danger of the Internet-crimes mainly results from the increasing role of system of the social relations, which are under its threat, and from its transnational and organized character (Chimiris, 2013). Not a single country can actively oppose this malice alone. Because of this, the intensification of international cooperation becomes of a high necessity.

Under the new realities of fast country entry to the single information space, it becomes necessary systematically and successively to counteract the cybercrime overall and the most popular types, and to minimize the harm to the economy. It is important to develop effective measures of the Internet-crimes combat and prevention, improve legal system of information security, including cyber safety.

## References

1. Official site of Europol. (2013). The EU Serious and Organized Crime Threat Assessment (SOCTA 2013). Retrieved August 5, 2013, from https://www.europol.europa.eu/sites/default/files/publications/socta2013.pdf.
2. Site of Security Service if Ukraine. (2003). A draft law of Ukraine amending Law of Ukraine "On the basis of national security" on cyber safety. Retrieved August 5, 2013, from http://www.sbu.gov.ua/sbu/control/en/publish/article?art_id=89497&cat_id=42924.
3. Problems of crime prevention in computer technologies (in Ukrainian). (2008). Donetsk, Ukraine.
4. Idov, R. (2013). Computer Crime Research Center. Peculiarities of information security in Ukraine. Retrieved August 5, 2013, from http://www.crime-research.ru/interviews/incidents22/.
5. Chimiris, M. (2013). Computer Crime Research Center. To limit anonymity in the net does not mean to oppress liberty of speech. Retrieved August 5, 2013, from http://www.crime-research.ru/interviews/maxim27/.
6. Darlington, Mwendabai. (2013). Zambia Daily mail. Cyber crime: threat to economies? Retrieved August 5, 2013, from http://www.daily-mail.co.zm/?p=5209.
7. Kavun, S. V., Mykhalchuk, I. V., Kalashnykova, N. I., Zyma, O. G. (2012). A Method of Internet-Analysis by the Tools of Graph Theory. Intelligent Decision Technologies. Smart Innovation, Systems and Technologies, 15 (1), 35-44.
8. International research portal (lab) of the information and economic security ES INFECO. (2013). Retrieved August 5, 2013, from http://infeco.net/infeco-overview/article/158-statistical2.html.
9. International research portal (lab) of the information and economic security ES INFECO (2012). Retrieved August 5, 2013, from http://infeco.net/test-security/statistic.html.
10. Kavun, S. V., Sorbat, I. V., Kalashnikov, V. V. (2012). Enterprise Insider Detection as an Integer Programming Problem. Intelligent Decision Technologies. Smart Innovation, Systems and Technologies, 16 (2), 281-289.
11. Kavun, S. (2009). Informational security. Kharkiv: Pub. of INZHEK.
12. Kavun, S., Mikhalchuk, I. (2009). Analysis of categorical apparatus in the field of economic and information security. Economic of development, 3 (51), 9-14.
13. Kavun, S. (2011). Structuring the normative and legal providing in the system of economic security of enterprise, Foreign Trade. Economic Security, 7, 22-28.

14. Kavun, S., Brumnik, R. (2013). Management of corporate security: new approaches and future challenges. Editorial Denis Galeta and Miran Vrsec. Cyber security challenges for critical infrastructure protection (pp. 141-151). Ljubljana: Institute for Corporate Security Studies. Retrieved August 5, 2013, from http://www.ics-institut.com/research/books/5.
15. Linde, Richard R. (1975). Operating System Penetration, Proceedings. NCC.
16. United Nations Crime and Justice Information Network. Centre for International Crime Prevention. (2000). Documents of 10th United Nations Congress on the Prevention of Crime and the Treatment of Offenders Vienna. Retrieved August 5, 2013, from http://www.uncjin.org/Documents/congr10/1e.pdf.
17. United Nations Crime and Justice Information Network. Centre for International Crime Prevention. (2000). Documents submitted for the consideration of the Commission on Crime Prevention and Criminal Justice at its 9th session (Vienna, 18-20 April 2000). Retrieved August 5, 2013, from http://www.uncjin.org/Documents/9comm/index.html.
18. Council of Europe - Documents database. (1989). Recommendation No. R(89)9 of the Committee of Ministers to Member States on Computer-related Crime (adopted by the Committee of Ministers on 13 September 1989 at the 428th Meeting of the Ministers' Deputies). Retrieved August 5, 2013, from https://wcd.coe.int/com.instranet.InstraServlet? command= com.instranet.CmdBlobGet&InstranetImage=610660&SecMode=1&DocId=702280&Usage=2.

# Methods of Formation of a Security Policy in Access Differentiation Processes

**Anatoliy Fedorovich Verlan[1], Salim Karimovich Ganiev[2], Durdona Yakubdjanovna Irgasheva[3], Aybek Turapbayevich Imamaliev[4]**

[1] Pukhov Institute for Modelling in Energy Engineering, NASU, Ukraine
a.f.verlan@gmail.com

[2] Tashkent University of Information Technologies, TUIT, Amir Temur Str. 108, Tashkent, Uzbekistan

[3] Tashkent University of Information Technologies, TUIT,
Amir Temur Str. 108, Tashkent, Uzbekistan, durdona.ya@gmail.com

[4] Tashkent University of Information Technologies, TUIT,
Amir Temur Str. 108, Tashkent, Uzbekistan, oybek.imomaliev@mail.ru

**Abstract.** This article describes methods conditioning of distributed security policy information - communication systems in the processes differentiation of access. Approach to research of questions of isolation of a subset of subjects and objects in a local segment for the organization of safety of the distributed systems is offered. Are investigated and provided the analysis of models of safety of a network, based on a combination of the discretionary, mandatory and role principle of differentiation of access.

**Keywords.** Distributed information and communication system, the local segment, subjects, objects, access, safety models.

## 1.Introduction

A characteristic feature of our time are intensively developing the processes of information in almost all spheres of human activity. They have led to the formation of a new information infrastructure, which is associated with a new type of social relations, with the new reality, new information technologies of various activities have radically changed and exacerbated the problem of information security[1].

## 2.The issue of security in distributed information - communication systems

Considered in [2] security model implicitly positioned with respect to the information-communications system comprising a unit having a single nucleus, and monitor security. However, the architecture and processes of the functioning of real information - communication systems (ICS) is always characterized by a distribution.

From a security perspective, there are three aspects of the allocation to the ICS:

- software for ICS typical modern modular software and hardware implementation of the kernel, and including a security monitor, whereby the individual security functions are distributed between the software and hardware components of the system;

- in the process of access control security monitor uses the associated data object that contains information on policies differentiation of access to a specific ICS, which can be implemented as a single data object, or as a distributed information structure;

- except for such species as isolated individual work stations, modern ICS are systems that combine physically distributed computing unit (computer networks and systems).

The modular nature of the problem determines the ICS kernel interface defenses identification/authentication, initialization processes, actors, memory management, access control, auditing, distributed on separate software modules kernel. The problem of coupling of distributed defense mechanisms is critical to ensure the safety of ICS, but beyond the scope of benefits [3].

If we consider the associated security monitor object containing information on the delimitation of access of subjects to objects of the system, as a kind of database, it is possible to distinguish structural and logical and physical distribution. Structural logic of the distribution of information and the logical structure and the specific model of access control. For example, in the discretion of such object models can be access matrix, which is implemented as one of the system tables in the database management system (DBMS) or as a set of distributed objects in the operating system access control lists. This, however, logically distributed structural elements associated with the security monitor object may reside within the same computer, or be set to the same and physically distributed across different computing settings.

Given the consideration above all CAS policy in the sequel we will adhere to the following interpretation of the concept of "Distributed ICS".

*Definition 1.* Distributed ICS is called a distributed system consisting of more than one local segment, representing an isolated set of subjects and objects of access[7].

It is easy to see that central to the concept of a distributed ICS is a way to isolate the aggregate of subjects and objects available in a separate independent entity, i.e. in the local segment.

We distinguish three ways to isolate a subset of subjects and objects in the local segment:

- grouping of a subset of subjects access based on their creation and management of a common entity, that is a system process;

- localization of a subset of subjects and objects of access under some technical components of ICS;

- assignment of all subjects and objects of a unique identifier in a single information space and the division of the space on the field, spun off the local segments.

The first way is characteristic for selection in the network of subjects and objects ICS, built in the ideology of "Client-Server", especially for its varieties with "thin" clients. In such systems, most of the processes controlled by the main subjects of the access system process in the form of a database server (DB) or application server. Therefore, the issues of security in such distributed systems are provided with a single entity, which functions as the server system. In particular, the processes of collective access to shared data are managed and regulated by a special software component, called a transaction monitor. Transaction monitor with the security monitor objects (SMO) provides all three aspects of data security - confidentiality, integrity and availability[5].

The second way is the most characteristic of computer networks, based on the binding of subjects and objects, or access their general terms of security and access control

information to one or more dedicated computing installations. May also be noted that in practice this way is isolated in most cases one computer installation or network segment.

The third method is in some sense a virtual second alternative method and used in most cases together with it. In particular, in this manner are protected virtual networks and physical location of the individual components of which has independent significance.

The principal and most important of all three approaches is that the localizable segment from the point of view of security and access control is regarded as retained ICS. Thus, in the future will be based on the following axiomatic assumption[5].

*Assumption 1.* In the local segment of the ICS functions common to all subjects and objects of the security monitor that implements a local access control policy.

The subject of access control policies in distributed information and communication systems is to consider the principles of organization and access mechanisms, including inter-segment, which has been implemented at least some common security policy, providing both Inter and inter-segment safety.

## 3.Model of distributed information - communication systems in the processes differentiation of access

Analysis of the practical aspects of the implementation of distributed ICS allows to select from the point of view of security processes are two main types of architecture (Fig.1):

- a system of interacting segments;
- a system of "local segment of the external environment".

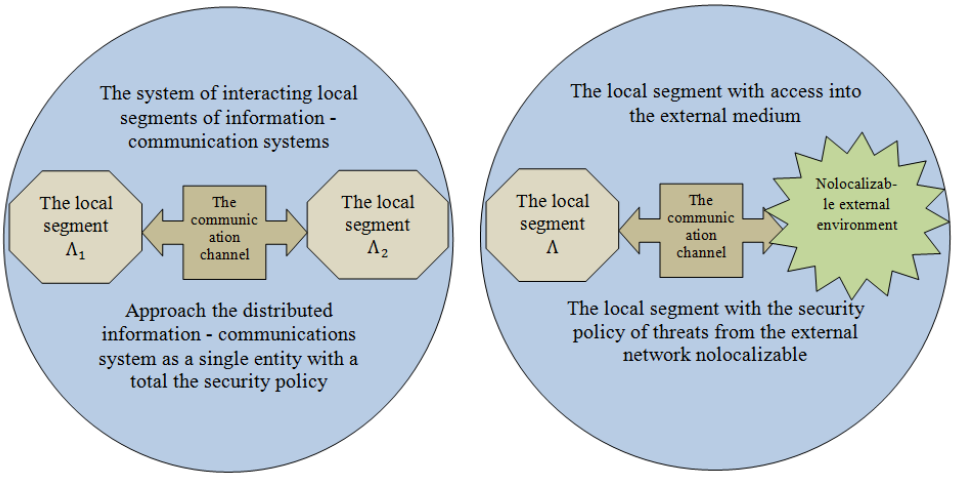



*Figure 1:* Types of distributed IR in terms of safety

The main difference between the first type of distributed systems is a theoretical possibility at any time decomposition of the composition and structure of a finite number of components, which allows capture of the border, and on this basis the set of interacting segments as an isolated distributed system. We note here that most of the architecture of these systems is based on the principle of openness, that is, with the option to add or exclude additional components. However, the possibility at any time fix the limits of distributed systems of the first type allows you to define the boundaries of their safety and on this basis

to build a common security policy in a distributed system, including possibly, but not necessarily having the specificity in the local segments.

In distributed systems, the second type of fix the composition and structure external to the local segment of the party system is fundamentally impossible for technical or organizational reasons. As a result, there is no possibility of defining the boundaries of a distributed system, and in terms of safety, it can only be conducted on the safety of the local segment of the external network, called in this case, the internal network[5].

The communication channel in Fig. 2 implies a special system process st, which provides an interface to local segment "outside world", i.e. segments with other local or external network. Self same interaction is a flow of information in the local segment of the outer segments and vice versa.

Based on the concepts of subject and object ICS, as well as on the notion of subject access to an object, we formulate crucial to the security of distributed systems, the concept of remote access.

*Definition2.* Remote Access $p^{out} = Stream(s_m, o_i) \to o_j$ of user subject $s_m$ in the local segment $\Lambda_1$ to object $o_j$ in local segment $\Lambda_2$ is called a product subject $s_m$ through information and communication subjects $s^{(1)}{}_t$ and $s^{(2)}{}_t$ of local segments $\Lambda_1$ and $\Lambda_2$ flow of information between the object and $o_j$ and $o_i$ of segment $\Lambda_1$.

Omitting some details, the concept of remote access to the subject-object model of ICS can be illustrated by the scheme shown in Fig. 2.

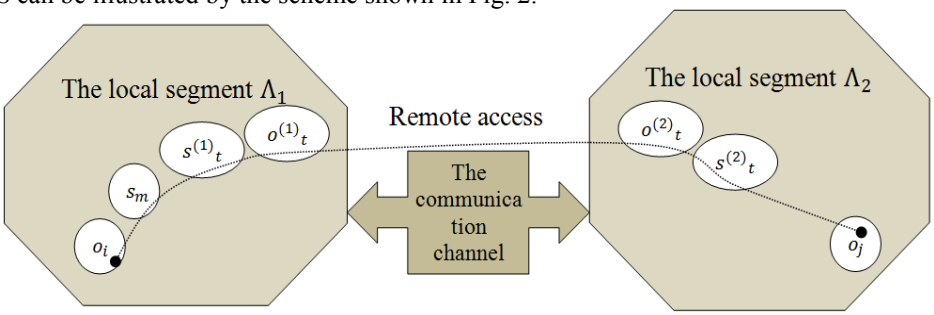



*Figure 2:* The structure of the flow of information for remote access

Legend Figure 2 are: $s^{(1)}{}_t$ and $s^{(2)}{}_t$ information and communication subjects $\Lambda_1$ segments and accordingly $\Lambda_2$; $o^{(1)}{}_t$ and $o^{(2)}{}_t$ associated with the subjects $s^{(1)}{}_t$ and $s^{(2)}{}_t$ information objects; $o_i$ - the object associated with the subject $s_m$.

Note that the data flow caused by the remote access may be either unidirectional, i.e. transfer of data from the local segment to another segment and the other side, to transfer information from another segment in the segment, and bi-directional with simultaneous data transfer in both side.

Another crucial aspect of distributed systems in terms of remote access is implicitly inherent in the *definition 2.* the assumption that the user is a local segment $\Lambda_1$ for some reason can not access the object $o_j$ directly in the local segment $\Lambda_2$. Otherwise there would be a need for remote access anymore. Thus, a reasonable one more axiomatic assumption.

*Assumption 2.* Many users U distributed information - communications system of interacting local segments $\{\Lambda_1, \Lambda_2, ..., \Lambda_K\}$ is the union of disjoint subsets $\{U_1, U_2, ..., U_K\}$ users of the local segments.

It should be noted that, strictly speaking, assuming it is 3, not users, and of their accounts, which are the source object for the generation of the primary subjects of the safety monitor users. There fore, in principle a single user may have multiple accounts distributed over different local segments. In this case, however, for a distributed system of actors one user-initiated on various accounts that correspond to the different users[7].

On the *assumption 3*, we also note that it is often made by users remote access from a local segment to the objects of another local segment, referred to as remote users.

When considering the interaction of the local segments from the perspective of remote access is allocated the top layer of security policy in a distributed ICS - the policy relationship between the local segments, expressed by the paradigm of trust or otherwise trusts.

*Definition 3.* Trust relationships between local segments $\Lambda_1$ and $\Lambda_2$ called part of a system-wide security policy that determines the possibility of remote access users on the same local segment to the objects of another local segment.

You can select the following types of trusts:

- sibling relationships of trust, subdivided, in turn,
  - one-way relationship of trust;
  - bilateral relations of trust;
- hierarchical relationships of trust.

With one-sided relationship of trust, for example, local segment $\Lambda_1$ to the local segment $\Lambda_2$ fundamentally enables remote access users on the local segment of subjects to objects $\Lambda_2$ local segment $\Lambda_1$, but the remote access of users to the local segment $\Lambda_1$ objects local segment $\Lambda_2$ fundamentally impossible. The term "possible in principle" means the need for more detailed and specific regulations - solve specific access under the local segment.

At the bilateral relations of trust are principally possible remote access of a local segment to the objects of another local segment, and vice versa, with the only proviso that the particular subset of the remote access permission should be regulated in the respective local segments.

Hierarchical relationships of trust relate to certain types of distributed systems architecture of the first type, when some local segments are embedded in other local segments. From the point of view of Definition 1, this means that an isolated set of subjects and objects inside local segment is both a subset of the set of subjects and objects outside the local segment. Hierarchical relationship of trust can be called unequal bilateral relations of trust. This means that as the user entities internal local segment is a subset of subjects of external local access segment, no fundamental restrictions for remote access "bottom up", ie, the subjects of the internal segment of the local objects of an external local segment does not exist. Remote access of external local segment of the objects inside local segment are possible in principle, but they should be regulated in the domestic local segment as it allowed remote access outside the local segment.

We can formulate two areas of security in distributed ICS:

- the allocation of a special system of the subject, providing an "external" security;
- to provide a common, including possible with a distributed architecture, security monitor, which provides a consistent security policy of a distributed computer system.

The first direction, derived primarily from the methods of solving security problems in the system "Local segment ICS (internal) - External environment (external)", and subsequently became widespread in distributed systems of interacting local segments as widely known at the present time firewalls. The method of firewall based on the analysis of

information flows and low-level filtering of the many hazardous and non-hazardous according to the criteria defined a priori.

In a broader interpretation in accordance with the definition of 2 remote access is not necessarily accompanied by the creation of a new facility on a particular side access. In addition, and most significantly, the remote access to the object is subject-driven user working in another local segment. In this case, if you go back to the diagram in Fig. 2, the chain "$s_m(\Lambda_1) - s^{(1)}{}_t - s^{(2)}{}_t$" can be interpreted as the subject of an internal access to the local segment $\Lambda_2$ remote user working on the local segment $\Lambda_1$.

Therefore, in the broader context of the second direction of the security of distributed ICS requires consideration of the conditions that allow the initialization of the local segment of the subjects access for remote users [7].

The studied works the distributed systems [5] devoted to the analysis and synthesis of models, safety mechanisms in showed that formal models of safety play an important role in the processes of design, development, certification and research protected ICS as provide the technical approach including the solution of the following major tasks:

- confirmation of properties of the developed systems by the formal proof of safety of system;

- drawing up the formal specification of a security policy as most important component of organizational and documentary support developed protected ICS.

Merits and demerits of formal models of management of access, proceeding from possibility of their application at creation of system of differentiation of access to ICS are given below.

| Name of model | Advantage | Shortcoming |
|---|---|---|
| **1** | **2** | **3** |
| Discretionary model | 1. Simplicity of realization;<br>2. Allows to operate powers of users to within operation over object;<br>3. The criterion of safety of this model is very strong on the practical level as allows to guarantee inaccessibility of a certain information to users who were initially not given out appropriate authority. | 1. In the general statement doesn't give guarantees of safety of system;<br>2. It is vulnerable in relation to attack by means of «Trojan Horse»;<br>3. Administration of system with a large number of objects and subjects of access is rather labor-consuming task. |
| Mandatory model | There is a formal proof of safety of system. | 1. All interactions are considered only at rather high level of abstraction at which details of realization of operations of access aren't considered;<br>2. Impossibility of the accounting of all requirements of model;<br>3. Absence in higher education institutions of strict classification of information on confidentiality levels. |
| **1** | **2** | **3** |
| Model of | 1. There is a formal proof of | We will apply only to a private task. |

| | | |
|---|---|---|
| thematic differentiation of access | safety of system;<br>2. Allows to organize access to information resources with difficult hierarchical structure. | |
| Model of role differentiation of access | 1. The role model doesn't allow one session to intensify other session;<br>2. The restrictions imposed for sets of roles for which the user can be authorized or on which he will become authorized during one session;<br>3. Provides bigger compliance to the technologies of information processing used in computer systems;<br>4. The account difficult organizational and regular structure with a large number of employees. | 1. Problem of hierarchy and globality of roles;<br>2. Absence of the owner of object;<br>3. Lack of limited completeness of functions of the user.<br>4. In role models there are no strict formal proofs of safety of system. |

*Table 1:* Results of the analysis of models of management of access

## 4.Conclusion

In summary it is possible will note that the models of safety of a network based on a combination of the discretionary, mandatory and role principle of differentiation of access a set of objects of system includes besides information objects and a set network a component having, including the classification. The structure of a set of operations included operations on establishment of communication of subjects with remote components of a network. Performance of operations is followed by passing of two phases by system - phases of access to system and communication establishment phases. On the basis of discretionary, mandatory and role mechanisms the conditions and restrictions imposed on operations at which safety of a network is ensured are considered.

## References

1. Zegzhda D.P., Iwashko A.M. Basics of security of information systems. – M.: Redline - Telecom, 2000.
2. Ganiev S.K., Irgasheva D.Y., Abramov A.S. The conceptual model of role-based access control for computer networks // International Conference on IT Promotion in Asia 2011 / Tashkent University of IT, 2011.
3. P.N. Devyanin, O.O. Michalski, D.I. Pravikov and other. Theoretical Foundations of Computer Security: Textbook for higher education institution - M.: Radio and Communication, 2000. – 192p.
4. Shcherbakov A.Y. Introduction to the theory and practice of computer security. Moscow: Publisher Molgachev S.V. - 2001.
5. Gaydamakin N.A. Restriction of access to information in computer systems. - Yekaterinburg edition of the Urals. University Press, 2003.
6. Court S.S. The theoretical framework for the protection of information: the manual. - M.: Helios ARV, 2004.
7. Devyanin P.N. Security models of computer systems: Manual allowance. - Moscow: Publishing Center “Academy”, 2005.

# Minimizing in Face Recognition Errors and Preprocessing Time

**Karimov Madjit Malikovich, Tashev Komil Axmatovich, Islomov Shahboz Zokir ugli, Khudoykulov Zarif.**

Tashkent University of Information Technologies, Uzbekistan.
zarif.xudoyqulov@mail.ru

**Abstract.** This paper describes two methods for face recognition. One of them is used to decrease image-repairing time for face recognition and second method is used to increase usefulness rate of face recognition methods. As experiments show, working default image (for example, with RGB model) is required more time. In this case, using just using shape and texture information for improving facial Recognition**.** Separately using face parameters in face recognition is not good way to recognize face. So calculate face parameters based on other parameters is efficiently way in face recognition.

***Keywords:*** face recognition, active shape models, pattern recognition, maximum-likelihood, principal component analysis.

## 1. Introduction

Face recognition is one of the most relevant applications of image analysis. It's a true challenge to build an automated system which equals human ability to recognize faces. Although humans are quite good identifying known faces, we are not very skilled when we must deal with a large amount of unknown faces. The computers, with an almost limitless memory and computational speed, should overcome human's limitations. The systems have to deal with varying face size and location in the frame, changing facial expressions, and a non-constant lighting. The systems were built on a regular PC using a simple video camera [1].

The difficulty in the face detection process stems mainly from a large number of potential face candidates per frame. The classification of each candidate is not negligible, and in order to reach desired detection and false alarm rates, an even more complex preprocess is needed.

## 2. Current face recognition methods and their problems

Automated face recognition is a relatively new concept. Developed in the 1960s, the first semi-automated system for face recognition required the administrator to locate features ( such as eyes, ears, nose, and mouth) on the photographs before it calculated distances and ratios to a common reference point, which were then compared to reference data. In the 1970s, Goldstein, Harmon, and Lesk used 21 specific subjective markers such as hair color and lip thickness to automate the recognition. The problem with both of these early solutions was that the measurements and locations were manually computed. In 1988, Kirby and Sirovich applied principle component analysis, a standard linear algebra technique, to the face recognition problem. This was considered somewhat of a milestone as it showed that less than one hundred values were required to accurately code a suitably aligned and normalized face image. In 1991, Turk and Pentland discovered that while using the Eigen

faces techniques, the residual error could be used to detect faces in images; a discovery that enabled reliable real-time automated face recognition systems. Although the approach was somewhat constrained by the environmental factors, the nonetheless created significant interest in furthering development of automated face recognition technologies. The technology first captured the public's attention from the media reaction to a trial implementation at the January 2001 Super Bowl, which captured surveillance images and compared them to a database of digital mug shots. This demonstration initiated much-needed analysis on how to use the technology to support national needs while being considerate of the public's social and privacy concerns. Today, face recognition technology is being used to combat passport fraud, support law enforcement, identify missing children, and minimize benefit/identity fraud. As one of the most successful applications of image analysis and understanding, face recognition has recently gained significant attention. Over the last ten years or so, it has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding.

This modern approach generated more proficient results in all aspects including mean square error, recognition rate and training time. They presented that single layer neural network can give 100% recognition accuracy if correct learning rate is assigned to it. They also proposed that BPNN is more sensitive to hidden neurons and learning rate [4].

A recognition method was suggested by Jong-Min Kim et al. [5], to improve the recognition rate that combines the multi-layer neural network with the method that computes the small number of Eigen-face feature vector through PCA. The proposed method showed the improved functioning with variations in illumination and noise so that it could minimize the recognition error. In the paper, the proposed face recognition system algorithm used the Automatic Facial Land-marking using Active Shape models to extract the face from the background image and after that it performed the face recognition by reducing image size and tracking facial landmarks from the image, then calculated the proper vector through PCA and finally applied the multi-layer neural network on the resultant image. Thus the performance of existing PCA matching method was compared with the combinational approach of PCA and MLNN.

Four baseline face recognition algorithms have been developed. They are [2]:

- A standard PCA or Eigen-faces algorithm;
- A combination PCA and LDA algorithm based upon the University of Maryland algorithm in the FERET tests;
- A Bayesian Intrapersonal/Extra-person Image Difference Classifier based upon the MIT algorithm in the FERET tests;
- An Elastic Bunch Graph Matching Algorithm that uses localized landmark features represented by Gabor jets. This algorithm is based upon the USC algorithm in the FERET tests.

## 3. Automatic Facial Land-marking using Active Shape Models (ASMs)

This is my primary research area. I am looking into ways of improving the fitting accuracy of ASMs to enable the automatic annotation of unseen faces. ASMs are deformable templates that can be trained to automatically position a pre-defined set of landmarks along the contours of an object of interest. They are most commonly used for facial land marking and if initialized well, using a suitable face detector, produce fairly accurate results.

Currently my implementation works fairly well with frontal faces and can automatically annotate them with a dense set of 79 landmarks. I am working towards extending it to handle pose and expression variation. There are several potential applications of this tool including face recognition (accurate registration can be made possible), expression analysis, pose estimation using regression models, generation of 3D facial models etc.

This model was able to apply ASMs to the task of tracking the previously mentioned 79 facial landmarks across the frames of different video sequences in which the subject showed rapid in-plane rotation and out of plane pose variation. This task was accomplished by Kalman filtering the landmark coordinates. The predictive mechanism of the Kalman filter ensured accurate initialization of the ASM on the next frame (without the need for face detection) while its corrective mechanism, treated the landmark locations produced by the ASM as noisy observations that were refined to produce more accurate results. We benchmarked our approach against naive methods that 1) did not harness any temporal information and treated each frame independently 2) initialized the ASM on frame n+1 using the fitting results on frame n but without Kalman filtering and found that our approach produced far lower fitting error. The applications of this work include tracking and facial recognition in surveillance footage. Our system is not currently capable of operating in real-time but can be used for post processing [9].

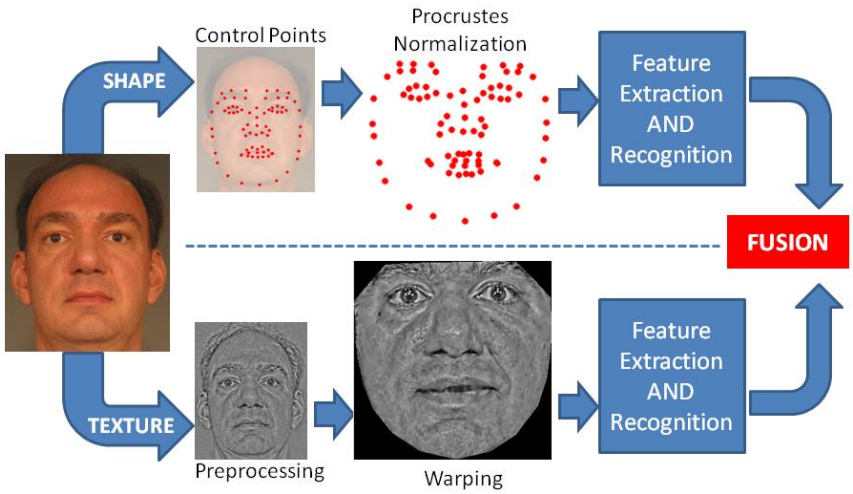



Fig.1. Using shape and texture information for improving facial Recognition

Conventional face recognition algorithms simply crop out the facial region, extract features from the crop and match based on these features. We aimed at a different approach to harness more information from images and hence boost facial recognition rates. Using our dense land-marking scheme, we performed experiments in which the shape information contained purely in the locations of these landmarks was harnessed while shape-free texture information was harnessed separately by warping all faces to a common mean face. Purely shape based features and purely texture based features were then used for matching and the results were subsequently fused. The results we obtained demonstrated that 1) purely shape based recognition is fairly discriminative 2) significant performance improvement can be gained by better registration of texture using shape information and 3) fusing shape and texture based methods consistently boosts the performance of subspace-based approaches such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) etc. To

the best of our knowledge, this is the first large scale study of the roles of shape and texture in facial recognition as we carried out our tests on the vast and challenging NIST Facial Recognition Grand Challenge (FGRC) version 2.0 database.

If detected faces are represented by the centers of the eyes (fig.2.), let's consider them to be correctly detected, if and only if detected eyes belong the area around the true eyes location with the diameter $D_{\text{Eyes}}$. Which depends on the distance between eyes' centers and *a,* has been taken equal to 0.25 (This criterion was originally used by Jesorsky et al.), and calculates as $D_{Eyes} = 2a \times l_{Eyes}$.

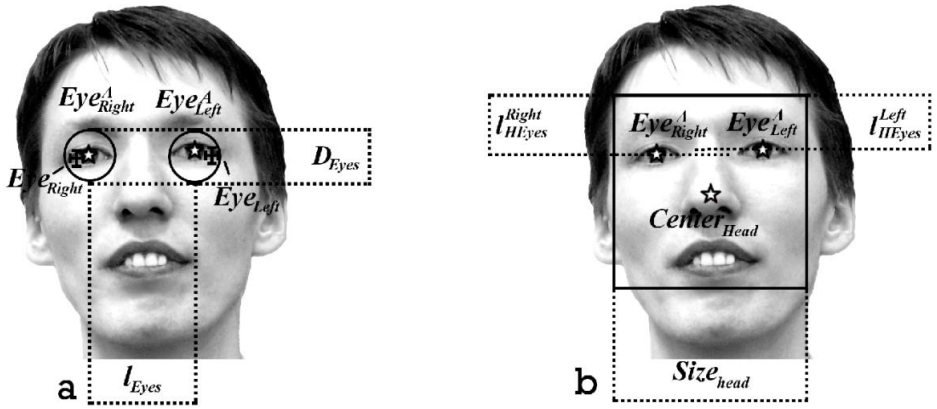



Fig.2. Schematic face representation

Assume there is a full face portrait image with no incline (fig.2.b), and the algorithm has found its center and size - *($Center_{Head}$* and *$Size_{Head}$* respectively). Obviously, the eyes on this image are located symmetrically about the vertical axis (i.e., at the half the distance between them: *$l_{Eyes}/2$)* and at the same distance *($l_{HEyes}$)* from the top border of the face's rectangle. Thus the absolute coordinates of eyes can be estimated as:

$$\text{Eye}_{\text{Right}} = \text{Eye}_{\text{Left}} = \text{Center}_{\text{Head}} + l_{\text{HEyes}} - \text{Size}_{\text{head}}$$
$$\text{Eye}_{\text{Right}} = \text{Center}_{\text{Head}} - 2 * l_{\text{Eyes}}$$
$$\text{Eye}_{\text{Left}} = \text{Center}_{\text{Head}} + 2 * l_{\text{Eyes}}$$

Let's try to estimate the parameters of the algorithm, namely $l_{Eyes}$ and $l_{HEyes}$, as an average of the huge amount of images with experts' labeled eyes. Based on such analysis, the following coefficients have been founded: *A* - average proportion of distance between top border of the face and center of the eyes *($l_{HEyes}$)* to the size of the face rectangle; and *B* - average proportion of the distance between eyes *($l_{Eyes}$)* to the size of the face rectangle *($Size_{Head}$).*

It should be noted that such "conversion" of face representation could deteriorate the localization accuracy for algorithms describing faces by rectangles.

**Pattern Recognition.** Due to variants such as viewing angles, illumination, facial expression and so on, the facial feature vector obtained from previous equations can have random variations and therefore it is better modeled as a random vector. If the incoming person is equally likely to be any person in the database (equal a priori probability), then according to Bayes decision theory, the minimum recognition error rate can be achieved if the recognition is following the *maximum-likelihood (ML)* criterion. That is, suppose Y =

f(X) is the feature vector and suppose that there are K persons in the database, the identity of the incoming person is assigned by

$$K_0=\arg\min \log p(Y/k),\ 1\le k\le K,$$

Where p(Y/k) is the likelihood density of Y conditioning on its being the *k*th person.

If we assume the variations in the facial feature vector are caused by zero-mean, additive white Gaussian noise (AWGN), then the ML matching becomes the common minimum distance matching. That is, the identity of the incoming person is *k* if the Euclidean distance between the feature vector of the incoming person and the mean vector of the *k*th person is the smallest among all people in the database. Unfortunately, in the real world situation the variations of facial images are much more complex than AWGN. In the following section we will describe various kinds of variants that may appear in the face recognition problem.

## 4. Conclusion

In this paper, we considered relevant issues related to one sample per person problem in the area of face recognition. We focused mainly on recognition efficiency of several methods working with single and multiple samples per subject. We researched techniques for enlargement of the training set by new (artificial, virtual or nearly synthetic) samples, in order to improve recognition accuracy. Such samples can be generated in many ways – we concentrated on modifications of the original samples by noise. We examined the impact of these extensions on various methods.

## 5. References

1. R.Chellappa, C.L.Wilson, S.Sirohey, Human and machine recognition of faces, Proceedings of IEEE (2010). 705-740p.
2. P. Chandra, M. Chandra, M. Kumar, B. Latha "Multi scale feature extraction and enhancement using SVD towards secure face recognition system," in Proc. International Conf. on Signal Processing, Communication, Computing and Networking Technologies. July 2011. 64 – 69pp.
3. Raman Bhati, Sarika Jain, Nilesh Maltare, Durgesh Kumar Mishra, "A Comparative Analysis of Different Neural Networks for Face Recognition Using Principal Component Analysis, Wavelets and Efficient Variable Learning Rate," Computer & Communication Technology(ICCCT), IEEE Int'l Conf on, 2010, pp. 526 – 531.
4. Jong-Min Kim, Myung-A Kang, "A Study of Face Recognition using the PCA and Error Back-Propagation," Intelligent Human-Machine Systems and Cybernetics, IEEE Second International Conference on, 2010, pp. 241-244.
5. http://en.wikipedia.org/wiki/Facial_recognition_system. "Facial recognition system".

# Configure Firewalls to Implement Special Filtering Modes

**Madjit Karimov[1], Sherzod Gulomov[2], Bakhtiyor Babamukhamedov[3]**

[1]A.Temur street 108, 100202 Tashkent, Uzbekistan
dr.mmkarimov@rambler.ru

[2]A.Temur street 108, 100202 Tashkent, Uzbekistan
sherzod.gulomov@rambler.ru

[3]A.Temur street 108, 100202 Tashkent, Uzbekistan
abbts@mail.ru

**Abstract.** This article describes the structure and size of packet an Ethernet, which are determined by type and standard network protocol and investigated Manchester encoding, representing periodic wave signal in the physical environment. The model of virtual transport connection, enabling the identification of traffic parameters and influence the fractal characteristics of the process. Were analyzed Hurst exponent, which is invariant to the characteristic scale of measurement that his continued input and output Firewall will be provided condition for complete with a special filtering mode. Proposed a parametric model in Firewall special filtering mode, allowing provide a range values of buffer size and performance.

**Keywords:** Queue system, special filtering mode, MAC-address, the factor of Hurst, maximum transfer unit-MTU**.**

## 1. Introduction

In the implementation of packet-based traffic filtering factor limiting the performance of Firewall, is performance of packet processing system that operates in accordance with the filtering rules. To evaluate the performance of network is commonly used parameter bits/s. When processed in firewall packet traffic, this parameter is not applicable. For example, a network device that handles 1500 byte packets at a rate of 100 Mbit/s can be handled packets of 40 bytes at a speed of 10 Mbit/s. In the first case processing accounts for approximately 8000 packets, in the second case, the traffic of about 32000 packets per second that is four times longer. Therefore, describing the intensity of the input and output flow must use the integral characteristic performance Firewall, measured as the number of packets per unit time calculated for a fixed packet size.

## 2. The structure of Ethernet packet

Functioning of a computer network consists of sending user data between workstations and servers. All messages sent through the network, broken into fragments of standard length [1]. The structure and size of the packet are determined by type and standard network protocol. In figure 1 shows the structure of packet in the Ethernet.

802.3 standard defines eight header fields:

Field preamble consists of seven bytes of sync data. Each byte contains the same bit sequence - 10101010.

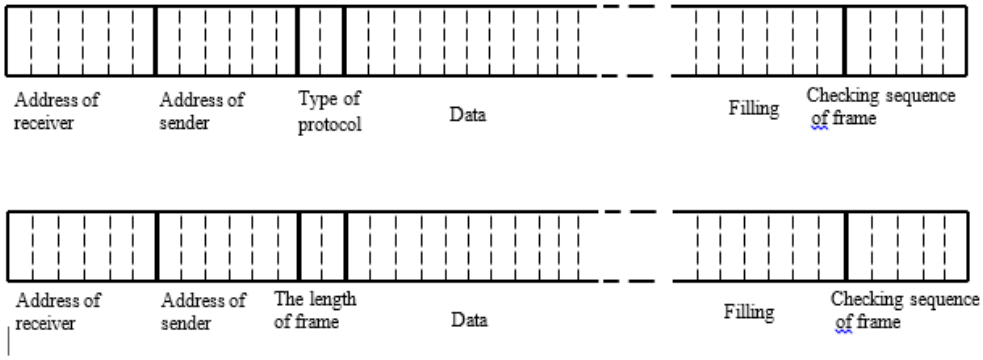



**Fig.1.** The structure of Ethernet II and IEEE 802.3 packets

With Manchester encoding, this combination appears in the physical environment periodic wave signal. Preamble is used to provide the time and opportunity scheme of a transceiver (transceiver) to come to a stable matching with the received clock signal.

1. Start frame delimiter consists of one byte with bit set 10101011. Appearance of this combination is an indication for the upcoming frame reception.
2. Address of receiver - can be as long as 2 or 6 bytes (MAC-address). The first bit of the destination address - this is a sign, is to address the individual or group, if 0, the address points to a specific station, if 1, the multicast address is a number (possibly all) of the network stations. When a broadcast address, all bits of the address field set to 1 is generally accepted that the use of 6-byte addresses.
3. Address of sender- 2 or 6-bytes of field, which containing the address of the sender station. The first bit - always has the value 0.
4. Two-byte length field specifies the length of data field in the frame.
5. The data field may contain from 0 to 1500 bytes. But if the field length is less than 46 bytes, then use the following field - a field filled to complement the frame to the minimum allowable length.
6. Field filling consists of an amount of bytes fillers, which provides a certain minimum length of the data field (46 bytes). This ensures the correct operation of the mechanism of collision detection. If the length of the data field is sufficient, then the field of filling in the frame does not appear.
7. Checksum field - 4 bytes containing the value that is calculated by a certain algorithm (polynomial CRC-32). After receiving the frame workstation performs proper checksum for that frame, compares the obtained value with the value of a checksum field, and thus determines corrupted or not the received frame.

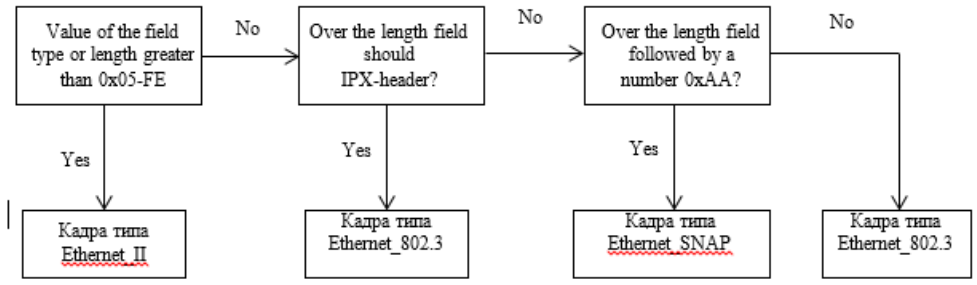



**Fig.2.** Algorithm for determining the format of an Ethernet frame

## 3. Method of calculating the invariant characteristics for a special filtering mode

In general a model Firewall in special filtering mode can be represented [2] in the form of buffer (figure 3).

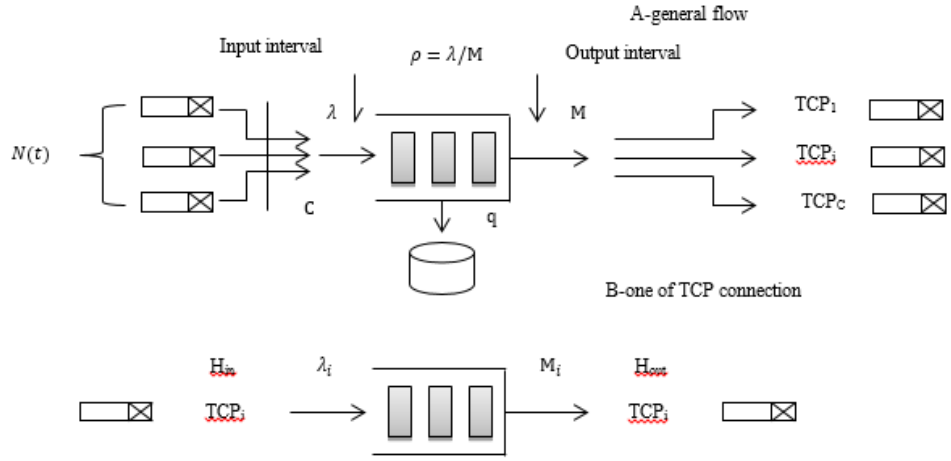



**Fig.3.** Generalized model of Firewall with two physical interfaces: A - general flow B-one of TCP connection

In figure 3 following notation: $M$ –the total intensity of the processing of all TCP connections (packet/sec), $\rho = \lambda/M$ –utilization, $q$ –buffer size (number of packets), $C$ –number of TCP connections, the existence of which allows to estimate the parameters $H, H_{in}$ и $H_{out}$ – indicators Hurst in input and output for Firewall allowed TCP connections, $\lambda_i$ – arrival rate for the *i*-th allowed TCP connections, $q_i$ – buffer size allocated to the *i*-th TCP connection and $m_i$ – treatment intensity of *i*-th TCP connections.

## 4. Parametric conditions for special filtering mode

When considering data in the transport layer model OSI, the interaction between the source and receiver data via a virtual connection. Although physically, network [3], through

which interaction may consist of a plurality of intermediate nodes: routers, switches, etc., equipment, including firewall and have a certain value of the Hurst exponent $H$ (figure 4). From the experimental data it is known that for WANs value $H$ −aggregated flow is in the range 0.6÷0.8.

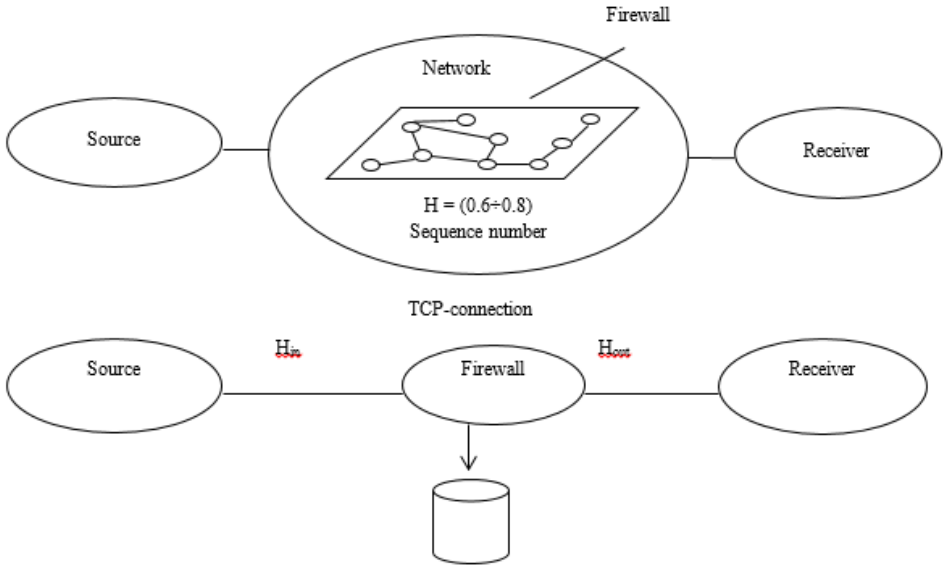



**Fig.4.** Displaying a virtual connection to a physical network

If the configuration of filtering rules Firewall will allow packets from the source to the receiver, then this virtual connection will be characterized by some $H$. Special mode filtering will be provided by the immutability of $H$ input and output Firewall for this virtual connection.

## 5. Procedures and conditions for implementation of a special filtering

When processing the input fractal process with the network device utilization factor $p$ and the Hurst exponent $H$ there is a level buffer size $q$, which does not happen to be dropping packets and Hurst exponent $H$ on the input and output Firewall remains unchanged. Queue system, taking into account fractal properties of network processes, there are increased demands on the buffer (figure 5), so to calculate its size using the ratio obtained with the diffusion approximation of the input flow applications.

This relation is as follows: $q = \frac{p^{1/2(1-H)}}{(1-p)^{H/(1-H)}}$ (1),

where in $q$ −buffer size, $p$ −utilization factor, $H$ −Hurst.

Given that the traffic handled by Firewall, is a set of packets from a variety of transport links, the above method of calculating [4] the size of the buffer are encouraged to apply for individual transport connections allowed to pass through the Firewall.

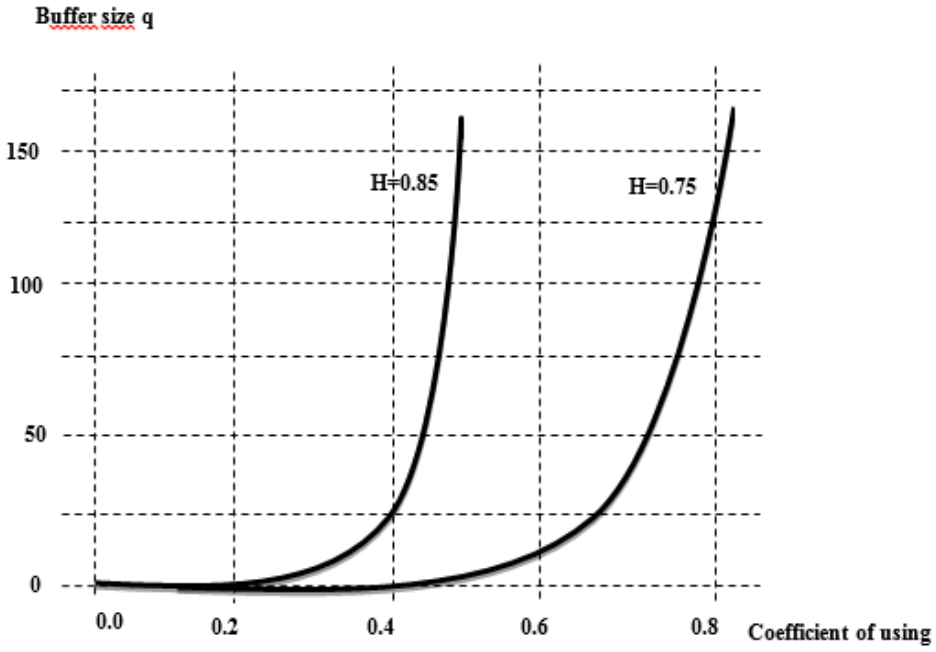



**Fig.5.** Dependence of buffer size from coefficient of using

## 6. Evaluation of the accuracy for performing the special filtering

To obtain acceptable values of the parameters $m$ and $q_i$ of Firewall need of expression (1) to obtain the dependence of the Hurst exponent $H$ for various values $\lambda_i$ −intensity, $q_i$ −buffer sizes and $m$ −performance (figure 6).

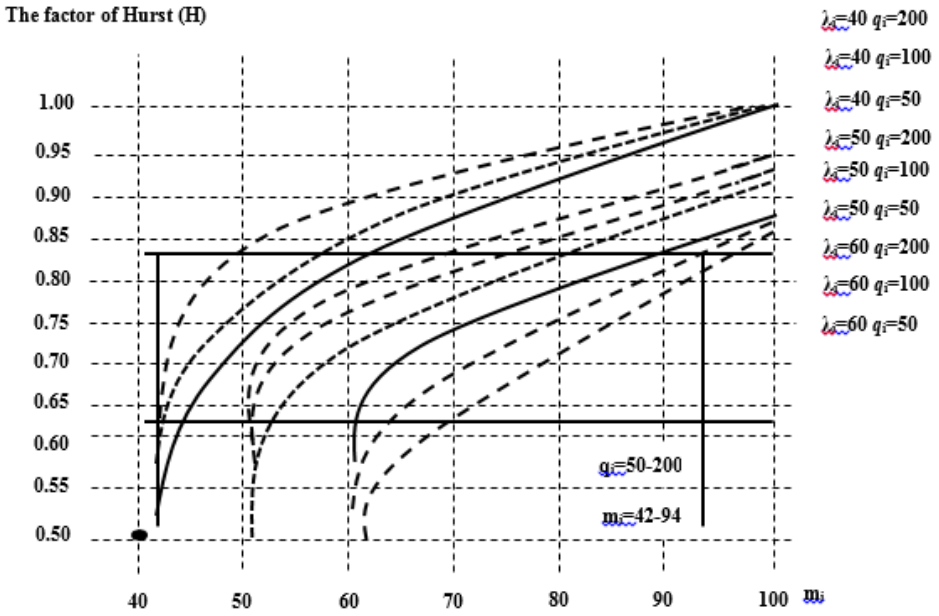



**Fig.6.** The dependence of $H$ from parameters settings of Firewall

In figure 6 following notation: $\lambda_i -$ load intensity of $i$ −th TCP connection in relation to the capacity of the physical channel, $m_i$ −processing performance of the $i$ −th TCP connections in Firewall, $q_i$ −buffer size in packets allocated to the $i$ −th connection [5].

By plotting the Hurst exponent $H$ for various $\lambda_i, q_i$ and $m_i$ defined ranges of $m$ and $q$ for the $i$ −th allowed TCP connections. In particular, in the case of network interfaces Firewall 100 Mbit/s mode special filtering will be provided at $q_i$ =50-200 packets and $m_i$ =4200-9400 pack/sec, which corresponds to the performance 6,3-14,1 Mbit/s.

In figure 7 represented by the choice of parameters, taking into account Firewall proposed model.

Firewall: $H_{in} = H_{out}$ at $q = 50 - 200$ packets and $m$ =14000-16000 pack/sec.

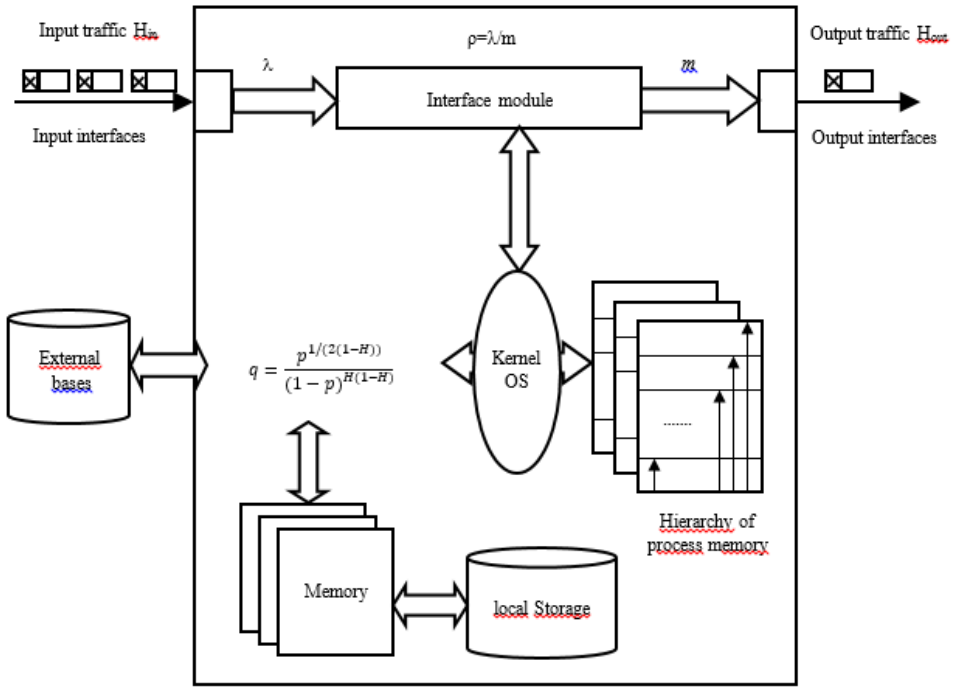



**Fig.7.** Firewall with the proposed model

## 7. Terms of special filtering mode

The packets are transmitted to the network card and stored in the buffer. When the buffer is filled, the card generates an interrupt and the NIC driver copies the data from the buffer card (mbufs) in kernel memory. Once a packet is transferred to the mbuf, execution time of all subsequent operations carried out with the packet does not depend on its size, as analyzed only its title. If you want to skip the packet, the packet will be given to the driver of the network card which will extract it from the mbuf and pass the line.

Most of these operations have a relatively high price per one packet, but a very low cost on the basis of packet size. Therefore, the processing of a large package is only slightly more expensive than processing a small packet.

Some restrictions are imposed by hardware. For example, the machine handle class x86-64 not more than 10,000 interrupts per second, regardless of the processor speed, which is caused by the limitations architecture. Some network adapters generate one interrupt for each packet. Consequently, the node starts losing packets when their amount exceeds about 10,000 packets per second. Other maps, for example, are more expensive, have large built-in gigabit buffers, which allows them to bind multiple [6] packets in a single interrupt. Therefore, the selection of hardware may impose some limitations on performance.

In environments transmitting Ethernet maximum frame size which can be transmitted or received adapter is 1538 bytes, which comprises:

- start of a frame, 8 bytes;
- Ethernet header, 14 bytes;

- data up to 1500 bytes;
- checksum 4 bytes;
- packet interval, 12 bytes.

The controller can send and receive Ethernet frames are complete:

• for 1 Gbit/s - every 12.3 microseconds, or approximately 81,000 frames per second (1,000,000,000/1538/8 ~ 81000);

• for 100 Mbit/s - every 123 microseconds or about 8100 frames per second (100,000,000/1538/8 ~ 8100).

During normal operation not all network packets are full size, so the actual number may be much larger value. Packet processing for such velocity requires considerable productivity, so the effectiveness of the physical layer is dependent not only the rate of transmission on network, but also the state of the whole system.

The work of network interface can be divided into 2 steps – this is transmit/receive packets and placing them in the buffers [7]. Both of these processes are interrelated - before the packet is sent to the network, first it is placed in the buffer of the network card in the case of receiving a packet from network, vice versa.

Driver allocates buffers in physical memory, which stores the network card newly arrived packets. To determine the size of the allocated memory is used, as a rule, two parameters - the number of buffers (one buffer - one packet), which are defined in the configuration of network card, and the maximum size of the transmitted segments (maximum transfer unit-MTU). The last parameter is used by the driver to determine the amount of memory you want to allocate a single buffer. If the MTU is not used, it may happen that the allocated buffer is less than that received packet, or the allocated memory will be greater. For example, some network adapters for MTU 1500 allocate 2048 bytes. So, if you set the number of buffers in 5000 for incoming packets, the driver will allocate about 10 MB of memory.

## 8. Conclusion

The proposed parametric model allows the Firewall to obtain ranges of buffer size and performance, in which the Firewall will provide a special filtering mode. Ranges of values obtained by analysis of the dependence of the Hurst exponent *H* as a function of performance for different values of buffer size and intensity.

## References:

1. F. Schneider, J. Wallerich and A. Feldmann, "Packet capture in 10-gigabit Ethernet environments using contemporary commodity hardware", (2007).
2. G. A. Cascallana and E. M. Lizarrondo, "Collecting packet traces at high speed", Proceedings of Workshop on Monitoring, Attack Detection and Mitigation (MonAM), (2006).
3. L. Deri and F. Fusco, "Exploiting commodity multi-core systems for network traffic analysis" (2009).
4. Cristian Estan, Stefan Savage, George Varghese "Automated Measurement of High Volume Traffic" // ACM SIGCOMM Internet Measurement Workshop 2012.
5. R.S. Liptser and A.N. Shiryaev, Statistics of Random Processes. New York: Springer-Verlag, 2004.
6. M. Christiansen, E. Fleury. Using IDDs for Packet Filtering. BRICS Report Series, RS-02-43, October 2010.
7. E. W. Fulp, "Optimization of Network Firewalls Policies using Directed Acyclic Graphs", IEEE Internet Management Conference, 2005.

# Models of Network Processes for Describing Operation of Network Protection Tools

**Madjit Karimov[1], Abdukhalil Ganiev[2], Sherzod Gulomov[3]**

[1]A.Temur street 108, Tashkent, Uzbekistan
dr.mmkarimov@rambler.ru

[2]A.Temur street 108, Tashkent, Uzbekistan
a.ganiev@tuit.uz

[3]A.Temur street 108, Tashkent, Uzbekistan
sherzod.gulomov@rambler.ru

**Abstract.** In this article is considered the model on the basis generalized random processes modulated by setting the parameters of random process generating load, It was brought a discrete message source with memory, that generated a message in discrete time and studied fractal models, allowing generate self-similar traffic, creating on the so-called chaotic mappings (Chaotic map-CMAPs). It was offered a method on the basis Markov modeling, let constructed initial model, consisting of a group states corresponding to each registered symbol of input flow and were analyzed process steps of measurement traffic on computer networks, which include data on circulating in network packets, that generated by different protocols.

**Keywords:** Data send networks-DSN, Generally modulated process-GMP, Chaotic map – CMAP, Generally modulated deterministic process – GMDP, Markov modulated process – MMP, Time To Live – TTL.

## 1. Introduction

Recently, in data send networks (DSN) with packet commutation is increased tendency expand of volume data traffic, which leads to the necessity of careful planning of network resources when building and operating DSN. The solution of such problems is impossible without building models that at minimal cost to evaluate efficiency of the test network. The modern DSN is a complex set of elements that interact with each other on different logical levels. In general, the DSN model should include a plurality of patterns, operating in a network environment. Models of individual elements of network are created on the basis analysis of algorithms for their work, as well as the results of testing real devices, communication channels and other elements of network infrastructure. An integrated approach to modeling the DSN calculation performance targets are made on the basis of the initial data on network load (traffic), generated by all users of network resource. Therefore, the importance of adequate modeling of network traffic is difficult to overestimate, as the information obtained as a result, is then used as input data for the calculation of algorithms used in the network model.

## 2. The purpose of the task

During operation of modeling network traffic may cause various problem situations:

- company's employees are unhappy with the long delays in the delivery of operational posts in the process of technological operations;
- remotely serviced customers express dissatisfaction with unstable information services;
- what kind of hardware are overloaded and need to be replaced?
- what changes should be made in the allocation of functions and organization of the servers?
- how to reduce the payment for traffic on leased communication channels?

One of the most effective methods for removal appearing problems and in answering these questions is to measurement traffic characteristics and definition of quantitative estimates on them to ensure informed decision-making.

To resolve problems that arise need to perform long-term measurements of traffic:

- traffic, measured in bits per second, and in frames per second, will reveal the monthly, weekly, diurnal patterns of change in load and set the intervals of high load. It is during these intervals typically arise failures in service users, particularly in those cases where there are "network bottlenecks";
- during the measurements, it is necessary for each transmitted frame data link layer to capture and save a set of parameters of the channel, network, transport and application levels. Only collected in real time the dump with a complete set of parameters may enable traffic analysis with the necessary detail;
- dump collected should be analyzed in several stages. At each stage narrows the range of temporal and spatial analysis and research increased detail.

This approach allows us to obtain quantitative estimates:

- temporary delays in the delivery and processing of data;
- the causes delays in excess of allowable limits;
- the causes erratic performance.

It should be noted that the study of modeling network traffics a number of difficulties, which include:

- the high volume of traffic at high speeds metabolism;
- random character of traffic, which depends both on the technological processes corporations and subjective characteristics of users;
- the need to define generalized, integrated numerical values dynamically changing random events on long-term range of operation, taking into account that through the test unit for 1 second can move a large number of transactions, the number of which can be measured in the thousands.

Thus, the task of creating universal models of traffic network applications is one of the key objectives in the development of modeling DSN.

Methods for modeling of network traffic conceptually can be divided into two classes - the analytical and simulation. Analytical modeling involves a formal description of the simulated objects and processes in the form of a set of mathematical equations and expressions. These models are useful for theoretical studies and formal manipulations, however in most cases the development of an adequate analytical model for many types of network traffic is a virtually impossible task. In that case, if modeling simulation aims to calculate (estimate) performance (parameters) of modeled system, the most preferred is use simulation models. Simulation models are a set of algorithms (usually implemented by

software), which step-by-step reproduce the events occurring in real system. The essence of traffic simulation is to create an algorithm (method) that let be would generate a sequence of values similar with sequence values for traffic intensities, observed in researching of DSN.

## 3. Methods for modeling of network traffic

Nowadays might select four classes of models used to modeling of network traffic (figure 1):

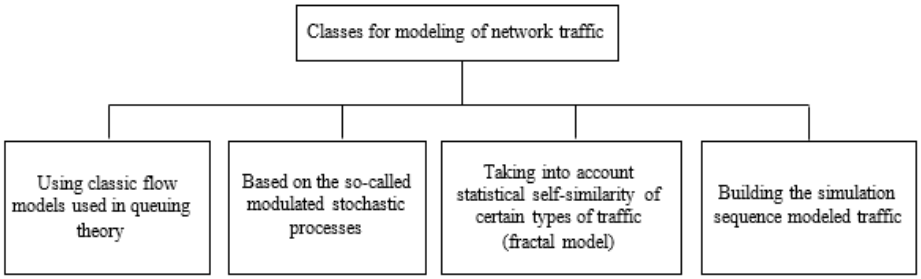



**Fig.1.** Classes for modeling of network traffic

The most promising today considered a model based on generalized random processes modulated (Generally Modulated Process - GMP) and fractal model based on chaotic maps [1] (Chaotic Map - CMAP).

## 4. Model based on generalized random processes modulated

On basis GMP lies of idea for control modulation distribution law with help of the auxiliary stochastic process. The source can be in one of a plurality of states, which set parameters of the random process generating the load [2]. Transition between states is caused by an additional modulating random process. Quotient event of class GMP are generalized processes with deterministic modulation (Generally Modulated Deterministic Process - GMDP) and processes with Markov-modulated (Markov Modulated Process - MMP). In GMDP models traffic source can be in one of $N$ states. The system remains in state $i$ in the current time interval $t_i$, wherein the source generates a load intensity $\lambda_i$. The values of $t_i$ distributed with geometrical law. In MMP models controlling process is a Markov process $M = \{M(t)\}^{\infty}_{t=0}$ with continuous time and discrete set of conditions$\{0,1 \dots m\}$. In this case, the MMP, the current law distribution for packets is completely determined by the current condition $k$ of processes $M$.

Also find application models (Markov Modulated Poisson Process - MMPP), representing a variety of processes and Markov modulated [3] source which generates a Poisson load with intensity $\lambda_k$, when the sub-process $M$ is in condition $k$.

## 5. Important variety of MMPP is MMPP-2

In MMPP-2, there are only 2 states - one of which corresponds to zero intensity of source. On depending from condition of sub process $M$ source generates a Poisson process, or inactive (figure 2).

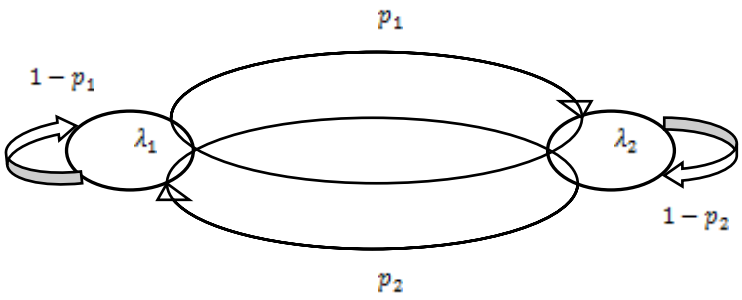



**Fig.2.** Model MMPP-2

where, $p_1$, $p_2$ – the transition probability; $\lambda_1$ and $\lambda_2$ intensity of a Poisson flow generated source located in the respective states [4]. Allowed value of either zero intensities. Recent research networks with commutation packet was being spoke about the statistical self-similarity of certain types of traffic and the effect of the presence of long-range dependence (Long Range Dependency - LRD).

Self-similarity is a concept that combines fractals, chaos and power laws. Self-similarity or invariance with respect to changes in the scale or size is a distinctive feature of many of the laws of nature and the countless phenomena in the world around us. Self-similar traffic [5] has the following statistical properties that are important from the point of view of modeling:

- distribution of time intervals of packets decrease slowly and have a so-called "Fat tails";
- distribution of time intervals of packets have infinite moments (since some order);
- slow speed $\backsim n^{-1}$ decreasing the dispersion calculated based on the sample traffic, increasing the length of the sample.

It should be noted that in the context of self-similar traffic classical methods of calculating parameters of a computer network (channel capacity, buffer capacity, etc.), based on Poisson flows, often give unduly optimistic decisions and lead to an underestimation of the load.

## 6. Fractal models

The most common models for simulating fractal traffic are:

- chaotic display;
- fractional Brownian motion (FBM);
- fractal Gaussian noise (FGN).

FGN-wide-sense stationary stochastic process $X = \{X(t)\}_{t=1}^{\infty}$ with a mean value $\mu_x$, dispersion $\sigma^2$ and the autocorrelation function

$$\rho_x(k) = \frac{1}{2}(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}), k \geq 1.$$

FGN a process a self-similar in a narrow sense (with self-similarity parameter $H$). FGN with zero mean is used as an incremental process in FBM. FBM - Gaussian process

with zero mean and continuous time $B_H = \{(t)\}^{\infty}{}_{t=0}$ with self-similarity parameter $H$ and autocorrelation function $\rho_x(s,t) = \frac{1}{2}(|s|^{2H} - 2|t|^{2H} + |t-s|^{2H})$.

FGN and FBM have been successfully used in hydrological studies, where their application was transferred to the modeling of traffic [6]. Complexity of their use in this field due to the definition of the best to get closer to the characteristics of source traffic parameters $\mu_x \sigma^2{}_x$ and $H$.

The most common and conceptually simple model that allows to generate self-similar traffic are models based on the so-called chaotic maps (CMAPs). These models use a smaller number of parameters than FGN and FBM, and their choice has a clearer interpretation.

Dimensional mapping, where $x_{n+1} = \begin{cases} f_1(x_n), & 0 < x_n \le d \\ f_2(x_n), & d < x_n < 1' \end{cases}$

where $d$ −display option called chaotic if the functions $f_1$ and $f_2$ satisfy sensitivity to initial conditions(sensitive dependence on initial conditions-SIC), distance between the trajectories through $N$ successive iterations can be written in the form $|f^N(x_0+\delta) - f^N(x_0)| = \delta \cdot e^{N\lambda(x_0)}$,где$\lambda(x_0) > 0$, for most values $x_0$.

When modeling the chaotic traffic display is often used in so-called ON-OFF models as generators of transitions between states of activity and passivity source. One of the most used maps is a sequence of "intermittency" ( Figure 2):

$$x_{n+1} = \begin{cases} \varepsilon_1 + x_n + c{x_n}^m(x_n), & 0 < x_n \le d, \\ \frac{x_n}{1-d}, & d < x_n < 1, \end{cases} \quad \text{где} \quad c = \frac{1-\varepsilon-d}{d^m}$$

Varying the parameter ε can change the length of time in the first state model, using the parameters $d$ and $m$ can control the average value of the traffic intensity.

A common disadvantage of currently used models of network traffic is blocked specificity (focus on any particular kind of traffic) and the lack of versatility. Furthermore, their application in practice leads to a large amount of research work required to adapt the model to the network configuration parameters that affect the statistical characteristics of traffic. All this greatly complicates the construction of a universal model, due to the large diversity of both the sources and network configurations that affect their work.

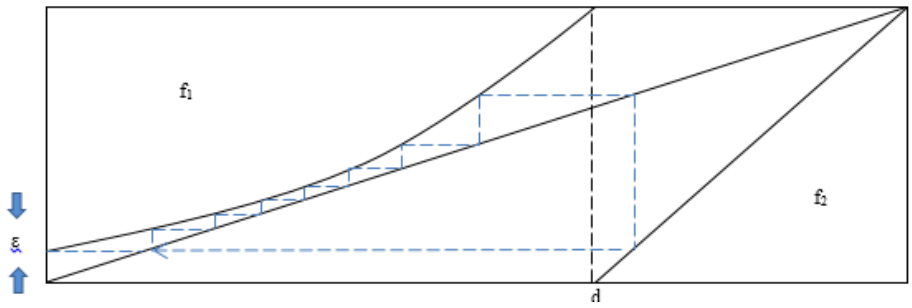



**Fig.3.** Illustration of the mechanism generating a sequence of "intermittency"

Context modeling methods (CM) are already used as methods of predicting the behavior of sources with memory, particularly in areas such as data compression and pattern recognition.

The CM methods can be on the basis for a universal simulation model (USM) of network traffic, able to adapt to network load, with a variety of statistical characteristics. Such USM is equally applicable for modeling the total of network traffic (traffic aggregate flows) and for the traffic simulation individual network applications and services [7].

The essence of context modeling methods is predicting the behavior of the modeled source of information in the future on the basis of statistical information about the behavior of the source in the past, in the context of different orders of magnitude, consistent with current contexts. The prognosis is usually a probability distribution of possible values in the future implementation of the simulated source. Having as a set of forecast probabilities, using a random number generator, the sequence can be generated to simulate the behavior of the simulated source. Indeed, the network traffic consists of a stream of individual discrete elements of information (packets, frames, cells). Mathematically, it can be described in several ways, namely by using:

- point process consisting of a sequence of values of $T_0,\ T_1, \dots, T_n, \dots, (T_0 = 0)$, which is the moment of arrival of information elements;
- non-negative integer-valued process with continuous time (counting process) $N = \{N(t)\}^{\infty}_{t=1,}$ где $N = \max\{n: T_n \le t\}$ –the number of occurrences of packets in the interval [0, t];
- non-negative integer process consisting of a sequence values of total load values calculated in equal non-overlapping intervals;
- non-negative process, consisting of a sequence of intervals between occurrences packets.

Choice of representation depends on the application model and method. Used to build the USM context approach determines the representation of traffic in form sequence dimension of values for total load, forming a sequence of "symbols", suitable for contextual analysis.

The message source are encoded with symbols of an alphabet $U$. In the case of traffic modeling applications, symbols will be measured by the intensity values of the load created by the application. Sequence of symbols $S = \{a_0, a_1, \dots a_{t-1}, a_t\}$, created by the source (the application that created network traffic) for the time interval $[0, t]$, where through $t$ is denoted by the current time.

The current context of order $i$ is a sequence of length $i$, which is a suffix of the sequence, including the current symbol. As part of the contextual approach as a method of constructing a universal context simulation model chosen method of dynamic Markov modeling (DMM), pre-adapted to the new domain, allowing for the simulation of traffic in packet-based DSN as Dynamic Markov Modeling in comparison with the rest of the CM is the most convenient for the subsequent generation of the simulation sequence.

## 7. The method of modeling traffic on the basis Dynamic Markov Model

In the modeling of network traffic is a problem a large alphabet, symbols which did not initially unknown. One of the solutions to the problem is the addition of pre-pass on the input sequence, which resulted in the initial model is constructed, consisting of a group of states corresponding to each registered symbol input flows.

During the subsequent (second) pass, creating simulation DM- model series handles all symbols of the training sequence, create a source of stress. At each step of DM-model is

represented by automaton with a finite set of states corresponding to the training sequence contexts. These probabilities are also the transition probabilities to other states [8]. The calculation of transition probabilities is implemented using frequency counters transitions $C_i$, realizing the count of the number of transitions in the corresponding states: $P_{\text{пер}A\to B} = C_{\text{пер}A\to B} / \sum_{\text{по всем } i} C_i$

Task of simulation model starts at equal values of the counters transitions in all states initial diagrams, moreover $C_i \geq 1, \forall i$.

Upon receipt of a character input stream is a transition model to the next state corresponding to the character who comes and modification of frequency counters, as appropriate, of which there was a transition: $C_{\text{пер}A\to B} = C_{\text{пер}\,A\to B} + 1$.

To determine the need for cloning states in the constructed model is used as a criterion for comparison of the number of transitions from the current state to be cloned to and the number of transitions to be cloned to all other states with predetermined constants $c_1$and $c_2$ being the parameters of the model. If the specified number of transitions exceed these thresholds are constants, then cloned the state in which the transition occurs.

In this case, the frequency counter symbols are divided between two states - cloneable and new, in proportion of the number of hits:

1) from the previous state to be cloned to;

2) in all other states to be cloned to.

Algorithm for dynamic construction of the state diagram is illustrated in figure 4.

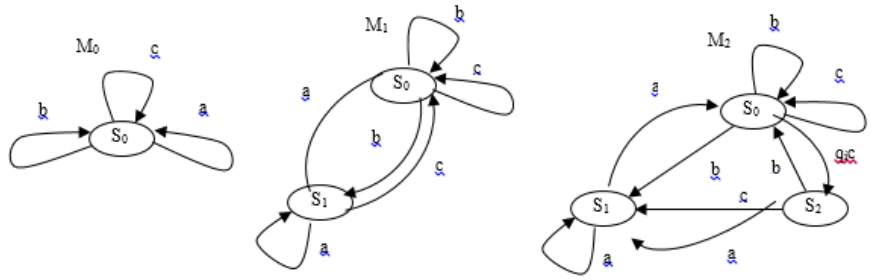



**Fig.4.** Dynamic construction of state diagram

Input symbol alphabet relies three$-a, b, c$. Three possible input symbols correspond to transitions represented by arcs. $M_0$ – the initial state of model; $M_1$, $M_2$ – Modeling the 1st and 2nd step, respectively. Bold characters received on the input of system occurred and, as a consequence, transitions after "learning" model is ready to work. In the simulation at each step of the output pattern will receive a set of probabilities of symbols computed from the values of the frequency counter at each step of above formula.

## 8. Measurement and synthesis models of network processes

On the basis offering method might possible to measure and syntheses of network traffic without loss packet in computer network.

In general, the process of measurement networks consists of two stages: data acquisition and analysis. In first step procedure of direct measurement of traffic characteristics, which include data on circulating in the network packets generated by

different protocols [9]. Data collection is carried out at any point in the network, by means of recording the traffic packets with information about the packet headers and timestamp of their registration. In second step is realized the procedure of processing gathering information by helping tools, that let convert the data in accordance with using [10] models and provide traffic characteristics in the form of numerical values or raw data for create graphing.

Agents used during the measurements: TCP dump (capture utility for IP-packets on a network interface - to dump the traffic), TCP trace (utility analysis of the resulting dump), xplot (plotting utility). The general scheme of measurement is shown in figure 5. As objects of measurement can be physical signals in the transmission channel, individual packets, virtual connections, consisting of series packets and aggregated traffic.

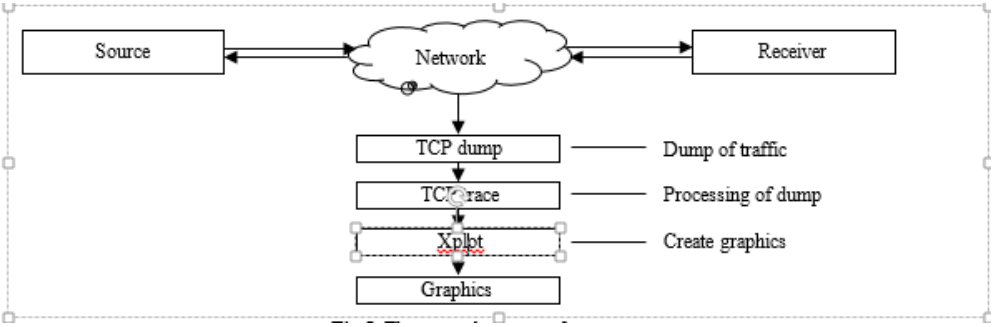



**Fig.5.** The general scheme of measurement

When measured, consider the following important factors:

- each measurement changes the status of the packet. When a packet passes through an intermediate network device is performed to reduce the time (Time To Live - TTL) life by one and recalculate the checksum for each packet;
- measurement packet "outraged network". Going to add to the delay in the transfer of packets, which in turn leads to the appearance of new properties (not stationary, originality, etc.);
- the measured traffic characteristic can be obtained over a time interval which is associated with the minimum packet size and the lifetime of the virtual TCP connection.

Measured characteristics are due to the hierarchical relatedness (figure 6) and the mechanism of formation of virtual connections of packets.

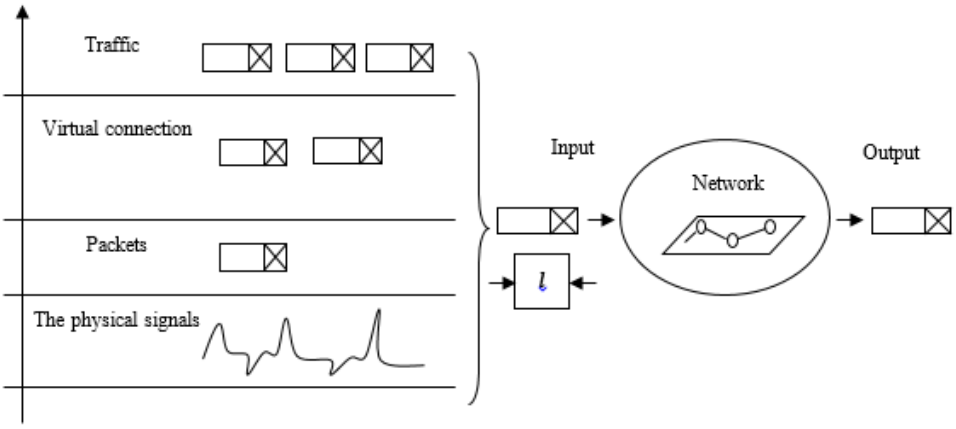



**Fig.6.** Hierarchical coupling processes

Evaluation characteristic scale of measurement:

$\Delta t = \frac{l_p}{\lambda} = \frac{1500 \times 8}{100 \times 10^6} = 120 \times 10^{-6}$, where $l_p$ −packet length: 1500 × 8 bits, capacity channel $100 \times 10^6$ bits per second.

## 9. Conclusion

As a result, network traffic monitoring was found that the sequence values intensity of network load, created by many network applications that have strong contextual dependencies. This leads to possibility of effective application of context methods for modeling their basic characteristics. On the basis offered method a set of probability packets by helping generator of random number generates a value of the next symbol (the value of network traffic on the modeling step) and according to received symbol is a transition model to the next state.

## References:

1. Kyrtsou C (2008). "Re-examining the sources of hetero scedasticity: the paradigm of noisy chaotic models". Physica A 387 (27): 6785–9.2008.09.008.
2. Werndl, Charlotte (2009). "Are Deterministic Descriptions and In deterministic Descriptions Observationally Equivalent?". Studies in History and Philosophy of Modern Physics 40 (3): 232–242.:2009.06.004.
3. T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process," Telecommunication Systems, vol. 17, no. 1-2, May/Jun 2001.
4. W. J. J. Roberts, Y. Ephraim, and E. Dieguez, "On Rydén EM algorithm for estimating MMPPs," IEEE Signal Process. Lett., vol. 13, no. 6, Jun. 2006.
5. Taqqu M, Willinger W, Sherman R. Proof of a fundamental result in self-similar traffic modeling. Computer Communication Review, 2007.
6. Herbin, E.: From N-parameter fractional Brownian motions to N-parameter multi fractional Brownian motions. Rocky Mountain J. Math. (2005).
7. Erramilli, O. Narayan, A. Neidhardt, and I. Saniee, "Performance impacts of multiscaling in wide area TCP/IP traffic," in Proceedings of INFOCOM 2010, 2010.
8. W. Wei, B. Wang, and D. Towsley, "Continuous-time hidden Markov models for network performance evaluation," Perform. Eval, vol. 49, 2002
9. L. Deri and F. Fusco, "Exploiting commodity multi-core systems for network traffic analysis", (2009).
10. Nottingham and B. Irwin, "Investigating the Effect of Genetic Algorithms on Filter Optimization within Fast Packet Classifiers", Proceedings of the ISSA 2009 Conference, (2009).

# Approximation of Planar and Spatial Experimental Curves by Splines that are Defined Parametrically

**Vitaliy Vishnevskey[1], Tatyana Romanenko[2], Vladimir Kalmykov[3]**

[1] Institute of Mathematical Mashines and Systems Kiev, Ukraine
vit.vizual@gmail.com
[2] Institute of Mathematical Mashines and Systems Kiev, Ukraine
romanenko@immsp.kiev.ua
[3] Institute of Mathematical Mashines and Systems Kiev, Ukraine
vl.kalmykov@gmail.com

**Abstract.** The method to approximate the planar and spatial experimental curves by the parametrically specified splines is offered. The curve to be approximated is represented by a sequence of readouts. The result of the approximation is invariant relatively to the number of readouts and determined only by the shape of the approximated curve.

**Keywords.** Approximation, experimental curves, parametrically specified splines.

## 1.Introduction

The experimental curves, which represent the results of data measurement, are generally distorted by noises. It is assumed that the measuring process is completely determined by some unknown function $f(x,y)$ for plane curves, or $f(x,y,z)$ for space curves. The result of measurement is a finite sequence $I$ of pairs $\{x_i,y_i\}$ for the case of plane curves, or triples $\{x_i,y_i,z_i\}$ for the case of space curves; $i = 1,I$. In other words, there is a tabular implementation of the function. Often, the experimental curves are plane or spatial graphics. In addition, it can be closed curves, for example, the object contours in the images. The contour of the object in the binary image is presented as a plane closed curve. The of object in the halftone image is presented as a spatial closed curve, and as the third coordinate values using the appropriate values of brightness image points. The graphics, the object contours of the images – this is seemingly the easiest and the long used method for cognitive representation of experimental measurements in various areas of human activity, which allow to evaluate the qualitative property of the process, the form of the object, despite noises, measurement errors.

The most basic feature of the graph or contour is its shape, which shows, as a rule, the unknown function that generates a visible realization of the curve. Different implementations differ in scale, the number of measurements and other affine transformations. Moreover, they are often corrupted by noise. However, the main features of the form are saved. For automatic or automated processing of graphics or contour analytic representation of the object must be obtained. The proposed method of analytical evaluation of any graph or contour is considered further.

## 2. The Description of the Experimental Curves

The description of the experimental curves, using parametrically given functions, allows at least to describe the curve of arbitrary complexity by the single-valued functions - Figure 1. The experimental curves representing one the same processes or objects may differ significantly from each other scales, the amount of measurement, noise, and so on. At the same time, the form of the graphical curve characterizes the parameters of the imaged object or process. The automatic processing of the graphic curves involves comparing their forms to determine whether different graphics or curves are the same or different processes or objects. Direct use of neural network techniques or methods of statistical pattern recognition to solve this problem is not possible, because different curves relating to the same object may differ from each other by parameters such as scale, noise, number of

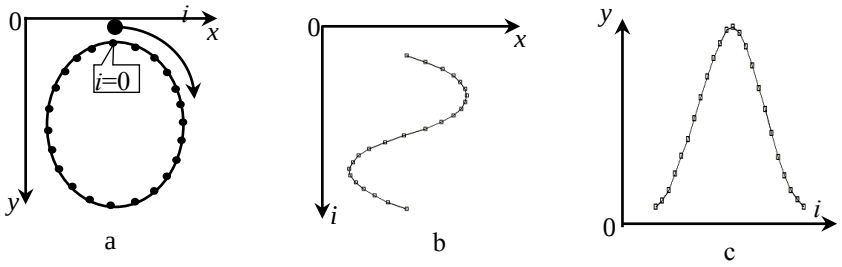



Fig.1. The contour of the simple figure – it is oval and its description by parameter functions $x = \varphi(i)$, $y = \psi(i)$, where $i$ – the point number in the sequence.
a) the initial contour; the measurement points are shown; the arc arrow indicates the start and the measurement direction of the contour points coordinates.
b) function $x=\varphi(i)$; c) function $y=\psi(i)$.

measurements, etc. Since the functions that define the experimental curves are unknown, they must be approximated by some other functions that are invariant to affine transformations in order to enable comparison of these curves. Parametrically defined splines [1,2] can be used to approximate the experimental curves. Typically, splines are used in computer graphics to produce smooth curves passing through the given points [3]. The number of points should not be too great that the required computing resources were in reasonable limits to solve the corresponding system of equations. However, the number of measured points for the experimental curve may be too large to use the traditional methods. In [4,5] the method is considered using splines for the approximation of the experimental curves, which are given in tabular form.

In these articles the new method for approximating the graphs that reflect experimental data is proposed. The proposed method differs from the known methods in that it is based on the assumption that points of the graph represent a function as a single whole, and therefore are connected with each other and therefore a connected path is formed.

There is proposed to use the size of the area between the approximating function and the graph in order to evaluate the their closeness.

To simplify the calculations, we propose to replace the approximating function to broken line, the nodes are the points of this function, and the nodes are chosen to ensure the desired accuracy of the approximation.

The offered algorithm allows finding the parameters of Bezier curve, as a special case of parametrically defined splines, which approximates the experimental graphs.

The curves, in particular the contours of objects can have a rather complex shape. It is difficult to find an explicit analytic function to describe such curves. At the same time, every curve can be represented in parametric form as: $X(t)$, $Y(t)$ - for the case of a planar curve, $X(t)$, $Y(t)$, $Z(t)$ - for the case of spatial curve, where $t$ - a parameter, $0 \le t \le 1$. Functions $X(t)$, $Y(t)$, $Z(t)$ are single-valued and can be represented as a polynomial of degree $r$.

However, in practice the measurement sequences are difficult to approximate polynomials. If the sequence of measurements, which are curves of complex shape, long enough, it makes sense to use the cardinal cubic splines [3] ($r = 3$). Cardinal cubic splines are a sequence of $N$ cubic polynomials which are defined by control points $u = \{x_n, y_n\}$ or $u = \{x_n, y_n, z_n\}$; $n = 1, N + 1$; $N << I$. Each value of $n$ in the sequence of control points corresponds to a value $i_n$. We consider only valid sequence of control points. Under a valid sequence mean a sequence that $i_n < i_{n+1}$. The number of polynomials in the sequence is determined by the complexity of the approximated curve shape. Subsequence $(i_{n+1} - i_n)$ of the measured values corresponds to each of the $n$ polynomials. A sequence of polynomials $X(t) = \{X_n(t_n)\}$, $Y(t) = \{Y_n(t_n)\}$, $Z(t) = \{Z_n(t_n)\}$ corresponds to each of the spatial coordinates. Coefficients of each of cubic polynomials in the sequence for coordinates $x$: $X_n(t) = a_{nx}t^3 + b_{nx}t^2 + c_{nx}t + d_{nx}$ is defined control points $x_{n-1}, x_n, x_{n+1}, x_{n+2}$:

$$a_{nx} = T\left(x_{n+1} - x_{n-1}\right) + T\left(x_{n+2} - x_n\right) + 2x_n - 2x_{n+1}$$

$$b_{nx} = -2T\left(x_{n+1} - x_{n-1}\right) - T\left(x_{n+2} - x_n\right) - 3x_n + 3x_{n+1}$$

$$c_{nx} = T\left(x_{n+1} - x_{n-1}\right)$$

$$d_{nx} = x_n,$$

where $T$ - tension. (at $T = 0$, we obtain a straight line segment, the greater the $T$, the greater the bending of the curve). For most practical applications, $T = 0.5$ is used.

For open curves for $n = 1$: $x_{n-1} = x_1$; if $n=N$: $x_{n+2} = x_{N+1}$. For closed curves $x_{N+1} = x_1$; $y_{N+1} = y_1$; $z_{N+1} = z_1$; for $n = 1$: $x_{n-1} = x_N$; if $n=N$, $x_{n+2} = x_1$.

The formula for determining the polynomial coefficients in the sequences for the coordinates $y$ and $z$: $Y_n(t) = a_{ny}t^3 + b_{ny}t^2 + c_{ny}t + d_{ny}$, $Z_n(t) = a_{nz}t^3 + b_{nz}t^2 + c_{nz}t + d_{nz}$ similar to the above formulas for $X_n(t)$.

Each sequence of control points defines a sequence of polynomial curves. Each of the polynomials in the sequence is defined in the interval between control points $n$, $n+1$. For example, the polynomial $X_n(t_n) = a_{nx}t_n^3 + b_{nx}t_n^2 + c_{nx}t_n + d_{nx}$ is defined by control points: $x_{n+1}, x_{n+2}$, on the interval $\left| x_n, x_{n+1} \right|$, and

$$t_{ni} = \frac{i - i_n}{i_{n+1} - i_n}$$

where $0 \le t_{ni} \le 1$, $i_n \le i \le i_{n+1}$.

The values of the approximating curve $X_n(t_{ni})$, $Y_n(t_{ni})$, $Z_n(t_{ni})$ correspond to each value of the $i$-th coordinate of the experimental curve $x_i, y_i, z_i$.

## 3.The optimal sequence of the approximating control points

For experimental curve which should be approximated most suitable sequence of control points must be defined, the sequence of polynomial curves for which displays the shape of the approximated curve the best.

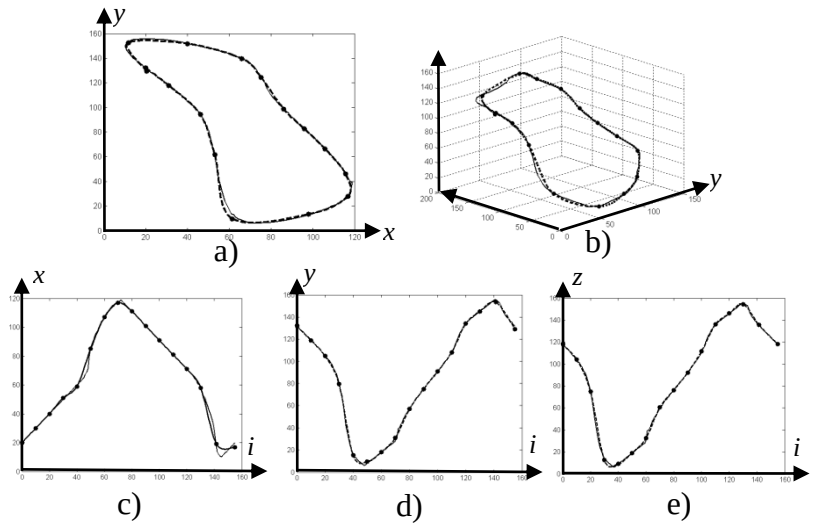



Fig.2. The examples of experimentally obtained closed contour lines and their approximation using the parametrically given splines.
The experimental curves are shown by solid lines; approximating curves are presented by broken lines; control points are represented as balls.
a) The measured points of the plane curve connected by straight line segments and approximating them parametrically given curve with applied control points. The measured parametric components of the curve and their approximate curves are shown on the graphs c) and d).
b) The measured points of the spatial curve connected by straight line segments and approximating parametrically given curve with control points. The measured parametric components of the curve and their approximate curves are shown on the graphs c), d) and e).
c),d),e) The graphs of the measured components for the plane and space curves, respectively, for *x*, *y*, *z* coordinates, depending on the number of measurement points *i* as a parameter and their approximating curves.

Let $(u)$ is the set of all possible sequences of control points. For some sequence of control points $u$ the distance $d(u)$ between the measured values of the experimental curve and the calculated values approximating curve can be represented as

$$d(u) = \sum_{i=1}^{i=I} (x_i - X_n(t_{ni}))^2 + (x_i - Y_n(t_{ni}))^2 + (x_i - Z_n(t_{ni}))^2 .$$

Then the optimal sequence of control points for cardinal spline which approximates the experimental curve can be defined as

$$u_{opt} = \arg\min_{\forall u} d(u).$$

The gradient descent method is used to find $u_{opt}$.
The approximation method of the experimental curves using the parametrically given cardinal splines is developed and experimentally verified. The examples of the experimental plane and space curves to be approximated and the resulting approximating curves are shown in Figure 2. The main advantages of the method are the next. The result of the approximation is determined by the shape of the experimental curve and does not depend on the number of measured points.
Approximating functions are unambiguous and orthogonal, which greatly facilitates the decomposition curves into classes with automatic recognition.

## 4.Conclusion

The method and the algorithm of the plane and space approximate representation curves is proposed. The method uses a previously proposed by the authors method of approximating the experimental plots by the parametrically given splines.
As before, the proposed method uses the assumption that all the experimental curve is an implementation of some unknown process. Thus the measured points must be connected in series to form a simply connected polyline. Then the area between the broken line and the approximating curves can be used as a measure of the differences between them. The best approximating curve is such that the area between it and the experimental curve is minimal.
The result of the approximation depends weakly on the number of measured points and noise.
The resulting analytical representation of the experimental curve is invariant with respect to some affine transformations and can be used for automatic processing of experimental data.

## References

1. J.H. Ahlberg, E.N. Nilson and J.L. Wash, The Theory of Splines and Their Applications, Academic Press, New York, 1967
2. D.F. Rogers, J.A. Adams, Mathematical Elements for Computer Graphics, McGraw- Hill, New York, 1976
3. Carl de Boor, A Practical Guide to Splines, Springer New York, Inc., 2001
4. Vitaliy Vishnevskey, Vladimir Kalmykov, Tatyana Romanenko. Approximation of experimental data by Bezier curves// International Journal "Information theories & applications". – Sofia. – 2008. – v. 15, № 3 – P. 235.
5. Tatyana Romanenko, Vitaliy Vishnevskey, Vladimir Kalmykov. Analytical Representation of Graphs by Means of Parametrically Defined Splines // Proceedings of the international conference on applications of information and communication technology and statistics in economy and education ICAICTSEE – 2013 December 6-7th, 2013, UNVE, Sofia, Bulgaria – P.536-542.

# Knowledge Models for Organizational Maintenance of Situation Centers

**Oleksii E. Kovalenko**[1]

[1] IMMSP NASU, Academician Glushkov Av., 42, 03680 Kyiv, Ukraine
koval@immsp.kiev.ua

**Abstract.** The paper reviews organizational models of situation centers and their relationship with knowledge bases used for adequate organizational maintenance. The model of integrated knowledge base for organizational maintenance is described. The knowledge base is viewed from situation center structure and business process standpoints. Overall knowledge base model will be a composition of these orthogonal models. Promising directions for future research are outlined.

**Keywords.** Knowledge model, ontology, organizational maintenance, situation center.

## 1.Introduction

The system approach to describing, analyzing and building of organizational structures that carry out purposeful activity, based on the use of standardized principles and tools for multidimensional representation of such structures. These principles and tools described in the relevant standards, generally accepted guidelines and frameworks and allow a unified position to consider various aspects of organizational structures and carry out their comparisons.

Functioning of organizational system (company, enterprise) is associated with ensuring of coordinated collaboration of staff to implement the mission and the objectives of the organization based on solving problems that are implemented through business processes and is supported by organizational mechanisms (fig.1).

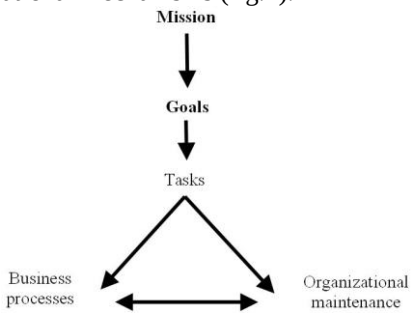



Fig.1. Organization functioning model

Collaborative activities of staff can be viewed from the point of view agents-based approach and use the behavioral model of agents for this. [1]. The behavior of staff agents is based on its knowledge. Hence, the development of adequate knowledge model for staff agents is actual problem.

## 2.The analysis of the current state of research and publications

A separate type of organizational systems is situational centers (SC) that perform the functions of situational management (SM). Situation management systems based on information technologies are used in various branches of human activity [2; 3]. Situation management is considered "as a framework of concepts, models and enabling technologies for recognizing, reasoning about, affecting on, and predicting situations that are happening or might happen in dynamic systems during pre-defined operational time" [3].

Various systems, including organizational, describes their architecture. There are various definitions of system architecture [4; 5; 6] and one of generalized definitions presented in the standard ISO/IEC/IEEE 42010:2011 "Systems and software engineering - Architecture description" [7] (the international revision of IEEE 1471:2000): **architecture** – fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution.

ISO/IEC 42010 (except the term "architecture") defines a number of terms:

- architecting: process of conceiving, defining, expressing, documenting, communicating, certifying proper implementation of, maintaining and improving an architecture throughout a system's life cycle (i.e., "designing");
- architecture description (abbreviation 'AD'): work product used to express an architecture;
- architecture description language (abbreviation 'ADL'): any form of expression for use in architecture descriptions;
- architecture framework: conventions, principles and practices for the description of architectures established within a specific domain of application and/or community of stakeholders;
- architecture viewpoint: work product establishing the conventions for the construction, interpretation and use of architecture views to frame specific system concerns;
- architecture view: work product expressing the architecture of a system from the perspective of specific system concerns;
- concern: interest in a system relevant to one or more of its stakeholders. A concern pertains to any influence on a system in its environment, including developmental, technological, business, operational, organizational, political, economic, legal, regulatory, ecological and social influences;
- model kind: conventions for a type of modeling. An architecture view consists of multiple models, each following one model kind;
- stakeholder : individual, team, organization, or classes thereof, having an interest in a system.

The ISO/IEC/IEEE 42010 standard makes a strict distinction between Architectures and Architecture Descriptions. Context for Architecture Description of ISO/IEC/IEEE 42010 depicted on fig.2 [7].

Other known definftions of architecture are the following.

**ZIFA architecture definition** [4]**.** Architecture is a structured set of descriptive representations relevant for describing an object and being employed such that an instance of the object can be created and such that the descriptive representations serve as the baseline for changing an object instance.

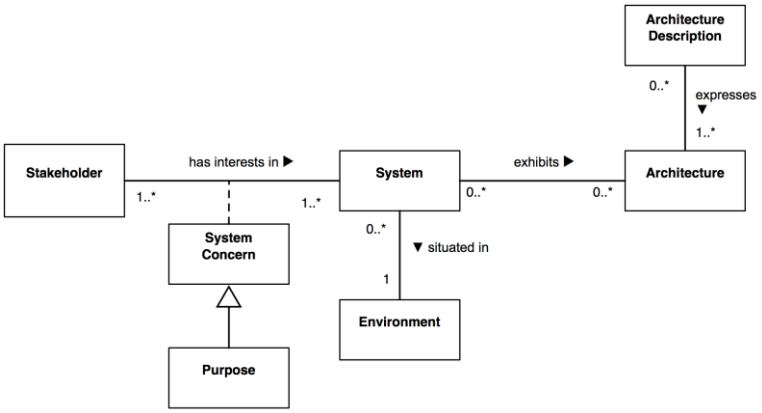



Fig.2. Context for Architecture Description of ISO/IEC/IEEE 42010 [7]

**Gartner enterprise architecture definition** [6]**.** Enterprise architecture is the process of translating business vision and strategy into effective enterprise change by creating, communicating and improving the key requirements, principles and models that describe the enterprise's future state and enable its evolution. The scope of the enterprise architecture includes the people, processes, information and technology of the enterprise, and their relationships to one another and to the external environment. Enterprise architects compose holistic solutions that address the business challenges of the enterprise and support the governance needed to implement them.

There are many specifications of architecture frameworks for modeling of enterprise architecture oriented on IT technologies [8]. The most popular of them are:

- Zachman Framework for Enterprise Architecture (ZIFA) [9];
- The Open Group Architecture Framework (TOGAF) [10];
- Federal Enterprise Architecture Framework (FEAF) [11];
- Gartner Enterprise Architecture [6];
- The DoDAF Architecture Framework [12].

Various aspects of enterprise architecture representation regulated by other standards in particular ISO 14258 [13] and ISO 15704 [14]. Models of enterprise architecture defined various possibilities of its description and use.

Separate place in the enterprises architecture occupy corporate knowledge and knowledge management technology. The corporate knowledge is part of enterprise architecture and defined its mission, goals, tasks, business processes, organizational structure, are considered in the context of interaction with the environment. Such a system will be called *knowledge architecture*.

## 3.Statement of the problem

Organization of corporate knowledge management (KM) is one of basic factors of effective enterprise activities. Therefore knowledge becomes the defining resource and knowledge management is an important component of the company. The model of corporate knowledge enterprise is determined by organizational model and its context (the interaction with the environment).

The specific of KM in SC is the presence of two knowledge areas. The first area – related directly to the activities of SC and the other – with problem areas of situational management, which is carried out using SC.

Knowledge management as a scientific discipline was initiated and developed in the works of Ikujiro Nonaka [15; 16; 17]. Transformation model of corporate knowledge proposed by Ikujiro Nonaka [16] depicted in fig.3.

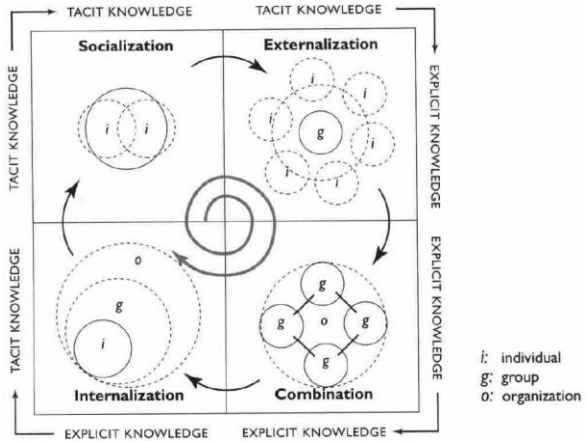



Fig.3 Spiral Evolution of Knowledge Conversion and Self-trancending Process [16]

Knowledge management is based on the enterprise architecture, its mission and so on. Elements of organizational structure ensure the implementation of the mission through the composition of motivations, knowledge (abilities and skills) and resources (or restrictions).

## 4.The main problem

Knowledge Model of organizational structure reflects the specificity each of its components and the nature of the relationship between them. Knowledge model of situational center is representing by a separate domain in knowledge system of organizational structures. Thus is the need for take account of the specific features and build knowledge models for organizational support of situational center.

## 5.Knowledge Models for Organizational Maintenance

Components of organization structure are depicted on fig. 4. Henry Mintzberg identified five structural models of organization and factors that determine them [18]. Structural models of organization are Simple Structure, Machine Bureaucracy, Professional Bureaucracy, Divisionalized Form, Adhocracy. The main features of organizational model there are the mechanism of coordination, a key part of the organization, design parameters (type of decentralization), and situational factors (motivations).

Model of SC organizational support is the basis for the implementation activities processes for situational management in a changing environment with adaptation to specific problems subject area of decision making support. Based on the characteristics of SC functioning Adhocracy is the most adequate model of the structural configuration of SC.

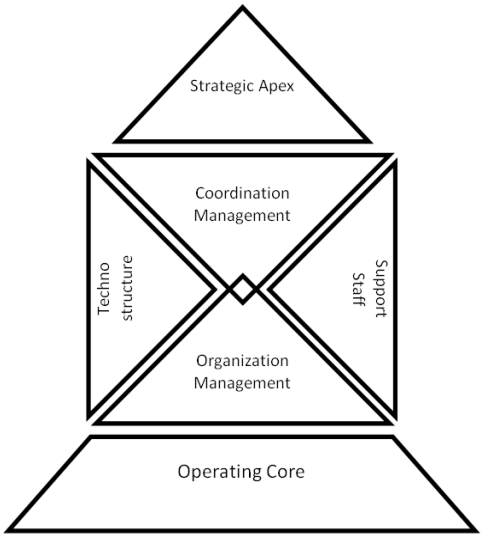



Fig. 4. Components of Organization Structure [19]

Model of organizational support is based on particular organizational structure. Formally, the model of organizational structure described by tuple:

$$M_O = <K, C, D, S>,$$

where $K$ – main coordination mechanism; $C$ – type of core (main part); $D$ – basic design parameters; $S$ – situational factors (motivation).

Specific parameters of organizational support model are defined by SC category. Categorize SC may be based on the classification of SC with using of the various classification criteria. In particular, by such SC classification features as: the composition of tools of situational modeling; the scale of tasks; the method of allocation; determinacy of the level of tasks; the goal direction; the method of processing of situational information; the universality; the subordination of management; the level of automation evaluation of the situation; by the number of premises; the number and structure of staff; necessary for decision making time; the composition of the tools; used technologies; the level of protection; the problem domain. Some of these classification features are basic (decisive), and the rest are additional or clarifying. Based on the model of the components of the organization (fig. 1) SC category is determined based on its mission, goals and objectives of activity.

We can offer the following hierarchy of classification criteria. The first (general) level includes criteria of goal direction, problem domain, scale and determinacy of tasks. Criteria of second level determine administrative aspects of the SC activity, in particular, the subordination, staffing, method of situational information processing, limitations of the time of decision making. Criteria of third level determine the specific design of SC and include a method of placement, universality, the number of premises. Criteria of fourth level determine the technical and technological aspects of the SC activity and include a list of technical equipment, structure of situational simulation tools, used technology, the level of security, and level of automation of the situation evaluation. The hierarchical category model is represented as a tuple:

$$M_I = <G, U, B, T>,$$

where $\boldsymbol{G}$ – general features categories; $\boldsymbol{U}$ – parameters of management aspects; $\boldsymbol{B}$ – the specific parameters of design; $\boldsymbol{T}$ – technical and technological parameters.

Thus, a complete knowledge model of situational center can be represented as a superposition model SC categories based on a hierarchical system of criteria and adequate models of organizational structure for this category:

$$\boldsymbol{M = M_I \circ M_O}$$

Components of defined knowledge model of situational center have poorly structured and informal nature. An adequate representation for them is the use of knowledge-oriented technologies.

The proposed knowledge model of organizational support of situational centers can be implemented as an ontology in which the set of concepts represented the union of the sets of organizational structures and features of SC category, the relationship between concepts and categories determine the properties of organizational structures and functions define the rules of interpretation superposition of the models.

## 6.Related Works

The purposeful activity described by trajectory, which includes events, situations, motivation, objectives, resources, constraints. The set of possible trajectories of purposeful system is described by a graph whose vertices represent states of the system, and arcs – events that cause the transition of the system from one state to another. Since the graph describes sequential in time changing of states, it is acyclic (and temporal). Combination the concepts of state-situation-conditions determine the situation as a context (semantics) of state defined by certain conditions.

Situation semantics is an information-based theory [20; 21; 22; 23], that seeks to understand linguistic utterances in tems of the information conveyed. Although work carried out by Devlin and Rosenberg [24] in the 1990s showed that situation theory could also be used to analyze language use from an action perspective. Barwise and Perry [25] began with the assumption that people use language in limited parts of the world to talk about (i.e., exchange information about) other limited parts of the world. Call those limited parts of the world situations.

In their 1980 paper The Situation Underground [25], the first published work on situation semantics, Barwise and Perry wrote of situations:

"The world consists not just of objects, or of objects, properties and relations, but of objects having properties and standing in relations to one another. And there are parts of the world, clearly recognized (although not precisely individuated) in common sense and human language. These parts of the world are called situations. Events and episodes are situations in time, scenes are visually perceived situations, changes are sequences of situations, and facts are situations enriched (or polluted) by language."

The appearance of the word "parts" in the above quotation is significant. Situations are parts of the world and the information an agent has about a given situation at any moment will be just a part of all the information that is theoretically available. The emphasis on partiality contrasts situation semantics from what was regarded by many as its principal competitor as a semantic theory, possible worlds semantics.

## 7.Conclusion

Creation of knowledge models for representing of organizational models of situation centers as a specific type of enterprise enables accurately determine the needs of various aspects of organizational maintenance. Ontology and knowledge bases used are adequate tools for describing of organizational maintenance. The model of integrated knowledge base for organizational maintenance is proposed. The knowledge base is viewed from situation center structure and business process standpoints. Overall knowledge base model will be a composition of these orthogonal models. Future research will be related with the integration of situation semantics and organizational models for building dynamic models of functioning and management of situation centers.

## References

1. Kovalenko, O. E. (2014). Models and Means for Service Agents Orchestration in Situation Management Systems. *Actual Problems of Economics*, *154*(4), 2014, (462-467).
2. Jakobson, Gabriel; Lewis, Lundy; Matheus, Christopher J.; Kokar, Mieczyslaw M.; Buford, John. (2005). Overview of situation management at SIMA 2005. In: Military Communications Conference, 2005. MILCOM 2005. IEEE, (1630-1636). IEEE Press.
3. Jakobson, Gabriel; Buford, John; Lewis, Lundy. (2007). Situation Management: Basic Concepts and Approaches. In: Popovich, Vasily V.; Schrenk, Manfred; Korolenko, Kyrill V. (eds.) Information Fusion and Geographic Information Systems 2007. LNG&C, vol. XIV, (18-33). Springer, Heidelberg.
4. ZIFA – https://www.zachman.com/ea-articles-reference/171-zachman-glossary.
5. FEA – A Common Approach to Federal Enterprise Architecture, May 2, 2012 – http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/common_approach_to_federal_ea.pdf.
6. Gartner Enterprise Architecture: A Home for E-Government – Gartner, 2003 – https://www.gartner.com/doc/405561/gartner-enterprise-architecture-home-egovernment.
7. ISO/IEC/IEEE 42010:2011 Systems and software engineering – Architecture description – http://www.iso.org/iso/catalogue_detail.htm?csnumber=50508.
8. Survey of Architecture Frameworks – http://www.iso-architecture.org/ieee-1471/afs/frameworks-table.html.
9. John Zachman's Concise Definition of The Zachman Framework™ (ZIFA) – https://www.zachman.com/about-the-zachman-framework.
10. The Open Group Architecture Framework (TOGAF) – http://www.opengroup.org/togaf.
11. A Common Approach to Federal Enterprise Architecture (FEA), May 2, 2012 – http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/common_approach_to_federal_ea.pdf.
12. The DoDAF Architecture Framework – http://dodcio.defense.gov/Portals/0/Documents/DODAF/DoDAF_v2-02_web.pdf.
13. ISO 14258:1998 – Concepts and Rules for Enterprise Models.
14. ISO 15704:2000 – Requirements for Enterprise Reference Architecture and Methodologies.
15. Nonaka I.; Takeuchi H. (1995) The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press, 1995 - 284 p.
16. Nonaka I.; Konno N. (1998) The Concept of "Ba": Building a Foundation for Knowledge Creation // California Management Review, 1998, vol. 40, No 3, (40-54).
17. Nonaka I., Toyama R, & Byosière P. (2001). A theory of organizational knowledge creation:Understanding the dynamic process of creating knowledge. In M. Dierkes, A. Berthoin Antal, J.Child, & I. Nonaka (Eds.), Handbook of organizational learning and knowledge, (491-516). New York: Oxford University Press.
18. Mintzberg, Henry. Structure in Fives: Designing Effective Organizations. – Englewood Cliffs, N.J.: Prentice-Hall, 1992. – 312 p.

19. Kovalenko, O. E. (2012). (in Ukrainian) Coordination of Interaction Between Program Agents Within Situational Management Systems. *Actual Problems of Economics*, *138*(12), 2012, (183-189).
20. Devlin, Keith. Situation Theory and Situation Semantics – http://web.stanford.edu/~kdevlin/Papers/HHL_SituationTheory.pdf.
21. Seligman, Jerry; Moss, Lawrence S. Situation Theory – http://fenrong.net/teaching/situationtheory.pdf.
22. Mechkour, Said Overview of Situation Theory and its application in modelling context – https://diuf.unifr.ch/main/pai/sites/diuf.unifr.ch.main.pai/files/education_situationtheory_smechkour.pdf.
23. Szczerbak, M. K., Bouabdallah, A., Toutain, F., & Bonnin, J. M. (2013). A Model to Compare and Manipulate Situations Represented as Semantically Labeled Graphs. In Conceptual Structures for STEM Research and Education (pp. 44-57). Springer Berlin Heidelberg, 2013.
24. Devlin, K. and Rosenberg, D. (1996) Language at Work: Analyzing Communication Breakdown in the Workplace to Inform Systems Design, Stanford University: CSLI Publications and Cambridge University Press,1996.
25. Barwise, J. and Perry, J. (1980) The Situation Underground, in Stanford Working Papers in Semantics, Vol. 1, eds. J. Barwise and I. Sag, Stanford Cognitive Science Group 1980, Section D, (pp.1–55).

# Grid-based System for Storage and Processing the Big Data of Digital ECG

**V. Vishnevsky, L.Kizub**
Institute of Mathematical Machines and Systems of NASU, Acad. Glushkova Avenue, Bl.42,03680 Kyiv, Ukraine
vit@immsp.kiev.ua

**Abstract.** The paper provides a general idea of the Grid-based system "Medgrid" for storing and processing big data consisting of depersonalized diagnostic information. The system is analyzed by the example of electrocardiograms. The paper emphasizes a special web service of "Medgrid" system. This service provides the joint between the digital storage of diagnostic information and hard copies of documents in a case history.

**Keywords.** Medical Grid system, Grid storages, digital electrocardiograms, web services, unique Web link.

Current state of the healthcare organizations in Ukraine is characterized by an apparent contradiction between the relatively widespread digital diagnostic equipment and the paperwork related to keeping hard copies of medical records in order, the latter being formally and essentially archaic.

This contradiction raises at least two questions:

1. Is it truly necessary to store the diagnostic information in digital form (including digital ECG) throughout the lifetime of the patient?

2. If nevertheless the diagnostic information is collected and stored, is it possible to provide access to it considering the aforementioned hard copies system?

The first question, while being actively discussed among the specialists, seems rhetorical for us. On the one hand, the benefits of keeping diagnostic information in digital form are obvious, being proven a long time ago. Losing these advantages by keeping the documentation purely paper-based is a luxury one cannot afford. Besides, storage of the entire diagnostic data during the patient's lifetime is rather a necessity long overdue than a subject for discussion.

We believe the answer to the second question has been given during the implementation of the project of the national program for development and implementation of Grid-based technologies called "Medical grid system for population-based investigations in cardiology based on cardiograms" or simply "Medgrid", which is discussed below.

Electrocardiograms are the primary source of information for the "Medgrid" project. They are produced by the certified digital cardiographs in the appropriate specialized medical facilities.

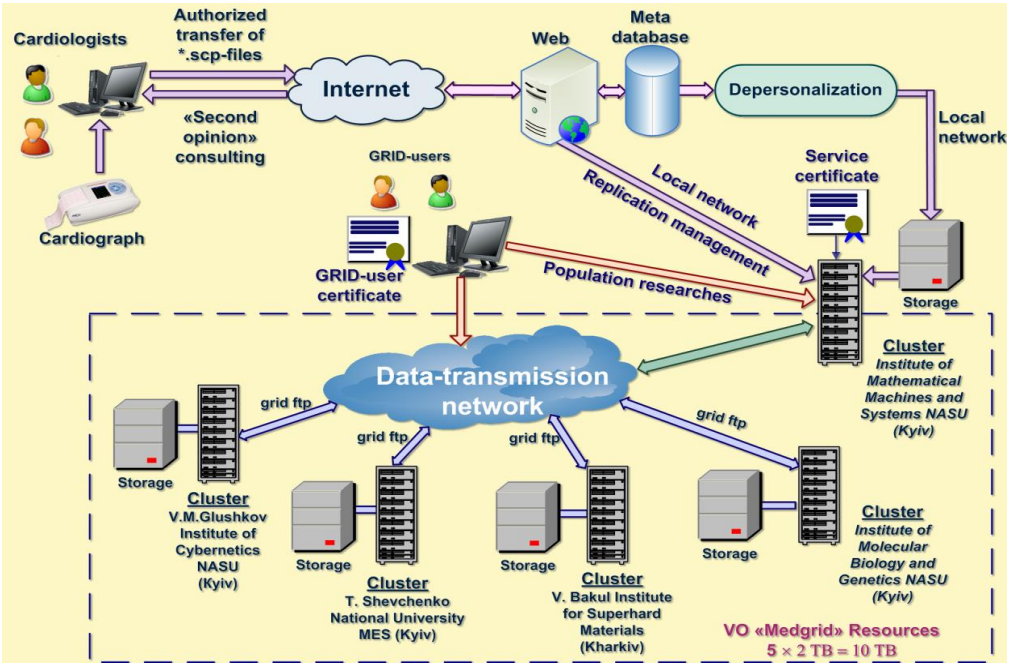



Figure 1. The technological design of the "Medgrid" project

The technological design of the "Medgrid" project combined with the computer resources of the similarly named virtual grid organization (VO) are shown in Figure 1.

Grid storages of the "Medgrid" project allow the accumulation of the depersonalized digital diagnostic data of the population on a national scale and provide the means to process this information for both applied healthcare and scientific research. The goals and potential of the "Medgrid" project are analyzed in detail in [1].

First of all it is essential to understand that digital ECGs are delivered to the distributed grid storages via http://medgrid.immsp.kiev.ua web portal by using our dedicated transport module. The module can operate either as a standalone program or an in-built element of the software system of the cardiologist workplace.

It is worth mentioning that working with depersonalized cardiograms is related to the certain difficulties regarding identification of the patients when generating the arranged meta-database of the system. Since these difficulties are fundamental, we have started examining the efficiency of the patient's biometric identification algorithms based on the cardiogram itself. The examples of such algorithms we are currently investigating are described in [2-4]. Due to the complexity of these algorithms, their discussion is beyond the scope of this report.

We use specially designed interactive web services to work with the digital ECG files. One of these web services generates a unique web link for each cardiogram placed to the grid storage of the project and forms a special PDF document for the physician to print or save for the patient. As shown in Figure 2, this document has a special QR-code storing the access link for the digital ECG in a machine-readable form. A QR-code is easily recognized by the standard software provided with webcams, including those of modern tablets and smart phones.


Український Національний Грід
Проект «МедГрід»
http://medgrid.immsp.kiev.ua

Пацієнт:

ПІБ: И_____

Рік народження: 1959

Стать: Мужской

Поштовий код: 03187

Медичний заклад: «Інститут проблем математичних машин та систем НАН України », 03187, 03187, м.Київ-187, пр. Глушкова, 42

Основний синдром: _____

Лікар _____ В.Вишневский

Файл електрокардіограми: 80_117_20091214_123100.scp

Дата зняття ЕКГ: 14.12.2009 Дата запису в МедГрід: 06.02.2012

URL електрокардіограми:

http://medgrid.immsp.kiev.ua/qr/ecg.jsp?qrcode=5cdc61c65b5cf0dd3d8638745652130641e45f1a

d2ee16cae8cbc9d0651e341080_117_20091214_123100.scp

Figure 2. A unique link to the cardiogram of a patient as a QR-code

We consider hard copy of a document with a built-in QR-code to be a valid solution merging modern digital diagnostic information storage technologies with the archaic method medical paper document filing principles. A unique link used for generation of the QR-code can likewise be employed for telemedicine consultations.

## References:

1. Vishnevsky V.V. Medical GRID-system based on the electrocardiogram: a new tool for clinical cardiology and population studies / Vishnevsky V.V., Tchaikovsky I.A., Kirzhner G.D., Ena L.M., Dordienko N.A., Frolov A., Vasilyev V.E. // International scientific-practical journal "Cardiologists: from science to practice." – 2012. – №2. – P. 108-116.
2. Israel S.A. ECG to identify individuals / Israel S.A., Irvine J.M., Cheng A., Wiederhold M.D., Wiederhold B.K. // Pattern Recognition. – 2005. – № 1. – P. 133-142.
3. Sasikala P. Identification of Individuals using Electrocardiogram / Sasikala P., Wahidabanu R.S.D. // International Journal of Computer Science and Network Security. – 2010. – Vol.10. – №12.
4. Agrafioti F. ECG biometric analysis in cardiac irregularity conditions / Agrafioti F., Hatzinakos D. // Signal, Image and Video Processing. – 2009. – V. 3, № 4. – P. 329-343.

# Situation Center as a Tool for the Situational Management

**Oleg Malyshev**

Institute of Mathematical Machines and Systems Problems of NAS of Ukraine,
42, Academician Glushkov Avenue, 03680, Kyiv, Ukraine
oleg_malyshev@ukr.net

**Abstract.** In recent years, in the developed countries in the governance structure are embedded the components of a new type - so-called "Situation Centers". As a rule, their performance is explained only by the fact of using of advanced information and communication technologies. In this paper we attempt to construct a platform for a broader and more in-depth study of the phenomenon of situation center as a tool for situational management by identifying the essential aspects of the activities of situation center and the factors influencing them. These include the types of tasks of situational management, the contribution of the situational center in their decision, the role of knowledge and information in the situation center activities, a place of situation center in the management infrastructure, the status of situation center, etc. Basing on this platform, it would be possible with a high degree of reliability to determine the profile of any created or functioning situation center.

**Keywords.** Situational management, situation center, knowledge, information, expert, decision making, service.

## 1. Introduction

In recent years in the developed countries to solve management problems the so-called situation centers are created and used. However, on the background of almost universal recognition that the situation center is "something very good", the scientific sources do not contain enough information to clearly understand the specifics of situation centers and the effect of their use.

In this regard, it is a pressing task to construct on the systematic base the classification space in which it would be possible to position the existing situation centers, given that the relevant and reliable information about them is available, and the newly created.

This work aims to contribute to the solution of this task.

## 2. Situational Management

The name "situation center" says that it is designed to meet the challenges of situational management However, the question immediately arises: a situational management - what is it?

Apparently, the first step to creating the foundations of situational management was made in [1]. Recognizing the fundamental nature of this work, however, we have to note a number of controversial issues, such as:

the basic concept of the situation is not determined (interpreted). Instead, the concepts of current and full situations are introduced. Their interpretations, in turn, use the concepts of information and knowledge, given the interpretation of which can not be found to be satisfactory;

in fact, the only one "universal" model of situational management is offered, which, upon closer inspection, turns out to be insufficient for many practical cases.

Unfortunately, since the publication of this work these and other contentious issues were not the subject of attention of researchers.

### 2.1 Situations and the Types of Situations

As was shown in [2], we use the word "situation" when in a certain process the state, falling under identifying effect of specific criterion or criteria system, occur or may occur and some time to be saved. It is possible to distinguish the following types of situations:

desirable situation;
undesirable situation;
actual situation, that is, the situation found and observed;
potential situation, that is, the situation which, presumably, may occur;
target situation, the occurrence of which is the purpose of the activity;
retrospective situation, that is, the situation which previously taken place;
conflict situation.

### 2.2 Situational Management Tasks

The fundamental task of situational management is to create the criteria bases to identify the situations in the various processes. Also on the base of the situations classification we can identify a number of common tasks of situational management and suggest the models to solve them:

**Detection and Development of the Actual Situation.** For the detection of the actual situation in the process the monitoring of this process should be organized to obtain values of certain of its characteristics and to evaluate each of them with the appropriate criteria. In case, if the situation is detected, it is necessary to ascertain whether it occurs. Responding to the current situation depends on what is its "sign" (desirable/undesirable). If this is desirable situation, we should consider how to use it. If this is undesirable situation, we should develop and implement measures to eliminate it (in terms of [3] – to carry out the "correction"). Furthermore, it may be appropriate to conduct a study to determine the causes of undesirable situation, and, if found, remove them (in terms of [3] – to implement the "corrective action"). Finally, it can be assumed that these same reasons led to negative consequences in other processes, and to investigate the matter.

A special case of the development of the actual situation is working out its maintenance, where the object of management is the cause of the situation.

Detection and development of the actual situation should be documented for the accumulation of information about retrospective situations.

**Forecasting and Development of the Potential Situations.** In relation to a particular process it may be profitable to predict the appearance of various situations (both desirable and undesirable) - when and where they may arise and with what probability.

To solve the problem of forecasting we must evidently not only to observe the development of the process, but the context in which it occurs, and to investigate in this context the causal relationships.

If the potential situation is desirable, then, determining the causes of it, it may be possible to enhance their effect.

If the potential situation is undesirable, then, determining the causes of it, it may be possible to develop and implement measures to eliminate them (in terms of [3] – to carry out the "preventive action"). At the same time, of course, we should study the effect of preventive action on other processes.

If the probability of undesirable situation is great, it is advisable, along with preventive action, to develop measures to correct the situation itself, as well as its consequences.

Forecasting and development of potential situation must be documented until the fixing the fact if it is occurred or not.

**Formation of the Target Situation.** The nature of this problem, apparently, is that we should not focus on the direct transition of process in the desired state, but on the creation of preconditions (reasons, causes) for this transition.

**Conflict Management.** Subject - party of the conflict situation in pursuit of its own goals should be aware of the other parties of the conflict, study and predict their activities, and with all this elaborate and implement its decisions.

As you can see, the response to each type of situation implies a creative work, in particular the study of causality in the observed processes and the development of appropriate solutions.

### 2.3 Cycle of Situation Management

Whatever the problem of situational management is solving, the management cycle necessarily includes the following interrelated steps:

decision making;

implementation of the decision, that is the direct material effect on the situation, its causes and consequences.

This pair of activities forms a complete management cycle that, if for any reason the management goal is not achieved, can be repeated.

The term "decision" we refer to information on how and what to do to achieve the management goal, that is, the program (plan) of activity.

It is obvious that in the decision making the key role is played by knowledge and information. And here it is necessary to make some adjustments.

## 3. Knowledge and Information in Situational Management

Currently, most experts shared paradigm DIKW [4], which with the concepts of **D**ata, **I**nformation, **K**nowledge, **W**isdom builds some kind of pyramid, where the base is data, and the apex is wisdom. Unfortunately, the paradigm does not give any clear definition of the terms used, nor any clear criteria for the transition from one level of the pyramid to another. So another paradigm is proposed, which establishes the primacy of knowledge in relation to information [5]. Information is regarded as a communicative phenomenon, as a manifestation of knowledge in the observable and tangible form.

With the invention of the computer the information has become not only a means of fixation, storage, sharing, replication of knowledge, but also the object of treatment.

Often the effectiveness of situation center is directly linked with the possibility of the application of advanced computer technologies. However, as it has been shown above, for any of the problems of situational management we can not offer universal method (plan, algorithm) to solve it. In practice, this method is individual for each situation; his choice and use are the prerogative of the expert who must possess a number of qualities, namely [6] the abilities:

to solve problems, the method of the solution of which is fully or partially unknown (creativity);

to see or be aware of non-obvious problems (to be heuristic);

to draw conclusions about the object, not even knowing which way his thought came to this conclusion (to have intuition);

to provide for or anticipate the future state of the object (to predict);

to oppose his own opinion of the majority opinion (independence);

to see a problem from different perspectives (comprehensiveness);

to formulate specific recommendations for action (constructiveness).

To this we must add professionalism - the ability to deal effectively with the problem of certain subject areas.

In addition, the work of the experts should take into account not only the above-mentioned quality experts, but also aspects such as the obligations, rights, ethics, morality and responsibility.

With regard to information, it also plays an important role in the process of decision making. The knowledge provided by experts to the information level hold a special place in it.

In the situation center expert knowledge and information are subjects to management.

A prerequisite for the organization in the situation center expert knowledge management is availability of its own pool of experts [7]. Then the knowledge management includes the following tasks:

formation of general-purpose pools of experts;

formation of specialized pools of experts;

formation of experts group "for the problem";

ratings of experts according to monitoring the decision making process;

ratings of experts according to monitoring the decision implementation;

training of experts;

extraction of expert knowledge and presenting them at the information level.

Information that the situation center could / should have and should be the object of management is extremely diverse. It can be divided into:

passive information - all sorts of databases, directories, classifications, descriptions, etc. In [8] the categories of information used in situation center was observed;

active information - software, including expert systems, simulation models, and so on, and the information technologies used to support activities [9].

## 4. Place of the Situation Center in Governance Structures

The space, in which it is necessary to position the situation center, is the existing management infrastructure. To simplify the picture, we can say that it is divided into [10];

horizontal - on the basis of specialization (ministries, departments, administrations, etc.);

vertical - in terms of delegated competence, forming the corresponding hierarchy.

The first question to which the answer must be sought is whether the situation center takes the position in the specialized substructure, or takes an intermediate position.

If the range of problems solved by situation center has signs of specialization, it is clear that it has to be implemented into a corresponding substructure. But how? There are two possibilities:

the situation center can be integrated with an existing element of the structure to be "dissolved" in it;

the situation center can be a separate entity.

In any case the situation center must have clearly defined functions, i.e. the duties to solve certain problems of situational management (see above). And there are other, no less important, issues arising from the expert's nature of the situation center:

who is a consumer of the results of the situation center?

what weight have (should have) the results of situation center for its consumer?

what rights have the situation centre in the relationship with the environment?

what is the responsibility of the situation center to the consumer?

These issues can not be ignored. They have to be answered in any particular case.

With regard to the positioning in the "vertical", it is clear that the situation center should take place somewhere in the higher levels of the hierarchy.

Perhaps the situation center can take an intermediate position in relation to the specialized management structures, created to solve complex interdisciplinary problems which occur permanently or suddenly.

Another possibility, which requires a detailed study, is the creation and using of the situation centers networks.

## 5. Status of the Situation Center

It is difficult to imagine a situation center, which is not part of the legal entity or a separate legal entity. That status of legal entity commits to clearly define responsibilities, rights, responsibilities of the situation center.

In addition, the situation center can be commercially profitable organization which provides customers with a certain range of services to support the decision making process with the involvement of experts. Obviously, this is not possible without a legal entity status.

## 6. Situation Center as a Service Organization

The situation center can be regarded as a service organization. In this regard, we can engage in the creation of "complete" list (catalog) of services that the situation centers can provide. For example, such list of services may include:

selection of experts for a particular purpose;
providing a specially equipped room for collective work of experts;
supporting the collaborations of experts in the remote access mode;
communications with the monitoring objects in online mode;
automated reliable identification of the participants of the experts meeting;
automated service line for queuing of speeches;
support of the processes of reports (speeches) with the provision of audio and video services;
simultaneous translation;
automation of the voting process;
documenting the process of the meeting of experts;
automated generation of meeting protocols;
ratings of experts according to monitoring the decision making process;
ratings of experts according to monitoring the decision implementation;
ensuring the confidentiality, etc.

Each specific situation center will match a subset of the services included in such list.

## 7. Conclusion

As a result of consideration the space of the essential aspects of the activities of the situation center as a tool for situational management was defined. It can be used to determine for each particular situation center the spectrum of its objectives in situational management, as well as its status and location in the management structure.

## References


1. Поспелов Д. А. Ситуационное управление: теория и практика. – М.: Наука. – Гл. ред. Физ.-мат. лит., 1986. – 288 с.
2. Малышев О. В. Задачи ситуационного управления. – В кн. «Сучасні проблеми інформатики в управлінні, економіці та освіті» : [матеріали XIII міжнародного наукового семінару] / за наук. ред. д.е.н., проф. М. М. Єрмошенка. – К.: Національна академія управління, 2014. – С. 27-33.
3. Quality management systems – Fundamentals and vocabulary. – International standard. - Third edition 2005-09-15 – 38 p.
4. Ackoff R. L., "From Data to Wisdom", Journal of Applies Systems Analysis, Volume 16, 1989 p 3-9.
5. Малышев О. В. Анти-DIKW. – Обчислювальний інтелект (результати, проблеми, перспективи): Матеріали 1-ї Міжнародної науково-технічної конференції (10-13 травня 2011 р. Черкаси). – Черкаси: Маклаут, 2011. – С. 104-105.
6. Маліцький Б. А., Попович О. С., Соловйов В. П. Методичні рекомендації щодо проведення прогнозно-аналітичного дослідження в рамках Державної програми прогнозування науково-технологічного та інноваційного розвитку України // Наукове видання Центру досліджень науково-технічного потенціалу та історії науки НАН України. – Київ, "Фенікс", 2004. – 52 с.
7. Малышев О. В., Симонов С. В. Формирование и использование пула экспертов Ситуационного центра. – Системи підтримки прийняття рішень. Теорія і практика: Збірник доповідей науково-практичної конференції з міжнародною участю. – 8 червня 2009. – Київ: Інститут проблем математичних машин і систем НАН України. – 2009. – С. 117-120.
8. Малышев О. В. Информационное обеспечение деятельности ситуационного центра. – Системи підтримки прийняття рішень. Теорія і практика: Збірник доповідей дистанційної науково-практичної конференції з міжнародною участю. – 3 червня 2013. – Київ: Інститут проблем математичних машин і систем НАН України. – 2013. – С. 56-59.
9. Малышев О. В. Комплексная информационная технология для ситуационного центра. – Тези доповідей II Міжнародної науково-практичної конференції «Інформаційні технології в освіті, науці і техніці» (ІТОНТ-2014): Черкаси, 24-26 квітня 2014. – У 2-х томах. – Черкаси: ЧДТУ, 2014. – Т. 1. - С. 102-103.
10. Малишев О. В. Позиціонування ситуаційного центру. – У кн. Матеріали Всеукраїнської науково-практичної конференції «В. М. Глушков – піонер кібернетики» (2014 р. м. Київ) / Укладачі: В. В. Новіков, А. А. Мельниченко, В. Д. Піхорович, І. В. Виселко, В. Ю. Пряміцин / – К.: Видавництво «Політехніка», 2014. – С. 106-108.

# Social Aspects of Shadow Information Economics

**Serghei Ohrimenco[1], Grigori Borta[2]**

[1] Laboratory of Information Security, Academy of Economic Studies of Moldova
osa@ase.md

[2]Laboratory of Information Security, Academy of Economic Studies of Moldova
grigori.borta@gmail.com

**Abstract.** This paper discusses the issues of illegal activity in the domain of information technologies. A definition is proposed and key factors of the struggle against the phenomenon are analyzed.

**Keywords.** Information security, shadow information economics, underground economics.

## 1. Introduction

This paper aims to draw attention to the problematic situation of the shadow information economics. The importance of shadow information economics study is demonstrated by the fact that it currently is one of the least explored segments of economics and is being paid much less attention than is required, even regardless of its importance for the researchers in the domains of economics and information technologies.

## 2. Definition of Shadow Information Economics

The following definition for shadow information economics is proposed: all the individual and collective activity that is illegal, related to research, design, development, distribution, support and use of information and communication technologies, that is hidden from society. Shadows information economics has characteristics, functions and descriptions in common with general shadow economics.

## 3. The Role of the State

According to the national strategy of information society development called "Digital Moldova 2020", information and communication technologies are an inseparable part of both state and private sector development policies. This kind of strategy is essential for the development of quality of the services offered to the population. The strategy approved in the Republic of Moldova is based on three main pillars:

- infrastructure and level of access development
- digital content and electronic services
- increasing the level of competence in information technologies.

It is expected of these improvements to benefit the society by means of offering much more comfortable conditions for living, business, and government structures.

The following problems are the most important among currently existing ones:

- lack of a well-developed infrastructure outside of the major cities

- limited access to information infrastructure
- psychological barrier when using this kind of services
- low development level of national information resources
- unavailability of national cultural and scientific heritage in digital format
- unavailability of adequate legal basis

Another major issue is inadequacy of information security controls that are currently in place. For example most of data is being stored in data-centers that do not correspond to all the appropriate requirements, and a lack of qualified personnel is obvious. Another important issue in data storage is very broad geographical distribution of data-centers, which is a major drawback when maintenance costs are considered.

The document also notes that around 37.7% of Internet users make online purchases, the vast majority of which are from foreign web-sites. It is important to note that this kind of statistics bear evidence to poor national internet resources development level. Another important issue is the fact, that little to none digital content is created in Moldova. This index is measured by the number of articles created in Wikipedia in the language of the country, number of blogs by its citizens, and the number of registered first level domain names. One of the reasons behind this situation may be low overall education level. For example, the document states that only 14% of the population aged 35 to 44 years old use Internet. In the 45-54 age group only 10% of the population is Internet users. The proposed solution to this problem may be the following:

- introduction of new educational programs
- qualified teaching staff preparation
- integrating information technologies into educational process
- creating digital educational content

One of the most important issues underlined in the document is the fact that the more a society is developed in information technology, the more it is susceptible to cyber threats. Therefore, providing a high level of cyber security is a major priority for each and every participant of the process, especially for the organizations responsible for developing policies in the domain. The document proposes the following definition: “cyber security is a state of normality, established as a result of proactive and reactive measures taken in order to establish confidentiality, integrity, availability, authenticity, and reliability for the information in electronic form, government and private resources, and services in the cybernetic environment”.

Vulnerabilities are categorized into the following three groups:

- human factor
- technical
- procedural

Along with vulnerabilities the following classification of threats is proposed:

- cyber attacks
- unauthorized access
- unauthorized modification
- cyber espionage
- property damage

The amount of attention drawn to this issue avidly demonstrates the need to inform all the users about the major risks when they make payments via the Internet or publish their personal data, and the measures that need to be taken in order to avoid them.

A similar strategy is being implemented in European Union. Digital economics grows much faster than all the other economic domains, yet still it is considered to have not reached it potential as of yet due to the lack of a corresponding policy on the European level. The potential is expected to be fulfilled owing to the new digital strategy. GDP is expected to grow by 5% or 1500 euro per capita in the following 8 years due to investment in information and communication technologies. The main priorities are:

- creating legislative base
- improving digital government services infrastructure
- creating work places in the information technologies industry
- revising the copyright legislation
- stimulating the development of cloud technologies
- developing an industrial strategy in the domain

One of major priorities is the development of information security strategy on the EU level.

In the US a Comprehensive National Cybersecurity Initiative was launched in January 2008. It aims to focus on the following domains:

- Forming a frontline defense against immediate threats by informing the government and all the levels of administration on state and private levels about existing threats as well as reacting to them as soon as possible and preventing intrusion.
- Defense against most of the threats by improving counterintelligence and fortifying supply lines of key technologies.

Strengthening the future climate of information security by improving education in the domain, coordinating research, and working on creating and improving a strategy to prevent malicious actions in cyberspace.

## 4. Conclusion

This paper does not cover all the multitude of problems raised by the phenomenon of shadow information economics. A much more profound research of the domain is required in order to develop effective struggle strategy. The research should range from the actions targeted at vulnerabilities research in software, to organized attacks directed at the corporate and public web sites and infrastructure, to the conditions necessary for the "offshore terrorism" to form and function.

## References

1. Буров, В. Исследование Теневой Экономики. *Известия.* [Online] 06 2010. [Cited: 09 05, 2014.] http://eizvestia.isea.ru/pdf.aspx?id=13936
2. Imperva. Hacker Intelligence Initiative, Monthly Trend Report #13. *imperva.com.* [Online] 10 2012. [Cited: 07 07, 2014.] http://www.imperva.com/docs/HII_Monitoring_Hacker_Forums_2012.pdf
3. Georgia Institute of Technology. Emerging Tyber Threats Report 2013. [Online] 2012. [Cited: 07 07, 2014.] http://www.gtsecuritysummit.com/pdf/2013ThreatsReport.pdf.
4. Panda Security. *The Cyber-Crime Black Market Uncovered.* 2010
5. Bilton, Nick. How Credit Card Data Is Stolen and Sold. *The New York Times.* [Online] 05 03, 2011. [Cited: 05 21, 2014.] http://bits.blogs.nytimes.com/2011/05/03/card-data-is-stolen-and-sold/?_php=true&_type=blogs&_r=1
6. Friedrich, Schneier and Johannes, Kepler. The Shadow Economy in Europe 2013. [Online] 2013. [Cited: 02 03, 2014.] http://www.protisiviekonomiji.si/fileadmin/dokumenti/si/projekti/2013/siva_ekonomija/The_Shadow_Economy_in_Europe_2013.pdf
7. Boehme, Rainer. Vulnerability Markets. What is the economic value of a zero-day exploit? [Online]
8. Schneier, Bruce. The Vulnerabilities Market and the Future of Security. *Forbes.* [Online] May 30, 2012. [Cited: July 22, 2012.] http://www.forbes.com/sites/bruceschneier/2012/05/30/the-vulnerabilities-market-and-the-future-of-security/

# About the Teaching Accounting and the Audit by Using the Information Technologies

**Tea Munjishvili**

Tbilisi, Georgia
tmunjishvili@gmail.com

**Abstract**. Teaching the electronic producing of accounting is discussed as the main part of teaching theory and practice of accounting. Is shown, that practical lessons must be held by the computer program of knowledge revealing and rating. Computer programs used in the workshops, laboratory and self-study. The study material by the program can be online in a teacher, and without it. Capabilities of our program: The study of e-learning materia; Online access to information of interest; Ability to osvoivat specificity of application of different options. Development of problem solving skills in accounting; Checking the level of knowledge of problem solving and testing; Development of problem solving skills in accounting.

**Keywords**: Accounting. Computer program; E-learning; Skills; Tutorials.

## 1. Introduction

Object of our research is the problematic of teaching the accounting and audit. Nowadays, in factories electronic producing of accounting is performed. Several types of computer programs are used. In Georgian market, are widely used computer programs: 1C, Oris, Oris-manager, Superfin, Global-exec and so on. In market, new programs, with some different parameters, are used.

Time has come, to create intellectual computer program of electronic producing of the accounting, by using the received results.

Now and in nearest future, no matter how perfect is computer program of accounting, the centre of it will be accounting specialists.

In Georgian economics, share of individual and tiny factories is 77% of current factories. Every head of the factory needs a high qualified accountant, which perfectly knows the accounting theory, practice, and electronic producing.

Request of the heads of factories is simple – Economist with diploma, by the profile of accounting, which perfectly knows theory and practice of electric producing. Completeness of the teaching methods in subjects of accounting and audit, to satisfy the market requests, is actual. Existing approach, when teaching is performed by explaining on the board, writing lection by the student, practical lessons, is not accepted.

## 2.Problem analysis

Nowadays, for "satisfying" the market requirements, new subjects are admitted – teaching the electronic producing of accounting with a specific computer program. We have a situation like: theory and practice are separated. Classic war between old and new

continues. In some subjects, like natural science, new method easily wins, in some other subjects the old method is better. This situation is in a studying field of accounting and audit.

Teacher of accounting, who knows theory of accounting, methods and he considers that knowing the utilities of electronic producing of accounting is a secondary. The result is, that not only old and high qualified teachers, also junior lecturers in economics, who choose an accounting as a narrow specialty, have general imagination about accounting computer programs and the more about creative use of producing the accounting by using them.

The majority of such teachers considers, that electronic producing of accounting limited by pressing the computer button and that it is enough to input the data and the program do the rest.

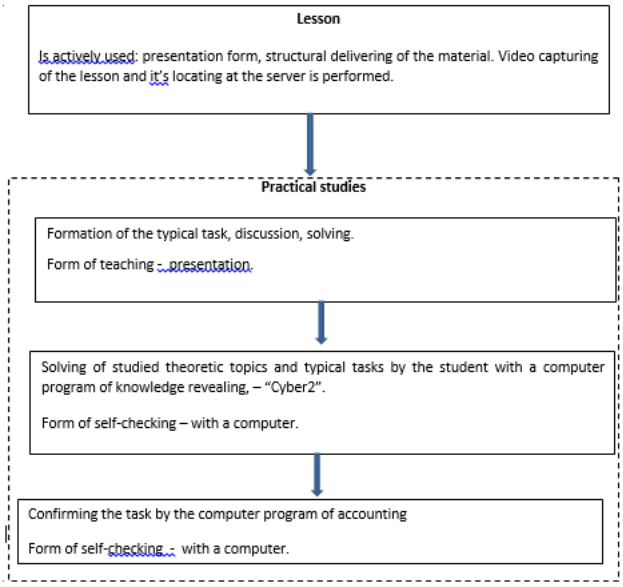
Since old times, it is known that accountants are hardworking and practical people, and they have a good flair to understand created situation. Among the subjects of accounting and audit, teaching the electronic producing, by the computer program they give a low priority, or find a way to create the special expensive curses to teach computer programs of electronic producing the accounting (which had to be learnt at university).

## 3. Solution ways

Creating the specialized curses or reducing the number of students in studying group, is not a way out. The existing method of teaching the accounting and audit should be principally changed – teaching should be performed in sphere of spread accounting in the market of the specific country, with a close relationship to the existing newest computer program. It should be whole with a theory. Nowadays, a spread tactic – teaching of the separate computer program, is harmful and unacceptable.

On the example, of teaching the subjects of accounting with the newest method, the issue of changing the teaching methods of economic profile subjects, can be questioned.

Our view of teaching the subjects of accounting and audit, in which we can't see the special thing, is following: /pic. 1/



Pic.1

Studying technologies are clear. As we see, the necessary conditions to held lesson are: Existing the video projector, local network or the internet, teacher, computers for students and server, in the auditory.

In the server, we have information about students, subjects, practical and theoretical (task) materials introduced by the text, graphic, video information, examples of solving the typical tasks and other. In the computer of client (student, teacher) are installed programs: the first – system of knowledge revealing and rating "Cyber2", with a help of which is performed the learning of materials, introduced by the teacher, self-checking and the second- the special program of accounting, like "Oris", or other.. With this program, generating of the skills of electronic producing of accounting is performed.

Fundament of lesson is the creating base of theoretical and practical materials, with tests and tasks, by the teacher.

Knowledge revealing with the computer program "Cyber2", consists of two parts: checking the theoretical preparation and practical skills. For this aim, we use two kinds of tests and tasks – generally tests: open and closed. Tests are called closed, where, from maximum 7 possible answers, must be chosen maximum 3 correct answers. Open tests are, where correct answer(s) must be written by student. In open tests (tasks), number of answers is not restricted.

From the view of knowledge rating, it is possible to rate answers with partial or full rating principles. In open tests, several answers can be rated with different points.

Answer can be a word, number, sentence presented: with narrative sentences, or with a combination of words and numbers. Words can be written in different declension, also – wrong variants of words.

As practice has shown (At Batumi Shota Rustaveli State University, the previous version of "Cyber2" is used, with which final and current exams are held), that rating the closed tests should not be partial. According to the number of correct answers, it is desired to have the following percentage distribution:

| *№* | *Product Name* | *Share %* |
|---|---|---|
| 1 | Tests with one correct answer from 3 – 7 answers | 10 |
| 2 | Tests with two correct answer from 3 – 7 answers | 30 |
| 3 | Tests with three correct answer from 3 – 7 answers | 60 |

Rating the open tests (tasks) is usually partial. According to teachers opinion it can be non-partial.

In the majority of subjects of accounting, an accounting handling of agricultural operations holds the enough share. The standard structure of accounting handling is =: <debt><account number><meaning><credit><account number><meaning>. In this structure <account number> is a positive integer, <meaning> - is positive integer or decimal. In the structure words: debt, credit or the variants of these words: d,k, deb, cr and so on, are used. Every accounting handling should be written with exactly analogical structure. Here replacing of words or missing/adding them is not allowed.

In "Cyber2" we are considering this moment and we perform the semantic analysis and diagnostic of answers, written by the student. During working with the system, student can receive a help in test or subject, from the teacher. Help on subject can be an accounting

reports, and help on test – solving or explaining the typical or specific task. It has a special meaning during working in the training mode.

The fundament of practical classes is, according to the syllabus of the subject:

1. Created opened and close tests (tasks)
2. Help on the subject
3. Help on tests (tasks)

Tests, created by the topic, can be ordered by the difficulty. Any test can be described with graphical or video files.

Before the class, teacher determines practical and theoretical materials of the lesson, chooses the tests according them and gives these tests to the system. Assimilating of materials is performed in the training mode. Duration of training is determined by the teacher.

During the training, with the teacher, discussing of subjects according the training, Finding the new way of task solving is performed. After training, in the presentation mode by the teacher, the teaching of electronic handling discussed typical task and assimilating of skills of using the exact accounting program.

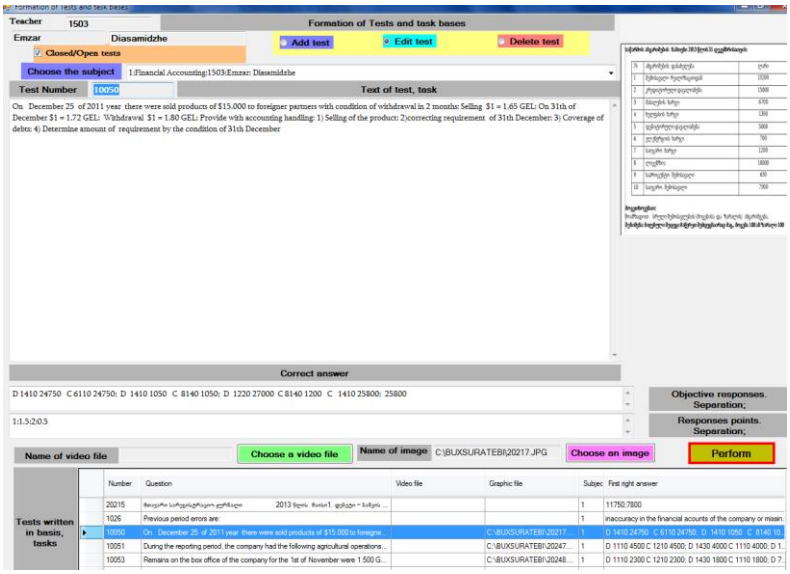
In next lesson, rating of assimilated topic/subtopic is performed with the "Cyber2". And knowing of the realization of assimilated materials is performed with using the electronic program of accounting.

Tests are chosen with a probability.

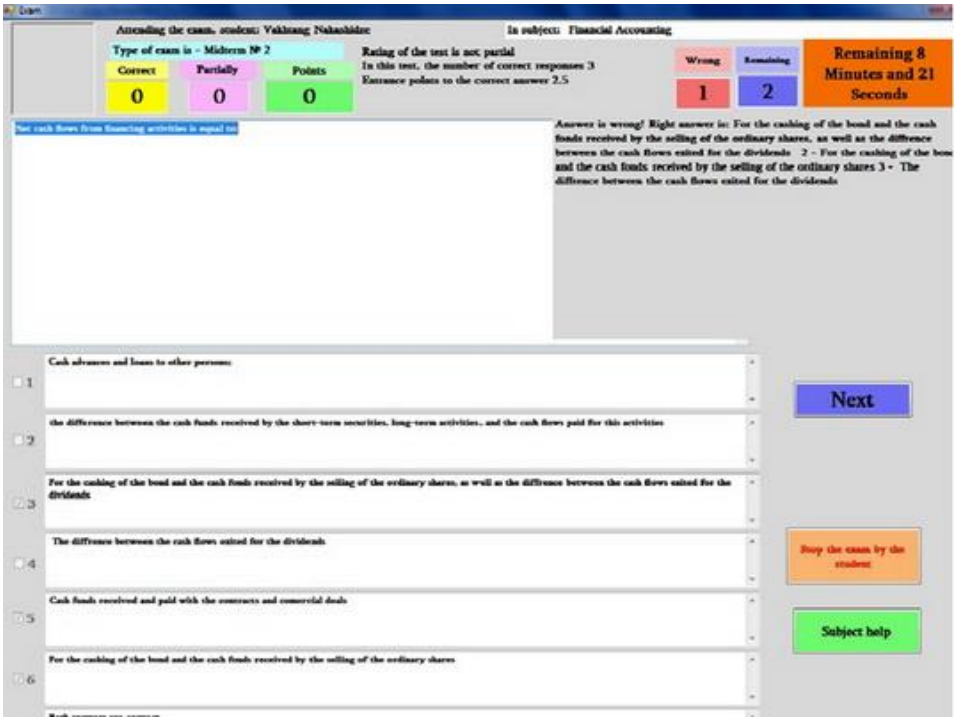
Training/exam can be finished:

Because ending the examination time, or after answering the all questions or by the student. In the program the stopping points, that means, that the stopped exam, because of technical or some other reasons, will be continued from the stopping place. After finishing the training/exam, student can have a review of the report, which reflects the duration of exam/training.

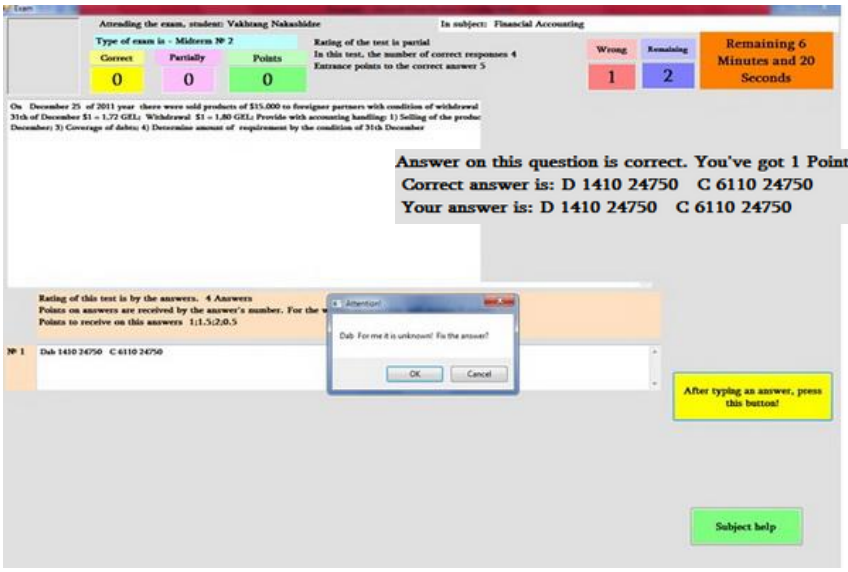
On the picture 2, 3, are shown dialogical windows, of tests/tasks, training, activities (in practical classes). Picture 4, shows us the process of answering the closed tests in the training/exam mode. On the picture 5, is reflected the moment, when in the open tests, student is typing the wrong word instead of the word "debt".


Formation of Tests and task bases
Teacher
1503
Emzar
Diasamidzhe
Closed/Open tests
Add test
Edit test
Delete test
Choose the subject
1:Financial Accounting;1503/Emzar Diasamidzhe
Test Number
10050
Text of test, task
On December 25 of 2011 year there were sold products of $15.000 to foreigner parmners with condition of withdrawal in 2 months: Selling $1 = 1.65 GEL: On 31th of December $1 = 1.72 GEL: Withdrawal $1 = 1.80 GEL: Provide with accounting handling: 1) Selling of the product: 2)correcting requirement of 31th December: 3) Coverage of debts: 4) Determine amount of requirement by the condition of 31th December
Correct answer
D 1410 24750 C 6110 24750; D 1410 1050 C 8140 1050; D 1220 27000 C 8140 1200 C 1410 25800; 25800
1:1.5;2:0.5
Objective responses. Separation;
Responses points. Separation;
Name of video file
Choose a video file
Name of image
C:\BUXSURATEBI\20217.JPG
Choose an image
Perform
Tests written in basis, tasks
Number
Question
Video file
Graphic file
Subjec
First right answer
20215
2013
1
11750;7800
1026
Previous period errors are:
1
inaccuracy in the financial acounts of the company or missin.
10050
On December 25 of 2011 year there were sold products of $15.000 to foreigne...
C:\BUXSURATEBI\20217...
1
D 1410 24750 C 6110 24750; D 1410 1050 C 8140 10...
10051
During the reporting period, the company had the following agricultural operations...
C:\BUXSURATEBI\20247...
1
D 1110 4500 C 1210 4500; D 1430 4000 C 1110 4000; D 1...
10053
Remains on the box office of the company for the 1st of November were 1.500 G...
C:\BUXSURATEBI\20248...
1
D 1110 2300 C 1210 2300; D 1430 1800 C 1110 1800; D 7...

Pic.2


Attending the exam. student: Vakhtang Nakashidze
In subject: Financial Accounting
Type of exam is - Midterm № 2
Rating of the test is not partial
In this test, the number of correct responses 3
Entrance points to the correct answer 2.5
Correct
Partially
Points
0
0
0
Wrong
Remaining
1
2
Remaining 8 Minutes and 21 Seconds
Net cash flows from financing activities is equal to:
Answer is wrong! Right answer is: For the cashing of the bond and the cash funds received by the selling of the ordinary shares, as well as the difference between the cash flows exited for the dividends 2 - For the cashing of the bond and the cash funds received by the selling of the ordinary shares 3 - The difference between the cash flows exited for the dividends
1
Cash advances and loans to other persons;
2
the difference between the cash funds received by the short-term securities, long-term activities., and the cash flows paid for this activities
3
For the cashing of the bond and the cash funds received by the selling of the ordinary shares, as well as the difference between the cash flows exited for the dividends
4
The difference between the cash flows exited for the dividends
5
Cash funds received and paid with the contracts and commercial deals
6
For the cashing of the bond and the cash funds received by the selling of the ordinary shares
Next
Stop the exam by the student
Subject help

Pic.3.




Pic.4

Program is multilingual. It is written on the VB.Net 2010. The database is organized on the platform of SQL Server 2008, Windows server 2008.

## 4. Conclusion

Teaching the electronic producing of accounting is the inseparable part of theoretical and practical teaching. Process of studying accounting subjects must be in the computer class with explaining the theory, practice, with self-checking of materials, received with the computer program of knowledge revealing and rating and with the special program, by generating the skills of electronic handling.

## References

1. Tea Munjishvili, Zurab Munjishvili, Knowledge demonstration and assessment system “Cyber1”, international Journal "Information Technologies & Knowledge" Volume 8, Number 3, 2014, pp. 271-279.
2. Munjishvil T., Munjishvil Z., Nakashidze V. System of knowledge revealing and rating – “Cyber 2”. 9th MIBES ANNUAL INTERNATIONAL CONFERENCE 2014 THESSALONIKI, GREECE, 30/5– 1/6 .CD ISBN# 978-960-93-6161-3. pp. 111-121.

# Macroeconometric Model of the US

**Andrew Shevelev**

IEIE SB RAS, Novosibirsk State University, Russia
andrew.shevelev@ieie.sbras.ru

**Abstract.** The paper reviews the large-scale macroeconometric model of the US economy. Ideology and methodology of S.M. Menshikov, Y.A. Chizhov, P.N. Teslya were used for the construction of the model. Trend-cycle analysis were applied to improve the accuracy of the model. The results of the model simulation and GDP forecasting are presented.

**Keywords.** Macroeconometric model, macroeconomics, econometrics, trend-cycle analysis.

## 1.Introduction

The state of the global economic system is characterized by consistently intensifying of the globalization process. The development of the modern state is impossible without taking into account its interaction with the outside world. Analysis of a large economy system and the study of the relationships, patterns of development and the role of government regulation of a country are necessary for a proper understanding of the processes taking place in the global economy.

The choice of the US economy as the object modeling and forecasting is due to the dominant role and the expansion of economic influence in the world economy during the review period.

Calculation of the model is made on the basis of quarterly statistics for the United States between 1947 and 2010 (256 quarters). The number of used variables in the model is 86. The number of exogenous variables and endogenous is 31 and 55 respectively. Among the endogenous variables, 30 of it were estimated based on the regression equations and other 25 on the identity equation.

## 2.Model construction

In the first phase of the model construction the set of economic indicators used in the model is determined.

The second step is a choice of its individual relationships (equations) the structure, evaluation of the parameters of these links and selection of the options that in accordance with the general structure of the model form a single system.

The third stage is a usage of Gauss-Seidel algorithm for solving nonlinear simulation system of equations and identities, with a need to clarify the links and adjusted models, checking the model of the economy that it describes and assessment of the applicability for economic research.

The list of indicators were chosen in such a way that allows you to create a closed system, which describes the basic relationship in the US economy. The US economy is

represented by five major and one additional unit that comprises from free to six econometric equations.

Dominant capitalist nature of demand in the US economy was supposed at the heart of the choice of the general structure of the model. Aggregate demand has been disaggregated to the consumer, investment, government and demand from foreign trade (balance of exports and imports of goods and services).

In turn, consumer demand has been split into qualitatively different kinds: the demand for cars, other durables, non-durables and services.

Investment demand disaggregated on the net investment in fixed assets, housing investment, depreciation and stocks investments, i.e., changes in inventories at the end of the current period.

State demand was subdivided into purchases of civilian and military products.

The demand for exports of goods and services makes from the part of foreign trade, while the part of domestic demand is met through imports. Therefore, foreign demand was presented by export and import of goods and services.

Because of different categories of generated demand, primarily on the basis of income received under the influence of factors such as prices, general conditions in the economy as a whole and its individual sectors, the model should include a control unit that describes the dynamics of these factors.

The main functions that show the dynamics of income are functions of personal and corporation income. The assessment of the dynamics of the general fund wages of that is divided into wages in the private sector and the public sector is one of the most important.

Wage fond, in turn, depends on the rates and the number of employed workers. Therefore, wages rate of the private sector and civil servants were included in the model separately. In addition, the model includes employment in the private sector and the public sector (civilians and military personnel).

The difference between the value of national income received by subtracting depreciation fund and indirect taxes from the total gross domestic product, and the wage fund represents unearned income. The most important element of unearned income is corporate profits out of that the dividends are paid to the shareholders. In addition, the category of unearned income also includes: income of unincorporated sector, interest and rental income.

Formation of the category of net income, the state revenue and expenditure occurs by the introduction of various taxes, transfer payments, as well as contributions and social security payments. The difference between income (taxes and other fees) and expenses (government procurement of goods and services) forms the balance of the state budget.

Assessment of the economy by the production was presented as the ratio of actual to potential product. The latter was obtained in the form of the production function in full capacity utilization and labor resources.

Accounting for the impact of accumulated reserves of capital goods and consumer durables for the current demand requires the formation of these parameters on the basis of summing up sales for a certain period are adjusted to their disposal.

A simplified overall structure of US econometric model is represented by the following flowcharts (Fig. 1). The solid and dashed lines indicate identical communication and functional respectively. Identical communication means that these figures are not dependent probability and obtained one from another by the definition. For example, the wage fund is calculated as the product of wage rate on the number of employees; net unearned income is the difference between the unearned income and taxes from them. In

contrast to the identical, functional relationships reflect the structure of the probabilistic dependencies, i.e. the regression equations.

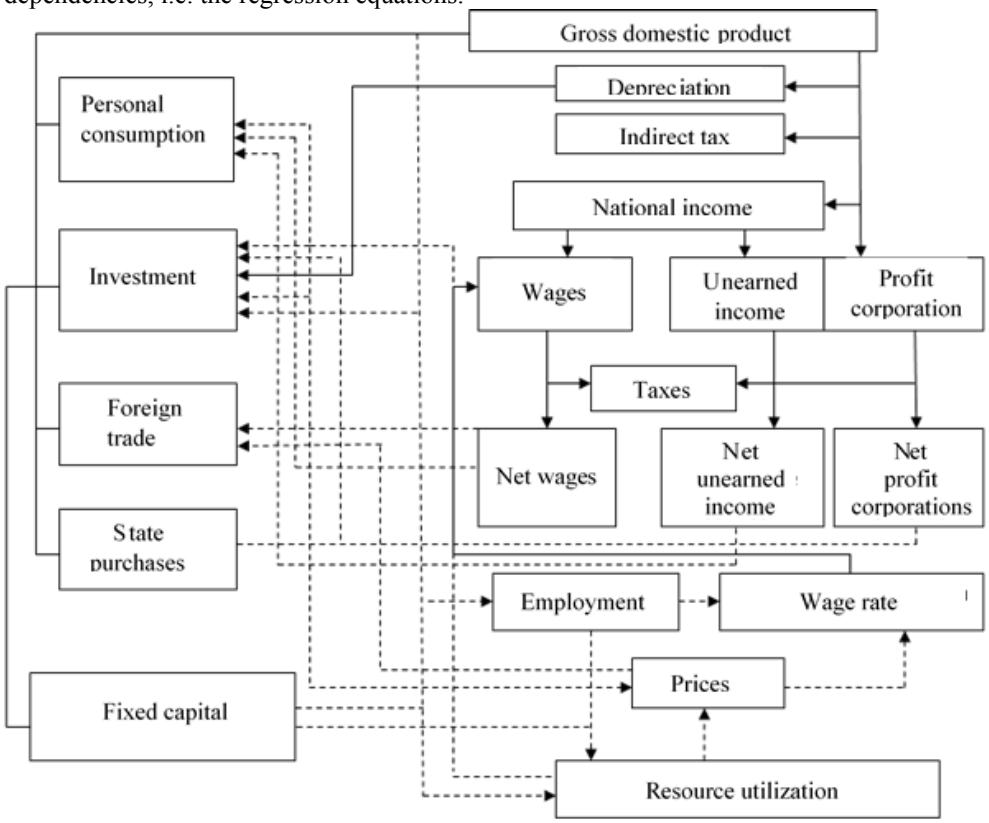



Figure 1 – The block structure diagram of a model US

Method of assessing each of econometric equations of the model is as follows. First, a set of variables that have an impact on the dynamics of the corresponding figure is chosen by the expert method. Lags of indicator values are considered for the time dependence of each other. Then, equation parameters are assessed by stepwise method (sequential linear regression includes the most important indicators based on partial correlations). The use of this tool allows us to not only identify the factors that influence the rate, but also to rank them in order of importance. At the same time, this method allows to find the optimal lag influencing factors, as well as to solve the problem multicollinearity between the independent variables.

## 3.Gauss-Seidel simulation algorithm and evaluation of constructed model

We represent the original data of model in the form of the matrix $A[i, j]$, where $i$=1,..,256 are numbers of quarter from 1947 to 2010, $j$=1,..,55 are numbers of endogenous variables, $j$=56,..,86 are numbers of exogenous variables.

Constructed model represents the function $v=f(x)$, where $v$ – endogenous vector, $x$ – row vector of the matrix original data $A$. The proposed algorithm can be represented by the following flowchart (Fig. 2).

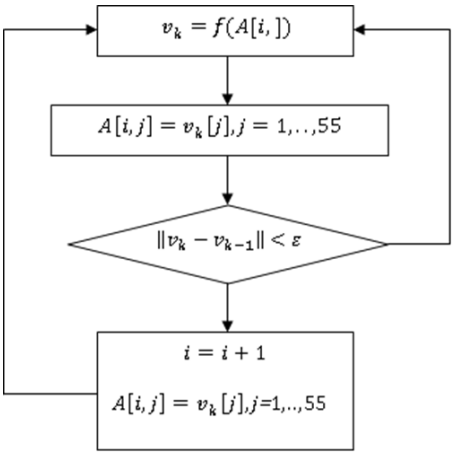



Figure 2 – Block structure diagram of Gauss-Seidel simulation algorithm

The original data are taken for the initial quarter. Each subsequent are calculated values obtained in the iterative process with prescribed accuracy. Exogenous variables retain their real values for each quarter.

In order to assess the general nature of the interaction of the endogenous variables, the model was implemented on 220 quarters ahead, starting with the first quarter of 1956, assuming that all this period of exogenous variables retain their real values. Despite the considerable length of the settlement period and significant effects of the global crisis, simulation of GDP was quite close to its real dynamics. The deviation does not exceed 11%.

## 4.GDP forecast

After debugging the model and evaluating its basic properties the leading direction of its applied use is the applying as an analogue of the economic system in simulation and forecasting tasks.

Simulation process with the first quarter of 2010 to the fourth quarter of 2012 was implemented for GDP forecast. Spectral analysis of error terms revealed patterns of movement not reproducible by the model and described them using the trend-cycle analysis that improved the quality of the forecast. Random deviations are interpreted as the effect of unaccounted factors in the model.

Constructed multifactor dynamic model of the US economy allows us to calculate forecasts for the 55 endogenous variables reflecting the state of the US economy.

According to the model calculations, GDP of US increases at 3% for 2 years (2010-2012), and almost 4% in reality. Forecast error in the 4th quarter of 2012 was 1.4%

## 5.Conclusion

Building a multifactor economic model of large dimension covering virtually the entire movement of the economic mechanism of the country over the long term is theoretical

interest in order to study the qualitative changes of the system itself and from the perspective of improving the methods of estimation of the market economy as a whole.

The constructed model reproduced the course of the US economy with high accuracy over the period. Accuracy EX-post forecast for the four quarters was 1.4%.

The ideology of submission of a functioning market economy (S.M. Menshikov, Y.A. Chizhov, P.N. Teslya, etc.) was confirmed by results of the model.

Expansion of the method at the expense of the trend-cycle analysis improves quality evaluation model.

Forecast calculations confirmed the steady growth of the US economy.

## References

1. S.M. Menshikov, Modeling of the US economy. Novosibirsk: Nauka, 1975.
2. Y.A. Chizhov, A.P. Ermilov, Economic forecasting of capitalist economy. Novosibirsk: Nauka, 1982
3. G.D. Kovaleva, Application of the theory of time series analysis in economic research. Novosibirsk State University, 2008.

A. Clopper, The craft of economic modeling. University of Maryland, 2008.

4. G.D. Kovaleva, I.V. Kozachek, P.N. Teslya, The report, "improving macroeconomic model short-term forecasting of the US economy." Novosibirsk, 1981.

# Tourist Behaviors in Online Booking: Empirical Evidence from Hotel Industry in Sri Lanka

**Thilini Chathurika Gamage[1], Fazeela Jameel Ahsan[2]**

[1] Department of Marketing Management, Faculty of Management Studies, Sabaragamuwa University of Sri Lanka, Sri Lanka
thilinicg84@gmail.com

[2] Department of Marketing Management, Faculty of Management and Finance, University of Colombo, Sri Lanka
dr.fazeela@gmail.com

**Abstract:** This paper seeks to establish the relationships between tourists' attitude, perceived ease of use, perceived usefulness, trust, perceived risk and online booking intention of tourism and travel related products by using an extended technology acceptance model. Primary data were collected using a self-structured questionnaire from 283 tourists visiting Sri Lanka in 2013. Seven out of eight hypotheses tested were supported. Perceived ease of use was not a significant predictor of attitude towards online booking. The model can assist destination marketing practitioners to better understand the online booking behavior of tourists.

**Keywords: Online Booking, Tourists, Behavior**

## 1. Introduction

Use of the Internet is becoming a positive trend and an important core competency for travel/ tourism industry over the last decade, and hotel industry is no exception (Heung, 2013). Research indicates that 76% of leisure and business travelers consider travel Websites when they are booking hotels, compared to only 24% who use the corresponding hotel's brand Website (Buhalis & Jun, 2011). Some of the top hotel booking sites include Hotels.com (18%), Expedia (15%) and Priceline (10%). However, the dramatic development of online booking has also brought forth many new challenges. Risks arising from lack of human contact, lack of previous experience, lack of trust inducing factors could impede on tourists' trust in online booking (Main, 2001). The full potential of online booking can only be achieved if the determinants of tourists' online booking intention are known to researchers and practitioners. So far these aspects have received little attention in existing tourism/ consumer behavior literature. This paper fills this research gap by extending the use of technology acceptance model (TAM) to investigate the online booking behavior of tourists specially referring to hotel industry in Sri Lanka.

## 2. Theoretical Background: Technology Acceptance Model (TAM)

Pavlou (2003) defines online purchase intention as a situation where a consumer is willing and intends to make online transactions. Online purchase intentions are explained in part by the TAM (Davis, 1989). The TAM model (See Figure 1) constitutes of two belief factors namely; perceived usefulness and perceived ease of use, which are predictors of user attitude towards using technology. In turn, the user attitude towards using technology influences behavioral intention. Perceived ease of use also influences perceived usefulness of technology.

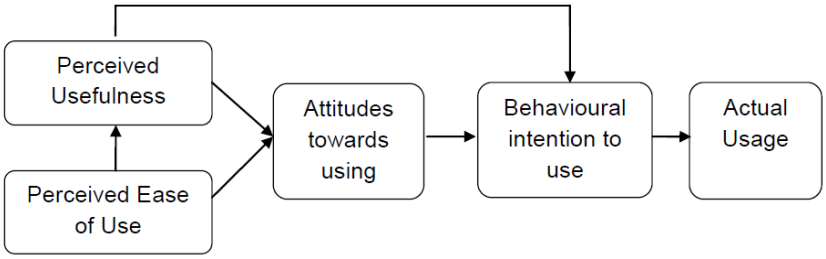



**Figure 1 Technology Acceptance Model (Davis, 1989)**

**Perceived Usefulness**

According to Davis, Bagozzi, and Warshaw (1989), perceived usefulness (PU) can be explained as the extent to which the consumer believes that using the system can increase his/ her performance. From a perspective of online booking in hotel industry, PU can be defined as the prospective tourist's subjective probability of using the Internet to efficiently facilitate his/her hotel service/ product purchasing. Childers, Christopher, Peck, and Carson (2001) and Chen (2006) found that PU affects attitudes towards online shopping. More specifically, in hotel industry context Heung (2013) and Khemthong and Roberts (2006) have also emphasized that PU influences the booking intention of potential tourists. Thus, it can be conclude that

***H1:*** *There is a direct positive relationship between perceived usefulness and attitude towards online booking.*

***H2:*** *There is a direct positive relationship between perceived usefulness and intention towards online booking of tourism and travel related products.*

**Perceived Ease of Use**

Perceived ease of use (PEOU) is defined as the degree to which a person believes that using a particular system would be free from effort (Davis, 1989). Many researchers (Aladwani, 2002; Moon & Kim, 2001) have studied the relationship between PEOU and PU. Nonetheless the relationship remains contradictory. Yet, past research on consumer adoption of online services in different research settings discovered PEOU as an important antecedent of user's adoption of new Web technology(Chen, 2006; Morris & Dillon, 1997; Pikkarainen, Pikkarainen, Karjaluoto, & Pahnila, 2004; Taylor & Todd, 1995). Hence, it is decided that

***H3:*** *There is a direct positive relationship between perceived ease of use of online booking Websites and perceived usefulness of the Websites.*

***H4:*** *There is a direct positive relationship between the perceived ease of use of online booking Websites and attitudes towards online booking*

**Trust in Online Booking Websites**

Kolsaker and Payne (2002) defines trust as the dimension of a business relationship that determines the level to which each party feels they can rely on the integrity of the promise offered by the other. Chen (2006) hypothesizes that a consumer's trust in a virtual store positively affects his/ her attitude toward using the e-store. Jarvenpaa, Tractinsky, and Vitale (1999) also found that trust in the online store directly influence attitude. Supported by the above literature, it can be hypothesized that:

***H5:*** *There is a direct positive relationship between the trust on online booking Websites and attitudes towards online booking.*

**Perceived Risks**

Perceived risk is defined as a consumer's subjective perception of the possibility to reach unexpected consequences (Dowling & Staelin, 1994). In this study, perceived risk refers to the uncertainties associated with possible negative consequences of online booking. Trust and risk are essential in explaining e-commerce adoption, as uncertainty is present in the technology-driven environment (Pavlou, 2003). Several studies show that perceived risk negatively influences trust in a given Website (Featherman, 2001; Pavlou, 2001). Thus, the following hypothesis is proposed.

***H6:*** *There is a direct negative relationship between perceived risk and trust in online booking websites.*

Perceived risks also impact on attitude towards online booking. Jarvenpaa et al. (1999) emphasize that risk perception affects consumer's attitudes toward an online-store. However, a study by Brown, Muchira, and Gottlieb (2007) highlight that even though perceived risk is noted as an important barrier to e-commerce usage, it does not seem to stop people from doing online booking. Yet, Wu and Wang (2005) study incorporated perceived risk into the TAM to evaluate consumers' adoption of mobile commerce, and reveals that potential risks of online transaction affect consumers' attitudes. Thus,

***H7:*** *There is a direct negative relationship between perceived risk and attitudes towards online hotel booking*

**Attitudes toward Online Purchasing**

Attitude is defined as a consumer's positive or negative feelings related to accomplishing the purchasing behavior on the Internet (Chiu, Lin, & L.Tang, 2005). In the context of this paper, attitude is defined as a tourist's salient belief of whether the outcome of his/ her use of the Internet for hotel booking will be positive or negative. Attitude is said to have a direct impact on purchase intention. The original TAM model (Davis, 1986), and the models of Taylor and Todd (1995) and Morris and Dillon (1997) indicate that attitude exerts a positive effect on behavioral intention. This relationship has also been proven in general e-commerce studies as well(Bruner & Kumar, 2005). Hence, it is decided that

**H8:** *There is a direct positive relationship between the attitudes and online booking intention*

## 3. Methodology

Following the positivistic research tradition and the quantitative research approach, in this paper survey research strategy was adopted. Self-administered questionnaire consist of five-point likert scale ranging from 5 = strongly agree to 1 = strongly disagree was used to collect the data. The questionnaire included items that were derived from the Davis (1989) and Taylor and Todd (1995) and were slightly modified to hotel industry context.

Hotel industry in Sri Lanka was used as the research setting of this paper due to its concrete effort to strategically position Sri Lanka in the international markets through its online initiatives. The list of hotels in Sri Lanka was obtained from the Accommodation Guide published by Sri Lanka Tourism Development Authority and Western and Southern region hotels were abstracted from the list. A Google search was done for all star-class

Western and Southern region hotels and the hotels holding a Website with online booking facilities were short listed from the abstracted list. Fifty hotels were then selected from the above list stratified across two independent variables; star category of the hotel and resort region as follows (See Table 1). The chosen sample represents the whole of hotels in Sri Lanka.

**Table 1 Sample Profile**

| Star Category | Number of Hotels | |
|---|---|---|
| | Western Region | Southern Region |
| 5-Star | 6 | 2 |
| 4-Star | 4 | 2 |
| 3-Star | 5 | 3 |
| 2-Star | 6 | 4 |
| 1-Star | 12 | 6 |
| **Total** | **33** | **17** |

A convenience sample of 350 tourists from selected hotels was used in the survey during the time span of 4 months period (October 2013 – February 2014). The survey generated 309 responses, out of which 14 were unusable as the respondents had not used online booking facility of the respective hotels. Additionally, 12 responses including over 10% missing values were excluded from further analysis. The final sample analyzed consisted of 283 responses. Sample demographics are presented in Table 2.

**Table 2 Sample Characteristics**

| Characteristics | Frequency | Percentage (%) |
|---|---|---|
| Type of the Customer | | |
| International | 204 | 72.08 |
| Domestic | 79 | 27.92 |
| Gender | | |
| Male | 163 | 57.59 |
| Female | 120 | 42.41 |
| Age | | |
| Under 19 | 15 | 5.4 |
| 20 - 29 | 48 | 17.0 |
| 30 - 39 | 114 | 40.1 |
| 40 - 49 | 41 | 14.3 |
| 50 – 59 | 37 | 13.0 |
| 60 & Over | 28 | 10.2 |
| Education (highest level completed) | | |
| Primary | 14 | 4.8 |
| Secondary | 36 | 12.9 |
| University | 137 | 48.3 |
| Postgraduate | 42 | 15.0 |
| Professional/ Vocational Qualifications | 54 | 19.0 |
| Purpose of Visit | | |
| Pleasure | 156 | 55.0 |
| Business | 42 | 15.0 |
| Visiting Friends & Relatives | 23 | 8.2 |
| Conventions & Meetings | 27 | 9.5 |
| Religious & Cultural | 21 | 7.5 |
| Others | 14 | 4.8 |

Reliability was estimated by assessing the internal consistency of the scale items using Cronbach's α. The alpha values were all above 0.7 and therefore the measures were

considered reliable (Nunnally, 1978). Multiple regression analysis was used as the main statistical analysis tool in hypotheses testing.

## 4. Findings

Table 3 shows the results of the multiple regression analysis as follows.

**Table 3 Results of Multiple Regression Analysis**

| Regression Models | b | t | p | Collinearity Diagnostic | | Model Summary |
|---|---|---|---|---|---|---|
| | | | | VIF | Tolerance | |
| ***Model 1: Predicting attitude towards online purchasing*** | | | | | | |
| PEOU | -0.066 | -0.863 | Not sig. | 1.629 | 0.614 | $R^2 = 0.48$; adj. $R^2 = 0.46$; $F = 32.92$; $p < 0.001$ |
| PU | 0.554 | 7.654 | < 0.001 | 1.448 | 0.691 | |
| TRUST | 0.173 | 2.347 | < 0.05 | 1.504 | 0.665 | |
| RISK | -0.221 | -3.276 | < 0.001 | 1.257 | 0.796 | |
| ***Model 2: Predicting online purchase intent*** | | | | | | |
| PU | 0.284 | 4.295 | < 0.001 | 1.622 | 0.617 | $R^2 = .60$; adj. $R^2 = .60$; $F = 111.661$; $p < 0.001$ |
| ATTITUDE | 0.568 | 8.580 | < 0.001 | 1.622 | 0.617 | |
| ***Model 3: Predicting perceived usefulness*** | | | | | | |
| PEOU | 0.539 | 7.778 | < 0.001 | 1.000 | 1.000 | $R^2 = .54$; adj. $R^2 = .29$; $F = 60.50$; $p < .001$ |
| ***Model 4: Predicting perceived risks*** | | | | | | |
| TRUST | -0.436 | -5.900 | < 0.001 | 1.000 | 1.000 | $R^2 = .19$; adj. $R^2 = .19$; $F = 34.81$; $p < .001$ |

Model 1 was a significant predictor of attitude towards online booking intention ($F$ = 32.92; $p < 0.001$) and explained 48 % ($R2$ = 0.48) of variance in the dependent variable. The relatively moderate percentage of variance explained in Model 1 is not surprising as in given research setting, the attitude towards online booking is generally affected by a multiplicity of factors, all of which have not taken into account in this model. PEOU variable did not significantly predict the dependent variable ($\beta$ = - 0.066; $t$ = - 0.863; $p > 0.05$). The other variables entered, namely PU ($\beta$ = 0.554; $t$ = 7.654; $p < 0.001$); TRUST ($\beta$ = 0.173; $t$ = 2.347; $p < 0.05$); RISK ($\beta$ = - 0.221; $t$ = - 3.276; $p < 0.001$) showed independent predictive effects on tourists' attitude towards online booking.

Model 2 was a significant predictor of online purchase booking intention ($F$ = 111.661; $p < 0.001$) and accounted for 60 % ($R2$ = 0.60) of variance in the dependent variable. Results suggest that PU significantly predicted online booking intention ($\beta$ = 0.284; $t$ = 4.295; $p < 0.001$). Thus, there was a significant positive relationship between PU and online booking intention. Attitude also significantly predicted online booking intention ($\beta$ = 0.568; $t$ = 8.580; $p < 0.001$). Hence, there was a significant positive relationship between attitude towards Online booking Websites and online booking intention.

Results from the multiple regression analysis suggest that Model 3 was a significant predictor of perceived usefulness of online booking Websites ($F$ = 60.50; $p < 0.001$). The model explained 29 % of variance in the dependent variable ($R^2$ = 0.290). PEOU significantly predicted perceived usefulness of online booking Websites ($\beta$ = 0.539; $t$ = 7.778; $p < 0.001$).

Model 4 significantly predicts the perceived risks associated with online booking Websites ($F$ = 34.81; $p < 0.001$) and explained around 19 % of variance in the dependent variable ($R^2$ = 0.190). The results confirm the significant negative relationship between

perceived risks and trust associated with online booking Websites ($\beta$ = - 0.436; $t$ = - 5.900; $p$ < 0.001).

## 5. Discussion

H1 which proposes a direct positive relationship between PU and attitude towards online booking was supported. This indicates that more the tourists perceived the Internet to facilitate his/her hotel booking, the more positive was their attitude towards these online booking Websites. These findings are in line with the study of Chau and Hu (2001) who found PU to be a significant determinant of attitude. Carey and Day (2005) also found a strong relationship between perceived usefulness and attitude. Based on the finding, PU was the strongest predictor of attitude. This outcome yields the implication that PU factor is of more concern to the travelers.

H2, postulating a positive direct relationship between PU and online booking intention of tourists was supported. This further explains that when more tourists perceived his/her online booking useful, the more likely he/she will engage in online booking. This is echoed in the findings of Tan and Teo (2000) who applied the TAM in e- banking and found that the greater the perceived usefulness of using electronic banking services, the more likely electronic banking will be adopted.

H3 which proposes a direct positive relationship between PEOU and PU was supported implying that hotels and online travel agents should make their Websites more useful and usable. To tourists, the degree to which the online booking Website is perceived to be easy to use strongly influences their perceived usefulness of online booking. The causal relationship of PEOU to PU is corroborated by (Hubona & Blanton, 1996). Morosan and Jeong (2008) study findings also reveal that PEOU of hotel reservation Websites had a significant influence PU of the Websites. Based on the result, both constructs are concluded to be closely linked in part, this is due to the fact that a tourist would inherently try to mould his/her perception of online booking based on his/her experiences in engaging in online booking and the ease in which the task was executed.

H4 which proposes a direct positive relationship between PEOU and attitude was not supported. This finding corroborates the study of Chong, Ooi, Lin, and Tan (2010) which emphasize perceived ease of use was not significant. Although PEOU was second to PU in importance concerning consumer attitude as per extant literature, the results of this paper reveal that it was not a significant predictor of consumer attitude. This is basically due to the fact that online shopping systems have become easier to use and users have become more technically savvy, variations in the perceived ease of use dimension are reduced. Similarly, this paper was conducted with tourists those who have already had the online booking experience. Therefore, perceived ease of use was found to have no effect on their attitude.

H5 which predicts a direct relationship between trust and attitude towards online booking was supported. Previous research has consistently argued that there is a positive relationship between trust and attitude (Grazioli & Jarvenpaa, 2000; Macintosh & Lockshin, 1997). Similarly, empirical evidence has underlines the direct positive influence of trust on shopping attitude (Wu & Wang, 2005). This further explains that trust in online booking

Website enables favorable attitudes since the Website is perceived to be reliable and no harmful consequences will occur if tourists use it for online booking.

H6 which proposes a direct negative relationship between trust and perceived risk was supported implying that when more tourists perceive the Websites to be trustworthy, the lesser he/she will perceive the risk associated with online booking. The inverse relationship between trust and perceived risk has been discussed by Kim, Ferrin, and Rao (2007) who argue that trust negatively affects a consumer's perceived risk in the online transaction. However, (Morosan & Jeong, 2008) assert that consumer's perceived risk is not only related to trust. Conversely, some other scholars argue that consumer's perceived risk in the online marketplace can be reduced by trust or by raising security of Websites (Heung, 2013).

The result of this paper support H7 where perceived risk was found to have a direct negative influence on attitudes towards online booking. The risk factor associated with online booking Websites can have an impact on attitudes of tourists. The result also coincides with the studies of Heijden (2003), who states that prior research has shown that perceived risk in e-commerce has a negative effect on attitude towards the behavior. Potential risks of online booking process will affect tourists' attitude of adopting it for instance, finance loss, divulgence of personal information or lack of product/ service quality warranty.

H8 which predicts a positive direct relationship between tourists' attitude towards online booking Websites and online booking intention was accepted. This result suggests that if tourists' attitude towards the Website is positive, his/her behavior is more likely to be optimistic. This finding is in line with that of Suh and Han (2003) (2003) who empirically validate behavioral intention is determined by the individual's attitude. Previous research has also examined purchase intention and attitudes to purchasing online (Heijden, 2003). According to the result, attitude would be favorable if appropriate strategies are adopted to increase perceived usefulness and trust and reduce perceived risks.

## 6. Conclusion

This paper provides further evidence on the appropriateness of the TAM to measure the different dimensions of online booking intention of tourists. The results confirm seven out of eight proposed hypotheses among which, perceived usefulness, trusts and perceived risks have been found to be significant in influencing attitude towards online booking. On the other hand, the results suggest that perceived ease of use is not a significant predictor of attitude. Yet, perceived usefulness impacts positively on online booking intention of tourists and perceived ease of use significantly influences perceived usefulness while trusts negatively impacts on perceived risks associated with online booking Websites. This study has its own limitations as it does not cover potential tourists, who have never conducted any online booking but have the intention to engage in online booking in future. Despite these limitations, this study is the first attempt in Sri Lanka to examine factors influencing online booking intention of tourists. Therefore, these limitations should be viewed as opportunities for future research.

## References


1. Aladwani, A. M. (2002). The development of two tools for measuring the easiness and usefulness of transactional Web sites. *European Journal of Information systems, 11*(3), 223 - 234.
2. Brown, M. R., Muchira, R., & Gottlieb, U. (2007). Privacy concerns and the purchasing of travel services online. *Information Technology & Tourism, 9*, 15 - 25.
3. Bruner, G. C., & Kumar, A. (2005). Applying T.A.M. to consumer usage of handheld Internet devices. *Journal of Business Research, 58*, 553 - 558.
4. Buhalis, D., & Jun, S. H. (2011). E-Tourism. *Contemporary Tourism Reviews*, 2-38.
5. Chau, P. Y. K., & Hu, P. J. (2001). Information technology acceptance by individual professionals: A model comparison approach. *Decision Sciences, 32*(4), 699 - 719.
6. Chen, C. (2006). Identifying significant factors influencing trust in an online travel website. *Information Technology and Tourism, 8*(197-214).
7. Childers, T. L., Christopher, L. C., Peck, J., & Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behaviour. *Journal of Retailing 77*, 511 - 535.
8. Chiu, Y. B., Lin, C. P., & L.Tang, L. (2005). Gender differs: Assessing a model of online purchase intentions in e-tail service. *International Journal of Service Industry Management, 16*(5), 416 - 435.
9. Chong, A. Y., Ooi, K., Lin, B., & Tan, B. (2010). Online banking adoption: an empirical analysis. *International Journal of Bank Marketing, 28*(4), 267-287.
10. Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly 13*(3), 319 - 340.
11. Davis, F. D., Bagozzi, R., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982 - 1003.
12. Dowling, G. R., & Staelin, R. (1994). A model of perceived risk and intended risk-handling activity. *Journal of Consumer Research, 21*, 119 - 134.
13. Featherman, M. (2001). *Is perceived risk germane to technology acceptance research?* Paper presented at the AMCIS Proceedings, Boston, MA.
14. Grazioli, S., & Jarvenpaa, S. L. (2000). Perils of internet fraud: An empirical investigation of deception and trust with experienced Internet consumers. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 39*(4), 395 - 410.
15. Heijden, H. V. d. (2003). Factors influencing the usage of websites: The case of a generic portal in the Netherlands. *Information & Management, 40*(6), 541 - 549.
16. Heung, V. C. S. (2013). Barriers to implementing E-commerce in the travel industry: a practical perspective. *International Journal of Hospitality Management, 22*(1), 111-118.
17. Hubona, G. S., & Blanton, J. E. (1996). Evaluating system design features. *International Journal of Human-Computer Studies, 44*, 93-118.
18. Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (1999). Consumer trust in an Internet store. *Information Technology and Management, 1*(12), 45 - 71.
19. Khemthong, S., & Roberts, L. M. (2006). Adoption of Internet and Web Technology for Thailand. *Journal of Business Systems, Governance and Ethics, 1*(2), 47-66.

20. Kim, D. J., Ferrin, D. L., & Rao, H. R. (2007). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support System, 44*, 544 - 564.
21. Kolsaker, A., & Payne, C. (2002). Engendering trust in e-commerce: a study of gender-based concerns. *Marketing Intelligence & Planning, 20*(4), 206 - 214.
22. Macintosh, G., & Lockshin, L. S. (1997). Retail relationships and store loyalty: A multi-level perspective. *International Journal of Research in Marketing, 14*(5), 487 - 497.
23. Main, H. C. (2001). The Expansion of Technology in Small and Medium Hospitality Enterprises with a focus on Net Technology. *Journal of Information Technology and Tourism, 4*, 167-174.
24. Moon, J., & Kim, Y. (2001). Extending the TAM for a World Wide Web context. *Information and Management, 38*, 217 - 230.
25. Morosan, C., & Jeong, M. (2008). User's perceptions of two types of hotel reservation Websites. *International Journal of Hospitality Management, 27*, 284 - 292.
26. Morris, M. G., & Dillon, A. (1997). How user perceptions influence software use. *IEEE Software, 14*(4), 58 - 65.
27. Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill Publishing.
28. Pavlou, P. A. (2001). *Integrating trust in electronic commerce with the technology acceptance model: Model development and validation.* Paper presented at the Seventh Americas Conference in Information Systems, Boston, MA.
29. Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce, 7*(3), 101 - 134.
30. Pikkarainen, T., Pikkarainen, K., Karjaluoto, H., & Pahnila, S. (2004). Consumer acceptance of online banking: an extension of the technology acceptance model. *Internet Research, 14*(3), 224 - 235.
31. Suh, B., & Han, I. (2003). The impact of customer trust and perception of security, control on the acceptance of electronic commerce. *International Journal of Electronic Commerce, 7*(4), 135 - 161.
32. Tan, M., & Teo, T. S. H. (2000). Factors influencing the adoption of internet banking. *Journal of the Association for Information Systems, 1*(5), 22 - 38.
33. Taylor, S., & Todd, P. (1995). Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. *International Journal of Research in Marketing, 12*, 137 - 155.
34. Wu, J. H., & Wang, S. C. (2005). What drives mobile commerce? An empirical evaluation of the revised technology acceptance model. *Information & Management, 42*(5), 719 - 729.

# Internet User's Attitude towards an Online Advertising on Hotel Booking, and its Impact on Purchase Intention

**Mr. A.M.G.S.PRIYANKARA**
Director - Serendivus International (Pvt) Ltd, Sri Lanka
Tele: +94 112 597 772, Mob: +94 715 279 696, Fax: +94 112 597 771
Skype: sampath.abe E-mail: sampath@serendivus.com

**Abstract.** Since the explosion of the Web as a business medium, one of its primary uses has been for marketing. Soon, the Web will become a critical distribution channel for the majority of successful enterprises. Many have speculated about the current state of Internet advertising (IA) or Online advertising (OA), how it affect to the purchase intention. Although many estimates exist regarding who uses the Internet as well as guidelines about how best to design IA, little is known about Internet users' attitudes toward IA, much less what characterizes these attitudes. To test this, a sample of 150 participants with exposure to the Internet was surveyed. The results exposed greater part opinion of IA: agree to neither agree nor disagree level in significant at the 95% confidence level, A regression analysis and correlations indicated that satisfaction of looking at Internet advertisements, it's useful, and it's utility for making purchase intention.

**Key words:** Online Advertising, Internet Advertising, Online Hotel Reservation

## 1. Introduction

The rapid developments in Information and Communication Technology (ICT) have greatly contributed in enhancing human living standards worldwide a s the advanced capability of the technology facilitates with extremely efficient collaboration and access to correct, consistent and effective information, which is fundamental to manage the quality in total. In successful world, most of the key economically effective environments are increasingly ICT dominant and therefore, investing on ICT developments in the 21st century has become a compulsory requirement for a country to survive. For justification and proper direction of the investments, continuous progress monitoring, again with timely and accurate information on ICT achievements itself is also essential (Computer Literacy Survey - 2009 Department of Census and Statistics). One of the most important things in ICT is internet. The Internet as a marketing medium offers many unique challenges to marketers. To assist marketers in their business enterprise on-line, comparisons and contrasts to existing marketing theory have been used to build a conceptual understanding of the current state of the Internet and its implications for consumer transactions (cf., Hoffman and Novak, 1996a; Hoffman, Novak and Chatterjee, 1995; Schlosser and Kanfer, forthcoming). To further understand the commercial possibilities of the Internet, several internet usage surveys have been conducted to document consumers' behavior online (the most notable being GVU, 1999 and the HERMES project by Gupta, 1995; see Hoffman, Kalsbeek, and Novak, 1996, for a review). Yet, in terms of assessing the commercial effectiveness of the Internet and the value of Internet advertising, most research has determined upon the company's rather than consumers' point of view (Berthon, Pitt, and Watson, 1996). As a result, many decisions regarding Internet advertising (IA) or online advertising (OA) are being made with relatively little specific knowledge about consumers' attitudes toward OA and how the structure of these OA attitudes compare to the structure of attitudes toward advertising in traditional

media. Consumers' attitudes toward advertising have been considered important to track because they likely influence consumers' exposure, attention, and reaction to individual ads (cf. Alwitt and Prabhakar, 1992) through a variety of cognitive and affective processes (Lutz, 1985).

The company may request the consumers' attention (e.g., through banner ads on others' Web sites or through hyperlinks), but it is up to the consumer to seek additional commercial content. Consumers can select whether, when, and how much commercial content they wish to view. That is, consumers "pull" for electronic advertising content. Because OA exposure is mainly under the consumer's decision, it is particularly important to understand the structure of one important driver of advertising exposure: attitudes toward OA. Tourism is one of main developing industry in Sri Lanka; it's including both foreign and local tourist. In rapidly developing this sector, travel agents, and hotels are highly utilize OA. Because it has many benefits like cost effectiveness, time is saving, information sharing facility and etc. Tourists want to plan their tour, for that they want to book or reserve room, bungalow, hotel or any place to relax. In that purpose they are going to search information using internet. At that moment, business use to improve their revenue through OA. So in this research proposal I'm going to keep in touch "Internet user's attitude towards an online advertising on hotel booking, and its impact on purchase intention".

| Country | Population | Internet usage | Percentage of Population (%) | Percentage of world (%) | Growth rate (2000/2007) |
|---|---|---|---|---|---|
| Sri Lanka | 19,796,874 | 280,000 | 1.4% | 0.1% | 130.5% |

*(Source: Vijaya computer magazine 2007 March page 8, 9)*

Research problem identified as;

Internet user's attitude towards internet advertising on hotel booking, and its impact on purchase intention. That mean as a promotional tool internet advertising success or fail.

The objectives of the present research are address to,

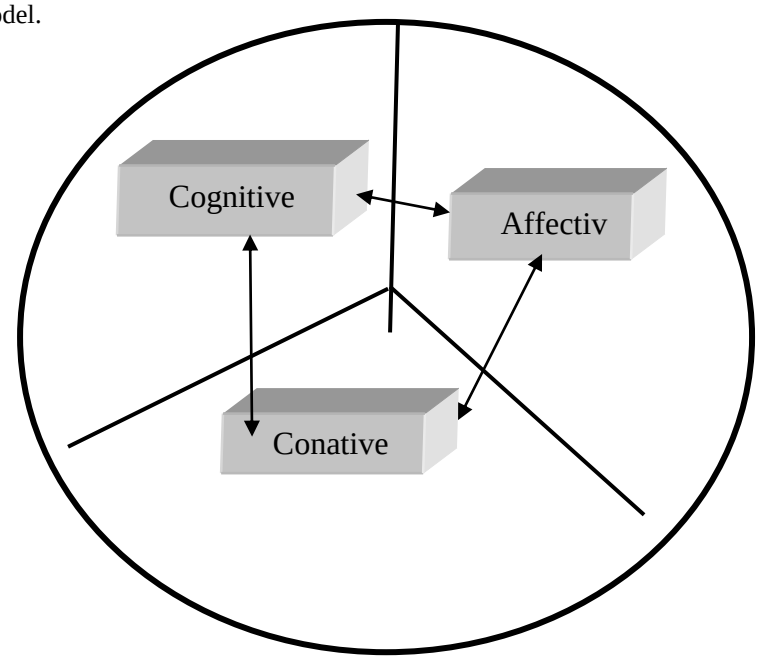
To identify consumers' attitudes toward online advertising

- To identify internet habit
- To identify how important the online advertising as information deliver

## 2.Literature Review

According to Ann E. Schlosser (1999) most of the direct-response measures administered to consumers have assessed consumers' perceptions and usage of the Internet and its services. Mehta and Sivadas (1995) assessed Internet user's attitudes toward advertising on newsgroups and through e-mail. They found that consumers held negative attitudes toward newsgroup and e-mail advertising, even when the message was directly relevant to the special interests of the group. However, their sample was limited to those who posted messages to the group. Those who merely read messages were not included in the sample. As a result, it is possible that these unfavorable attitudes are due to this verbal sample's perception that they are competing with electronic advertisements for the group's attention. Another notable exception is research conducted by Ducoffe (1996) studying the experience of consumer's attitudes toward Web advertising. It was found that a sample of 318 business executives in New York City perceived Web advertising to be generally informative and entertaining, although more informative than entertaining. Furthermore,

Mehta and Sivadas, (1995) said that they, interviewed executives found Web advertising to be useful, valuable and important. Although these previous findings on attitudes toward e-mail and Web advertising shed light upon the Internet population's attitudes toward specific types of IA, a larger and more representative sample as well as an examination of IA in general would be useful. Therefore understanding the above factors that motivate IA attitudes would also be important for both theory development and apply goals. The idea that affective (entertainment, irritation) and cognitive (informativeness) experiences with Web advertising contribute to people's judgments of Web advertising is steady with tricomponent attitude model.




**Cognitive component**

Tricomponents attitude model consists of a person's cognitions, that is the knowledge and perceptions that are acquired by a combinations of direct experience with the attitude object and related information from various source. This knowledge and resulting perceptions commonly take the form of beliefs. That the attitude object possesses various attributes and that specific behavior will lead to specific outcomes.

**Knowledge**

According to Alba and Hutchinson (1999), Consumers vary in their knowledge about products and brands, depending on their personal experiences. For example, an individual might be highly knowledgeable about computers but not dishwashers, or about pet grooming but not broker services. These differences in knowledge have important implications for consumers' pre-purchase information search activities.

**Perception**

Perception is the process by which people select, organize, and interpret information to form a meaningful picture of the world.

**Affective component**

A consumer's emotions or feelings about a particular product or brand constitute the affective component of an attitude. These emotions and feelings are frequently treated by

consumer researchers as primary evaluative nature; that is, they capture an individual's direct or global assessment of the attitude object. The emotional status may enhance or amplify positive or negative experience and that later recollection of such experience may impact what comes to mind and how individuals acts.

**Conative component**

This concerned with the like hood or tendency that an individual will undertake a specific action or behave in a particular way with regard to the attitude object. According to some interpretations, the conative component may include the actual behavior itself. This is frequently treated as an expansion of consumer's intention to purchase.

**Intention to purchase**

According to Authors Carole, attitude is described as the psychological tendency of a person to respond, or behave, in a consistently positive or negative manner with respect to a stimulus as a result of their attitude toward the stimulus.

**Internet Advertising**

IA is described broadly as any form of commercial content available on the Internet that is designed by businesses to inform consumers about a product or service. Hence, IA can be delivered via any channel (e.g., video clip, print or audio), in any form (e.g., an e-mail message or an interactive game), and provide information at any degree of depth (e.g. a corporate logo or an official Web site). According to researchers, Studies of consumers' reactions to IA typically have quantified customers' judgments of Web sites in terms of consumers' behavioral traces at the site (i.e. counting the number of "clicks" and "hits" (Berthon, Pitt, and Watson, 1996). These measures have been shown to both overestimate and underestimate the number of visitors and exposures (Riphagen and Kanfer, 1997).

Another important exception is research conducted by Ducoffe (1996) studying the antecedents of consumer's attitudes toward Web advertising. It was found that a sample of 318 business executives in New York City perceived Web advertising to be generally informative and entertaining, although more informative than entertaining. This is consistent with findings regarding people's perceptions of the Web in general (Diaz et al., 1996). Furthermore, in contrast to attitudinal findings toward e-mail advertising (Mehta and Sivadas, 1995), the interviewed executives found Web advertising to be useful, valuable and important. Although these previous findings on attitudes toward e-mail and Web advertising shed light upon the Internet population's attitudes toward specific types of IA, a larger and more representative sample as well as an examination of IA in general would be useful.

**Abbreviations & Definitions**

HTML - Hyper Text Mark-up Language
IA – Internet Advertising
ISP - Internet Service Provider
Link(s) - Hyperlink (Method of navigating to content)
Click-through rate - The number of hyperlinks accessed from a campaign
Image link - Hyperlinks accessed from a campaign.
Pay-per click (PPC) – Per click ads has to pay
SEO – Search Engine Optimization

## 3.Methodology

Most of the Sri Lankan are very close to the globalize environment and market. As a result of internet usage, the environmental process also highly improved. In current situation Sri Lanka has rapidly growth in the internet literature, during 2000- 2007. Also recent

research found out that 13 percent of the population aged 5 to 69 years has used the Internet facility at least once during the last twelve months. This chapter gives a clear picture on the methodology used in the study specifically the sample selection and data collection.

**Independent variable**

The dependent variable of this study was cognitive factors and affective factors affect purchase intention on online hotel booking. Twenty questions measuring internet user's cognitive factors and affective factors affect purchase intention on online hotel booking were included in the survey.

**Dependent variables**

Independent variables of this study is conative factor (purchase intention)

**Types of the study**

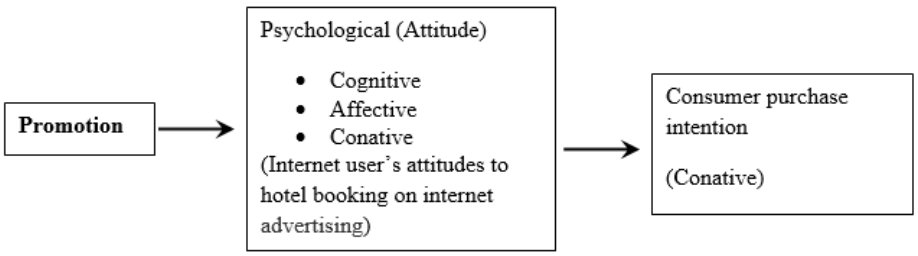
This study is a descriptive study. To using this descriptive study Tricomponent attitude model is analysis according to objectives. Major objectives of the descriptive research is to describe something usually market characteristics or functions. So my research is describe cognitive (knowledge, perception), affective (feelings), conative (purchase intention).

**Theoretical framework**

Based on the proceeding literature and research question, the conceptual model for this study is shown follows. It was conducted based on casual relationship of some variables. Initially it has described the relationship between Internet user's attitudes to regarding hotel booking through internet advertising and their purchase intention.

**Conceptual frame work**

| S | **BBB** | **R** |
|---|---|---|
| Marketing stimuli | Buyer characteristics | Buyer decision |




(Source: Page 67 MAR2301 cause manual and CB shiffman &kanuk)

**Approach**

The main approach is online survey research method for data collection.

**Time horizon**

A cross sectional research design used to collect information from respondents. A cross-sectional or one shot study has followed to gather the data from sample of internet users who are dealing with internet. The data was gathering through a 5 days from them.

**Population and sampling**

Population is internet usage customers.

Total Sample size is 150. Random sampling method was used to collect required data. Respondents are meeting from e mail campaign that design from Google form / free online survey.

**Data collection**

- Primary Data

The major part of the study is primary data. Primary data is collect by using an online market survey for internet and email users.

- Secondary Data

Number of related research articles, books, newspapers, trade journals industry portals, government agencies, trade associations and the information related to the topic.

**Data analysis method**

The data will be analyzed by using the SPSS Statistical package as well as Microsoft Excel. Methods used for analyzing this, Likert Scale, and data will be analysis through mean, standard deviation, ANOVA test, correlation and T-test.

## 4.Data Analysis and Findings

**Analysis of the Sample**

Selected sample consists with 150 respondents and questions consist with cognative, affective and conative variables. Hence firstly this chapter was attempting to identify the reliability of the respondents by scale test. The sample reliability was 0.877 (Table 01 appendixes).

Model was accept, because 0.05 < 0.6, model relationship;

$$\mathbf{Y} = \hat{\beta}_0 + \hat{\beta}_1$$

**Table 02 Model Summary**

Model Summary

| Model | R | R Square |
|---|---|---|
| 1 | .774[a] | .600 |

a. Predictors: (Constant), Purchase Intention

**Y=** Internet advertising that regarding hotel room / bungalow booking is creating purchase intention

$\hat{\beta}_0 = 0.05$

$\hat{\beta}_1$ = Cognative and Affective question including question number one to twenty.

Model regression presented 0.774 so it said strong positive relationship.

**Table 03 Anova**

ANOVA[b]

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 46.948 | 1 | 46.948 | 221.812 | .000[a] |
| | Residual | 31.325 | 148 | .212 | | |
| | Total | 78.273 | 149 | | | |

a. Predictors: (Cons
Purchase Intention

b. Dependent Variable: Internet advertising that regarding hotel room / bungalow booking is creating purchase intention

**Table 04 Coefficients**

Coefficients[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -.381 | .187 | | -2.032 | .044 |
| | Purchase-Intention | 1.234 | .083 | .774 | 14.893 | .000 |

a. Dependent Variable: Internet advertising that regarding hotel room / bungalow booking is creating purchase intention

In above table researcher had identified cognative and affective components as a significant factor of purchase intention. In following table indicate relationship between cognative and affective with regard to Purchase-Intention. According to that there is significant relationship between purchase intention and cognative and affective factors (Table 04, 06, 07, 08 appendixes).

As same as the cognative part respondents attitude towards online advertising were significant at the 95% confidence level. So there can be identified some difference between cognative and purchase intention, 0.041<0.05 (Table 06 appendixes). As the affective part respondent's attitude towards online advertising were significant at the 95% confidence level, (Table 08 appendixes).

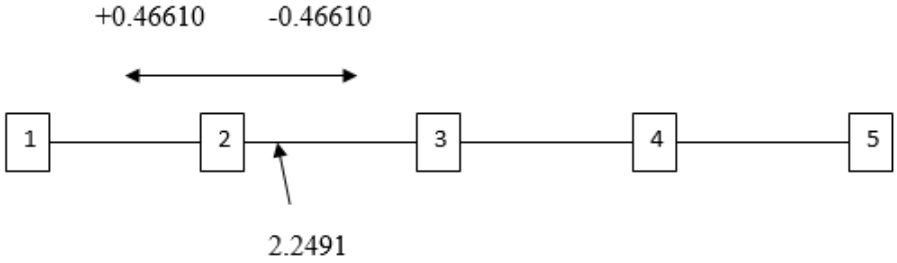
**Descriptive analysis**

**Table 09 Descriptive analysis**

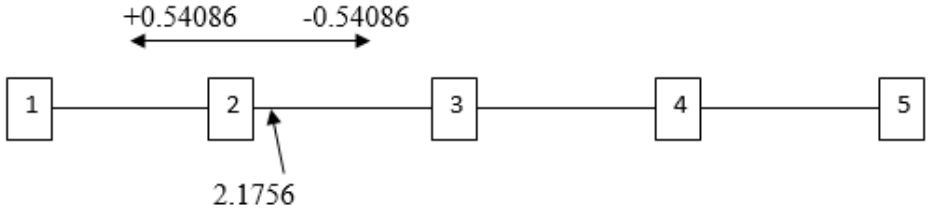
Descriptive Statistics

| | N | Min | Max | Mean | Std. Deviation | Variance |
|---|---|---|---|---|---|---|
| cognitive | 150 | 1.64 | 3.73 | 2.2491 | .46610 | .217 |
| affective | 150 | 1.44 | 4.00 | 2.1756 | .54086 | .293 |
| Purchase Intention | 150 | 1.60 | 3.70 | 2.2160 | .45496 | .207 |
| Valid N (list wise) | 150 | | | | | |

According to the above table overall mean value of the cognative is 2.2491. It can be varying from 0.46610 (2.2491 ± .46610). According to the decision criteria this mean value close to agree to neither agree nor disagree fallen level of attitudes.




According to the above table overall mean value of the affective is 2.1756. It can be varying from 0.54086 (2.1756 ± .54086). According to the decision criteria this mean value close to agree to neither agree nor disagree fallen level of attitudes.




According to the above table overall mean value of the conative (purchase intention) is 2.2160. It can be varying from 0.45496 (2.2160 ± .45496). According to the decision criteria this mean value close to agree to neither agree nor disagree fallen level of attitudes. Higher level of attitudes towards online advertising is significant at the 0.05 significance level.

**Correlations**

**Table 10 correlation of cognitive and affective to purchase intention**

| | | | Internet advertising that regarding hotel room / bungalow booking is creating purchase intention | Purchase Intention | cognitive | affective |
|---|---|---|---|---|---|---|
| Spearman's rho | Internet advertising that regarding hotel room / bungalow booking is creating purchase intention | Correlation Coefficient | 1.000 | .731** | .537** | .693** |
| | | Sig. (2-tailed) | . | .000 | .000 | .000 |
| | | N | 150 | 150 | 150 | 150 |
| | Purchase Intention | Correlation Coefficient | .731** | 1.000 | .794** | .836** |
| | | Sig. (2-tailed) | .000 | . | .000 | .000 |
| | | N | 150 | 150 | 150 | 150 |
| | cognitive | Correlation Coefficient | .537** | .794** | 1.000 | .379** |
| | | Sig. (2-tailed) | .000 | .000 | . | .000 |
| | | N | 150 | 150 | 150 | 150 |
| | affective | Correlation Coefficient | .693** | .836** | .379** | 1.000 |
| | | Sig. (2-tailed) | .000 | .000 | .000 | . |
| | | N | 150 | 150 | 150 | 150 |

**. Correlation is significant at the 0.01 level (2-tailed).

Moderate positive correlation (0.537) between cognative and purchase intention that, question number 21 said; Internet advertising that regarding hotel room / bungalow booking is creating purchase intention. Also moderate positive correlation (0.693) between affective and purchase intention

## 5.Conclusion & Recommendation

**Conclusion**

This chapter is present Final Conclusion and Recommendation of the data from the Survey. This chapter consists with two sections and in first section is reserved to present finding of the study. Second part was given some recommendation relating to the finding of the study. According to the findings of the research following conclusions can be done.

Generally Sri Lankan internet users were developing an internet purchasing behavior. Cognative and affective factors affect the online advertising on hotel booking which represent were moderate positive relation. That mean those factors could be impact on purchase intention on online hotel booking. Sri Lanka is rapidly developing country; so many internet users were around 60% of respondents said online advertising was benefit to hotel booking. In addition to the fact that all the scales used in the study showed high reliability, those of shopping enjoyment, perceived usefulness, involvement, challenges, and skills also affect the online booking. From a substantive point of view, examined how emotional and cognitive responses to the Web advertising could be influence online consumer behavior.

**Recommendations**

- To use Web site investment effectively, however, managers should identify the reason most consumers visit their site. Above results indicate that web site investment is effective in boosting online purchase intentions when visitors are searchers. Also online marketers should establish more invest in web site design to establish trust online.
- Consumers appear to gather that a firm with a well-designed site can be trusted privacy / security statement would have if its exposure was allowed to vary. Perhaps the first hurdle is to establish trust in the firm's ability.
- Sometimes customers have giant requests to book hotel online, but they have some extent to fear to do it, because relevant details may not be provide by advertisement. So when hotel or bungalow advertises online the campaign should be improve the relevancy that they can gain more advantage through online advertising. As example; when we advertise facebook.com they give us to respondents summary sheet. That sheet can be present our web site as relevancy copy.
- This type of customer and market is opportunity to the companies. Multimedia format is much considers respondents. Because multimedia advertisement can be give more details regarding products. Ex: shape, color, etc. It helps to express their feelings regarding product at that moments. Ex; video massage, blink ads, slide show etc.
- Online advertising always targeted the audience and this helps in making campaign more profitable and getting more relevant leads.
- Blogs are the latest internet advertising trend. It's too soon to tell just how much money and business can be generated from a blog but early reports indicate that blogs are becoming an important part of online advertising.
- Link exchanges still work. At one time, link exchanges were a big deal. People have since moved on to other methods but link exchanges can still help improve rankings if done in appropriate quantities with relevant sites. If companies have little on cash, find a link exchange program or approach some webmasters and ask to exchange links.

- Create affiliate advertisement programme and answering machines of the internet also sponsor a chat room.

**Limitation**

The main limitation of the study is that the use of an online survey may result in some problems regarding the generalizing of the results. It is possible that certain types of respondents may be more likely to participate in an online survey. This limitation is common to many online surveys.

## References

1. Ann E. Schlosser,Sharon Shavitt,Alaina Kanfer (1999).Survey of internet users' Attitudes toward internet Advertising. Journal of interactive marketing c volume 13 / number 3 / summer 1999.
2. Ann E. Schlosser, Tiffany Barnett White, & Susan M. Lloyd (April 2006). Converting Web Site Visitors into Buyers: How Web Site Investment Increases Consumer Trusting Beliefs and Online Purchase Intentions. Journal of Marketing Vol. 70 (April 2006), 133–148.
3. Arvenpaa,SirkkaL.,PeterA.Todd (1997a).Is there a future for tailing on the Internet? R.A.Peterson,ed.ElectronicMarketing andtheConsumer.Sage,ThousandOaks,CA.
4. Berthon, P., Pitt, L. F., & Watson, R. T. (1996). The World Wide Web as an Advertising Medium: Toward an Understanding of Conversion Efficiency. Journal of Advertising Research, 43–54.
5. Briggs, R. & Hollis, N (1997). Advertising on the Web: Is There Response Before Click-Through? Journal of Advertising Research, 37, 33–45.
6. Computer Literacy Survey (2009) http://www.statistics.gov.lk. Department of Census and Statistics. Sri Lanka ISSN 2012-6565 December – 2009.
7. Corritore, Cynthia L., Beverly Kracher, and Susan Wiedenbeck (2003), "Online Trust: Concepts, Evolving Themes, A Model," International Journal of Human-Computer Studies, 58 (June), 737–58.
8. Cyberatlas (1998). Sharp Rise in Online Ad Spending: Computer Industry Leads Way. http://cyberatlas.internet.com/segments/advertising/ad_spending.html
9. DECK, S. (1998) , Study sees growth in online shopping . Computer world,. Available online: http :/ /www.computer world. com / home/ online9 69 7 .ns f/ all /9 80 5 21 study IF9 56
10. Hoffman, D. L. & Novak, T. P. (1996a). A New Marketing Paradigm for Electronic Commerce. The Information Society, 13, 43–54.
11. INTERNET ADVERTISING BUREAU, " Report on Online Advertising Revenue," (April, 1998). I/PRO CyberAtlas, 1996.
12. Jupiter Communications (1997). Jupiter Model Defines Online Ad/Direct Marketing Convergence [online]. http://www.jup.com/jupiter/release/aug97/ online.shtml.
13. Janiszewski, Chris (1998), "The Influence of Display Characteristics on Visual Exploratory Searching Behavior," Journal of Consumer Research, 25 (December), 290–312.
14. Layton, J. & Kanfer, A. (1996). Summary of Internet Surveys, [on-line]. http://www.ncsa.uiuc.edu/edu/trg/survey/.

# STD Implementation in Automated Learning Systems through Dialog Modeling

**Yurii Furtat**[1]

[1] G.E.Pukhov's IMPPE NAS of Ukraine
15, General Naumov Str., Kiev, 03164, Ukraine
saodhar@ukr.net

**Abstract.** During the learning process, students often encounter difficulties when dealing with the test text material. These difficulties arise in cases where the student works with text, the complexity of which is either beyond his readers' skills, or outside its conceptual basis. The main reason for these difficulties is often limited or inadequate support necessary for understanding, manipulating, studying and applying the knowledge gained from the text material to real-world problems.
It is necessary to provide support to the student when working with educational materials (texts). One of the promising approaches to providing such support is the creation of intelligent interaction with the text, which is based on the concept of "supported text document" (STD) and the theory of question-answering dialog.
STD is a specially adapted form of hypertext, designed to provide students in the process of using the semantics of the studied text or multimedia material.
Using STD with integrated intelligent component will make it easier for the the student to work with electronic teaching materials due regard to its cognitive preferences and facilitate access to the auxiliary and reference materials.
To effectively integrate STD into automated learning systems though, a dialog-modeling system has to be considered, developed and implemented.

**Keywords.** learning process, automated learning system, supported text document, dialog modeling.

## 1. Introduction.

During the training process students often face difficulties while working with studied text material. The difficulties appear if student are working with text that is too complex for the reader's skills, or lies outside of his conceptual basis. One of the main reasons for these difficulties is limited or inadequate support needed to understanding, manipulating, studying and applications of knowledge gained from text material to real world tasks.

Is is necessary to provide support to students while working with training materials (texts). One of the perspective approaches to provide such support is creating intelligent interaction with the text, based on the concept of "resource-extended text "and the theory of question-answer interaction.

Supported Text Document (STD) is a specifically adapted Hypertext developed to provide assistance to students in process of forming studied text or multimedia material semantics.

STD consists of three main components: source text, set of keys and set of resources.

Source text - is the text document with which student is working and which one it tries to understand and associate with earlier formed cognitive schemes.

The keys represent the parts of source text that can cause for students the difficulties in perception and understanding in process of semantics forming. Keys are marked in text with key markers which, when activated, provide students with access to resource packages. Trained students can work with source text, almost without using key markers. For such students resources are "folded" and invisible and do not violate natural material perception. However, if student, teacher, or built-in intellectual component needs a resource, key marker activates and provides access to specific resource.

The keys are divided into two categories. Syntax keys are source text physical components: technical or context dictionary, complex or confusing sentemces or paragraphs, characters, titles, links, etc. Since these are the keys are part of the source text, they normally play key markers role.

Semantic keys cannot be marked directly using source text elements as they are not physical text elements themselves. They relate to cognitive processes with understanding of the text and to possible need in additional knowledge.

STD resource categorization and description are available in [1] and [2].

## 2. Built-in intellectual component objectives and tasks.

Experience of STD applications to different groups of students allows formulating two top reasons for necessity of intellectual component embedding into its structure.

First reason - wrong use of STD resources. Students either forget that STD is an interactive system, and ignore resources, or focus attention exclusively on resources and ignore source text. These problems is usually resolved by the teacher's intervention, however, they can also be resolved by utilizing STD built- in intellectual component.

Second reason - the need to personalize resources because of the differences between primary knowledge and the student's cognitive abilities. Preference files can be used for it, formed either by students or by teacher. However, because a student can rarely clearly formulate their cognitive preference it would be helpful to conduct automatic diagnostics and create preferences file with assistance of intellectual component.

Considering these reasons one can formulate two top tasks in front of intellectual component embedded in STD:

- compensation of the students' cognitive skills lacking while working with extensive STD resources and navigating them;
- adaptation STD components to students' personal cognitive characteristics.

Thus, the STD built-in intellectual component should help solve problem of personalized training environment creation for a student with in advance unknown cognitive abilities.

## 3. STD personalization

STD can be personalized by its adaptation to the student's current cognitive abilities from a point of view of a technological paradigm that researchers from Massachusetts Technology Institute Media Lab called "thinking things" [3].

"Thinking things" behavior examples analysis [3], allows to list distinctive characteristics of such systems:

- "thinking things "are in dialog with their users;
- "thinking things"activity can be defined as dialog behavior;

- "thinking things"dialog behavior is purposeful and adaptive.

In this manner we can define STD with built-in intellectual component as a subclass of built-in intelligence systems that differs from simple STD by following distinctive characteristics:

- intellectual STD is in constant communication with students;
- intellectual STD acts semi-autonomous, and its activities can be qualified as behavior;
- intellectual STD behavior is purposeful and adaptive.

Intellectual STD implements its behavior by generating a sequence of SmartPages, which it provides to student. SmartPage description reflects cognitive characteristics of trainee and is generated by STD built-in conversational component in accordance with information accumulated in it during the dialog with students.

## 4. Specifics of dialog with the student.

Natural dialog agents' roles distribution in case of STD with built-in intellectual component - is reactive dialog agent role implementation by students and active dialog agent role implementation by STD. The student logically depends from STD. Student cannot give a free-form answer, and has to provide STD with relevant fragment of declarative knowledge (Anderson [4] called this declarative Knowledge fragment «chunk»). The student cannot provide aa arbitrary response, because in this case dialog process logic breaks, and the dialog becomes two independent monologs.

Thus, in one step STD and student transmit each other interconnected fragments of declarative knowledge. SmartPage that STD transfers to the student has logical structure of a question, and knowledge fragment transmitted from student to STD - logical structure of a response.

Dialog behavior description in form of a sequence of interconnected dialog steps is stored in interactive knowledge database (DiKB - Dialogue Knowledge Base). The main DiKB elements are questions memory (in the considered case it is SmartPages memory) and questions memory access method – DiAM.

Dialog behavior "materialization" can be implemented with assistance of a dialog-modeling system.

## 5. Dialog-modeling system.

Let's consider and option of such a system implementation in a software agents commands form.

An agent is defined as any part of consciousness, which structure and behavior are simple and clear enough, while interaction of a "team" of such agents can produce a reasonable behavior phenomenon, much more complicated to understand [5].

Arbitrary agent shows on each dialog step specific behavior that is determined by its internal structure and *internal and external dialog-modeling system world* states. The system's external world is defined as the student's knowledge model and internal world – an STD with built-in intellectual component knowledge model (dialog scenario) stored mainly in DiKB.

The proposed concept of dialog-modeling system is developed with purpose to process a random step, and each agent's behavior is determined by dialog process logic, and not by logic of the subject area. All subject area features are stored in the model's internal

world (dialog scenarios). Agents act asynchronously, sequentially and cyclically transmit activity one to another [6].

Combination of knowledge exchange process and deductive process is described in dialog script and reflected in the DiKB structure.

Fig. 1 shows the structure of the system represented by four agents located in cyclic order that determines sequence of their activation.

All of the dialog-modeling system internal world components are located within the cycle.

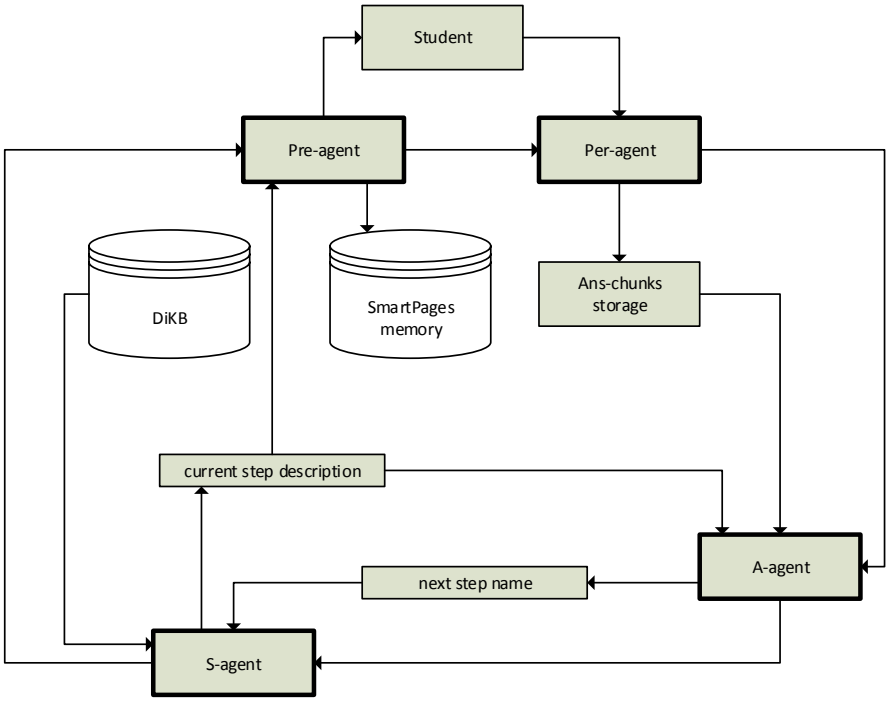



Fig. 1. Four agent dialog-modeling system architecture.

Presentation agent delivers SmartPage to the student. For example, SmartPage can relate to a question designed to receiving facts with purpose to make conclusions about source document text fragment understanding degrees by students. The interaction with external world is one-directional, from Pre-agent to the student. Interaction with internal is done in such a way: from a memory block that stores current step description, Pre - agent reads the SmartPage name, and then - all information required to deliver this SmartPage.

Perceptional agent receives Ans-chunks from student and accumulates them in Ans-chunks storage. So, the Ans-chunks storage contains the history of the student's answers. Interaction with external world is one-directional, from student to Per-agent. Reactive dialog agent's answers are received and encoded using standard peripheral computer equipment. Interactions with internal world are performed so: Ans-chunks are accumulated in storage.

Analyzing agent recognizes received Ans-chunk by its alternative comparison to answer sets elements, defines the next step name. A-agent does not interact with external world. Internal world interaction: A-agent reads Ans-chunk from the Ans-chunks storage and answers sets from memory block that sores current step description, then determines next step name and places it into the next step's name memory.

Step agent updates current step description. The S-agent does not interact with external world. Internal world interaction: S-agent reads the next step name from a corresponding memory block and then this step's descriptions p from DiKB. Read descriptions are placed into current step descriptions memory.

## 6. Conclusion.

Using STD with built-in intellectual component allows to facilitate student work with electronic training materials through accounting his cognitive preferences and simplifying access to auxiliary and reference materials.

The considered dialog-modeling system is a concept practical sale such intellectual component and his interaction with learners (students).

## References

1. Chemyr I. A., Horney M. A., Anderson-Inman L. SmartText: Using Agents Supporting Personalized Reading Comprehension. - Personal Technologies, Vol. 2, no. 3, 1998. - pp. 152-161.
2. Anderson-Inman L., Horney M. A. Transforming text for at-risk readers. In D. Reinking, L. Labbo, M. McKenna & R. Kieffer (Eds.). - Handbook of literacy and technology: Transformations in a post-typographic world - Mahwah, NJ: Lawrence ErlbaumAssociates, 1998. - pp. 15-43.
3. Michael Hawley R., Dunbar Poor and Manish Tuteja. Things That Think. - Personal Technologies, vol. 1 no. 1, 1997. - pp. 13-20.
4. Anderson, J. R. The Architecture of Cognition. - Harvard University Press, Cambridge, Massachusetts, in 1983.
5. Minsky M. The Society of Mind. - Published by Simon and Schuster, 1988.
6. Верлань А.Ф., Чмырь И.А. Системы со встроенным интеллектом на базе архитектуры машины диалога. – Электронное моделирование, т. 23, № 1, 2001. – с. 75-83.

# Banks Bankruptcy Probability Forecasting on the Bases of Survival Models

**Olena Sergienko[1], Nadiya Morozova[2]**

[1]Kharkiv institute of banking of the University of banking of the NBU, Kharkiv, Ukraine
ser_helen@mail.ru
[2]Kharkiv institute of banking of the University of banking of the NBU, Kharkiv, Ukraine
morozova.nadiya@khids.edu.ua

**Abstract.** The model of survival analysis of commercial banks by the Kaplan-Meier method which makes possible to determine the hierarchy of importance of factors that influence on commercial banks survival as well as to make comparative analysis between the designated groups in terms of financial stability and to find differences within groups is suggested in this paper. Using the proposed model in the management system makes possible to predict future situations for development and implementation of preventive measures and to confront bankruptcy.

**Keywords.** Bankruptcy, commercial bank, financial stability, the Kaplan-Meier method, survival model

## 1. Introduction

Modern development of the global banking system occurs in conditions of increased competition and crisis in financial markets. The change of social and economic policy course in Ukraine, rapid economic development, European integration intentions and actual bankruptcy of a large number of commercial banks poses new challenges to banking system such as competitiveness improvement of Ukrainian banks, regulatory and competitive environment improvement, and search for effective mechanisms for improving banking management and for increasing survival. Because of the severity of bank insolvency problems, studies of the causes and analysis of consequences are examined by a lot of companies, government agencies and research organizations.

Banks activity indicators (profits, capital profitability, banks cash, etc.) also prove the difficult situation of Ukraine banking system. Today each bank must not only adapt to rapidly changing environmental conditions, to counter the negative factors impact, but also to find new ways for survival. It confirms that the number of bankrupt banks and the banks that have moved to a temporary administration increases. According to the NBU, during the last 10 years 67 banks were removed from the state registry and the liquidation procedure was completed (10 banks were liquidated as a result of reorganization). In the middle of May 2013 23 banks were in the process of liquidation because of financial indicators deterioration. For two of them the decision was made in 2012; for one of them in 2013; 3 banks have been liquidated since 2009; 9 banks since 2010, and only one since 2000 ("Gradobank") [1, 2]. In 2013 and early 2014 Bank Tavrika, Erde Bank, Bank Daniel were in liquidation process, Odessa Bank, Bank Armagh, East European Bank, Land Bank, Ukrainian Industrial Bank and others became bankrupt [8, 9], which proves the need for

methods of banks survival that will anticipate and counteract bankruptcy. So today the vital question is the construction of mathematical and economic models for research and identification of key factors survival of Ukraine's banks.

**Analysis of recent research and publications.** The degree of scientific development of these problems is very low. V. Malyugin and E. Pytlyak [7] studied the issue of bank bankruptcy model development. O. Bystritskaya and O. Kramarenko also study the problem of distance supervision as an essential component of banking supervision. Such authors as A. Bayko, A. Brishtelev [**Error! Reference source not found.**] A. Haliv and P. Kallaur considering the banking sector stability factors describe the steps required for its increase, and examine the role of the National Bank for provision of the state financial stability.

Researches of such scientists as A. Karminsky, A. Peresetskiy and S. Golovan are devoted to modeling banks defaults by using of binary choice models [5]. It should be noted that the most developed methods of assessment and analysis of bankruptcy probability are focused on enterprise bankruptcy, and only a small part is focused on lending institutions. Among the popular methods of financial monitoring technique is ratio analysis, ratings or rating systems, etc. [6]. The existing methods and approaches to modeling, forecasting of commercial banks bankruptcy cannot always predict the crisis situation, and even more so, to determine the factors for its overcoming and for commercial bank survival in these conditions. Therefore, the application of survival analysis methods is very important for commercial banks.

**Description of the method.** Methods of survival analysis are mainly applied to the same statistical problems as other methods, but their peculiarities reside in the fact that they are applied to the censorial or incomplete data [3, 5]. It should be noted that more often than normal distribution function the so-called survival function is used in these methods, that is the probability for an object to live longer than time $t$ [**Error! Reference source not found.**].

Thus, complete and incomplete or censored observations are distinguished in the survival analysis. This approach was originally developed for medical and biological research and insurance, but then these methods have been widely used in social and economic sciences and in engineering problems (reliability analysis and failures time).

The most common way of describing survival in the selection is to build the lifetimes tables. The area of possible time of critical events occurrence (deaths, failures, bankruptcy, etc.) is divided into a number of intervals. For each interval, we calculated the number and part of objects that at the beginning of the reporting interval were "alive", the number and part of objects that were "dead" in this interval and the number and part of objects that were deleted or censored in each interval.

Some additional statistics are calculated on the bases of such values as [**Error! Reference source not found.**]:

- the number of studied objects is the number of objects that were "alive" at the beginning of the considered time interval minus half a number of deleted or censorial objects.

- the part of objects that are "dead" is the ratio of the number of objects that are dead in the corresponding interval to the objects that are studied in this interval.

- the part of those survived is equal to one minus the part of objects that are "dead".

The cumulative part of those survived (survival function) is a cumulative part of survivors before the relevant time period. Since the probabilities of survival are considered

independent at different intervals, this part is the multiplication of the part of objects that survived at all previous intervals. The resulting part as a function of time is called as the ability to survive or survival function

Density of probability is the probability of failure in the relevant interval, which is determined as:

$$F_i = \frac{P_i - (P_i + 1)}{h_i}, \quad (1)$$

where $F_i$ - the evaluation of the failure probability in the *i*-th interval, $P_i$ - cumulative part of objects that survived (survival function) before the *i*-th interval, $h_i$ - the width of the *i*-th interval.

The intensity function is defined as the probability that an object that has survived till the beginning of the corresponding interval fails or dies during this interval. The evaluation of intensity function is calculated as the number of failures per time unit corresponding to the interval divided by the average number of objects that survived till the time moment in the middle of interval.

The median of expected lifetime is a point on the time axis in which the cumulative survival function is equal to 0.5.

It is recommended to use at least 30 observations on order to get reliable estimates of three main functions (survival function, density of probability function and intensity function) and their standard errors at each time interval.

For censorial, but not grouped observations of lifetimes, the survival function can be evaluated directly (without table of lifetimes) by using Kaplan-Meier estimates. A file in which each observation contains exactly one time slot must be created. Multiplying probability of survival in each interval, we obtain the following formula for the survival function [3]:

$$S(t) = 1\frac{n - j}{n - j + 1}, \quad (2)$$

where *S(t)* is an estimation of survival function, n is a total number of events (end time), j is a serial number of a particular event; d (*j*) = 1 if the *j*-th event means failure (bankruptcy) and d (*j*) = 0 if j-th event means the loss of observation (censored), *n* means multiplication of all observations *j*, which ended by the time *t*.

Kaplan and Meyer (1958) first proposed the estimation of survival function, which is called, duplicating assessment.

The advantage of Kaplan-Meier method (compared with life table method) is that the estimates are independent from partitioning the observation time into intervals, i.e. from the grouping. The method of duplicating assessments and method of tables lifetimes lead to the same results if the time intervals contain a maximum of one observation each [3].

In survival analysis lifetimes can also be compared. In principle, since lifetimes are not normally distributed, nonparametric tests on ranks can be used. Non-parametric statistics offer a wide range of nonparametric criteria that could be used to compare the lifetime. However, these criteria do not "work" with censored data.

In survival analysis there are five different (mostly non-parametric) criteria for censorial data: Wilcoxon test generalized by Hehan, Cox F-test, the log rank test and Wilcoxon test generalized by Peto (R. Peto and J. Peto). Most of these criteria have corresponding z-values (the standard normal distribution); these z-values can be used for statistical test of any differences between the groups. However, the criteria give reliable results only at sufficiently large volumes of samples. For small volumes of samples, their "behavior" is less understandable [**Error! Reference source not found.**].

The most common regression model for analysis of survival rate is Cox proportional intensity model because it is not associated with any assumptions about the distribution of survival time. This model assumes that the function has a level of intensity that is a function of the independent variables. No assumptions about the form of the intensity function are made. Therefore, the Cox model can be considered in some sense as a non-parametric. The model can be written as follows:

$$h\{(t),(z_1,z_2,...,z_m)\} = h_0(t) \times \exp(b_1 z_1 + ... + b_m z_m), \quad (3)$$

where $h\{(t)...\}$ is the resulting intensity at the corresponding observation values *m* covariant $(z_1, z_2,..., z_m)$ and the corresponding lifetime (t);

$h_0(t)$ is a basic intensity function equal to the intensity when all independent variables are zero.

The exponential regression model assumes that the distribution of life is exponential and is associated with the values of independent variables set ($z_i$). Intensity parameter exponential distribution is expressed as:

$$S(z) = \exp(a + b_1 z_1 + b_2 z_2 ... + b_m z_m), \quad (4)$$

where *S(z)* is a lifetime, *a* is a constant, $b_i$ is the regression parameters.

The value of xi-square can be calculated as a function of the likelihood logarithm for the model with all evaluated parameters ($L_1$) and the logarithm of the likelihood model where all covariates are converted to 0 ($L_0$). If the value of xi-square is statistically significant, reject the null hypothesis and accept that the independent variables gravely affect the lifetime [**Error! Reference source not found.**].

Let us consider the normal and lognormal regression type. This model assumes that the lifetimes (or their logarithms) follow a normal distribution. The model is generally identical to the usual multiple regression model and can be described as follows:

$$t = a + b_1 z_1 + b_2 z_2 ... + b_m z_m, \quad (5)$$

where *t* is a time of life. If it is lognormal regression model, it is replaced by *t ln t*.

Normal regression model is particularly useful because the data often can be converted to normal by using normalizing approximations. Thus, in a sense, this is the most general parametric model (as opposed to the model of Cox proportional to the intensity, which is nonparametric). Its estimates can be obtained for a large variety of initial distributions of lifetimes.

The value of xi-square can be calculated as a function of the likelihood logarithm for the model with all independent variables ($L_1$) of likelihood logarithm for the model where all independent variables are replaced by 0 ($L_0$) [3].

Thus, the construction of tables lifetime, survival distribution fitting, evaluation of survival function by using the Kaplan-Meier procedure is a descriptive methods for censorial data research. Some of the proposed methods make possible to compare survival in two or more groups. Finally, survival analysis includes the regression models for assessing the relationships between the multidimensional continuous variables with values of lifetimes type [3].

**The results of constructing a model of survival analysis.** In the model of commercial banks survival analysis, which is offered in this paper the input data are the crisis dates and the bankruptcy date or commercial banks connection loss. For example, the first bank crisis situation happened on the 6 of January 2009, and on the 1 of January, 2014 the connection was lost, so it means the incomplete observation (variable censored). The input data also include the assets of commercial banks, the level of banks financial stability and the stability of the group name to which the commercial bank belongs. Statistical data are obtained on the official website.

Density of probability is a density of bankruptcy probability at this interval when survival function for the next period is subtracted from the survival function of this interval and then divided by the length of the interval (Fig. 1).

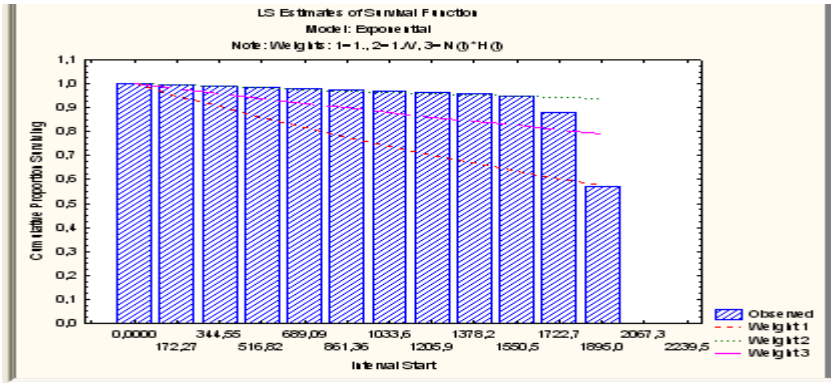



Fig. 1. Graph of density probability estimation

Survival analysis model is made by using STATISTICA.

The graph density estimation (Fig. 1) shows that the probability of bankruptcy is almost equal, but 1378 days later it drops. That is the main banks hold out many days after the crisis, and then the situation can be stabilized.

Fig. 2 shows the Kaplan-Meier estimation of survival functions.

Kaplan-Meier (Product-limit) analysis (Kursovaya1)
Note: Censored cases are marked with +

| Case Number | Time | Cumulativ Survival | Standard Error |
|---|---|---|---|
| 28+ | 1607,000 | | |
| 86+ | 1612,000 | | |
| 39 | 1616,000 | 0,959694 | 0,023113 |
| 3+ | 1676,000 | | |
| 12+ | 1678,000 | | |
| 49+ | 1678,000 | | |
| 4+ | 1685,000 | | |
| 25+ | 1685,000 | | |
| 34+ | 1687,000 | | |
| 74+ | 1691,000 | | |
| 33+ | 1691,000 | | |
| 24+ | 1691,000 | | |
| 60+ | 1693,000 | | |
| 50 | 1694,000 | 0,940877 | 0,029336 |
| 73+ | 1694,000 | | |
| 97+ | 1695,000 | | |
| 85+ | 1695,000 | | |
| 61+ | 1699,000 | | |
| 11+ | 1699,000 | | |
| 45+ | 1704,000 | | |
| 52 | 1715,000 | 0,919493 | 0,035620 |
| 14+ | 1721,000 | | |

Fig. 2. Kaplan-Meier estimation of survival function

Fig. 2 shows, for example, that the probability for the bank to operate for more than 1616 days is equal to 0.9596, the probability for the bank to continue the activities for more than 1694 days is equal to 0.94, etc. The first column of the table (Fig. 2) shows the numbers of observations where at any given time some event occurred, the sign "+" means that the bank is censored (continued operations). Graph of survival function is shown in Fig. 3.

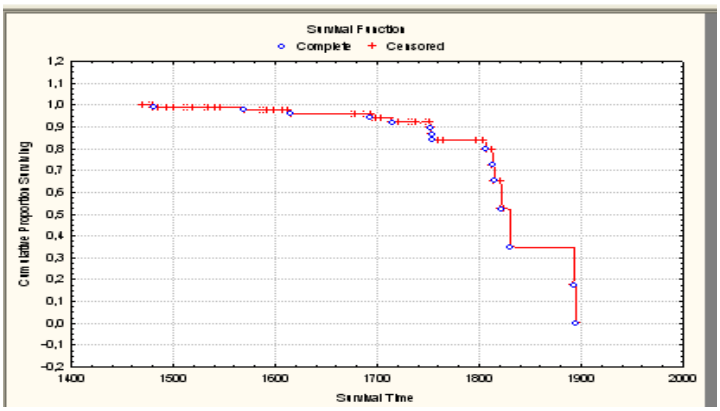



Fig.3. Graph of survival function

The advantage of the Kaplan-Meier method (compared with life table method) resides in the fact that the estimates are independent of partitioning lifetimes in intervals. The research also compares pairs of banks groups with different levels of financial stability, such as high, low and sufficient obtained from official rating of National Bank financial stability of banks. Comparison of banks with a high level of financial stability and a group of banks with low levels of financial stability is shown in Fig. 4.

```
Variable: number of days computed from data
Variable with censoring indicator: CENSORED
Grouping variable: GRUPPA UST.
Total number of valid observations: 72
    uncensored: 13 ( 18,06%)          censored: 59 ( 81,94%)
Valid observations:  Group 1 (Vys      ): 13      Group 2 (Nizkiy   ): 59
         Uncensored:       2 ( 15,38%)                11 ( 18,64%)
           Censored:      11 ( 84,62%)                48 ( 81,36%)
```

Fig. 4. Results of the comparing survival of banks groups with high and low levels of financial stability

As you can see, the percentage of survival is higher in the banks groups with a high level of financial stability. Comparison of groups with a high and sufficient level of financial stability is presented in Fig. 5.

```
Variable: number of days computed from data
Variable with censoring indicator: CENSORED
Grouping variable: GRUPPA UST.
Total number of valid observations: 25
    uncensored: 4 ( 16,00%)          censored: 21 ( 84,00%)
Valid observations:  Group 1 (Vys      ): 13      Group 2 (Dost     ): 12
         Uncensored:       2 ( 15,38%)                2 ( 16,67%)
           Censored:      11 ( 84,62%)               10 ( 83,33%)
```

Fig. 5. Results of the comparing survival of banks groups with high and sufficient levels of financial stability

Percentage of survival in the high-resistance is still higher. Thus, we can conclude that the current level of financial stability of the bank has a significant impact on the commercial banks survival in crisis situation.

**Conclusion.** The proposed model of the commercial banks survival analysis makes possible to determine the factors that affect the commercial banks survival, as well as to compare survival between several groups.

This model can be adapted to any group of banks, and it can take into account different parameters, depending on the analysis nature. It will help to determine the factors of survival in conditions of instability, financial crisis, and competition and so on.

## References:


1. Law of Ukraine "About restoring debtor's solvency or recognition the bankrupt» // http://zakon2.rada.gov.ua/laws/show/ 2343-12 (Ukrainian).
2. Analysis of Ukraine banks: reviews, graphics, facts // http://bankografo.com/analiz-bankiv/bankivska-statystyka (Ukrainian).
3. Borovikov V. STATISTICA. Art of data analysis on the computer: for professionals. 2nd ed. – St. Petersburg, 2003. (Russian).
4. Brishtelev A. Theoretical and methodological basics of banking system economic security / A. Brishtelev // Banking Journal, 2009. – №10. (Russian).
5. Karminskiy A.M. Default models of Russian banks / A.M. Karminskiy, A.A. Peresetskiy, S.V. Golovan. – M.: RES, 2004 – 11 p. (Russian).
6. Malugin V. Banks stability estimation based on econometric models / V. Malugin, E. Pytlyak // Banking Journal. – 2007. – №4. – P. 30–36. (Russian).
7. The National Bank of Ukraine – official website – http://www.bank.gov.ua/control/uk/index (Ukrainian).
8. Tishchenko V. Formation of methodological approach for estimation of Ukraine banks financial stability / V. Tishchenko, I. Popov // Collection of scientific papers «Financial and credit activity: problems of theory and practice» KhIB UB NBU, – Kharkiv, – 2013. – №1(14) – P. 24-31. (Ukrainian).
9. Kobylin A. Application of information technology for monitoring and forecasting of a particular bank and stability of the banking system as a whole / A. Kobylin, R. Sementsov// Collection of scientific papers «Financial and credit activity: problems of theory and practice» KhIB UB NBU, – Kharkiv, – 2012. – №1(12). – website – http://fkd.khibs.edu.ua/pdf/2012_1/26.pdf (Ukrainian).

# The Use of Multimedia Components in Modern Technology of Mobile Learning

**Yevhen Hrabovskyi[1],**

[1] Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine
Jenya37@yandex.ru

**Abstract.** A comparative analysis of the efficiency of various multimedia components in modern technology mobile learning. Detail the possibilities of learning through mobile telephony. A conceptual model of mobile learning. Analysis of specific forms and methods of implementation of mobile technology in the learning process.

**Keywords.** Mobile learning, conceptual model, multimedia components, learning process.

## 1.Introduction

Each modern universities from school and finishing higher education institutions, students are often using different mobile devices, which is why the question arose use mobile devices in learning, this is marked development of mobile learning. Trends contributing to the development of functionality of portable electronic devices encourage scientific world community to reflect on their use for educational purposes.

In [1-7] analyzes the advantages and disadvantages of mobile learning, there are uses of mobile learning into practice, Higher Education, examined mechanisms of mobile learning specific academic disciplines. However, in the specialized literature is lacking a comprehensive consideration of the specifics of multimedia components using modern technologies of mobile learning.

The purpose of this section is to systematize mechanisms of using multimedia components in modern technology mobile learning.

## 2.Definition of Mobile learning

Mobile learning is associated with electronic and distance learning, the only difference is the use of mobile devices. Training takes place irrespective of the location and is using portable technology. Mobile learning facilitates access to learning materials, so that removes geographic limitations through a variety of mobile devices.

Mobile learning (m-learning) - is learning in an environment where the student has mobile access to educational resources, can interact with the teacher and other students. [1]

Mobile learning implements the principles of open education: flexibility, modularity, independent of time and place, the use of modern information and communication technologies. The mobile learning come first didactic principles such as multimedia, interactivity, accessibility [1].

## 3.Current Mobile learning Trends

The possibilities of mobile learning:

• Students use laptop computers, PDA or portable voting systems in the classroom or lecture room.

• Students who use mobile phones and handheld computers in the classroom, increase collaboration between students and teachers.

• The industrial training those who have access to education on mobile phone, gain knowledge just in time to find a solution to the problem that occurred.

• Training in museums and galleries using portable technology.

• Training outdoors, such as production practices.

• Using technology to support personal informative or lifelong learning, such as using handheld dictionaries and other resources for language learning.

• Improving literacy development thinking and participation in education among the younger generation.

• Provision of audio and video support, to improve training, provide in a corporate environment or in another classroom.

Schematic view of the concept of mobile learning is presented in Fig. 1.

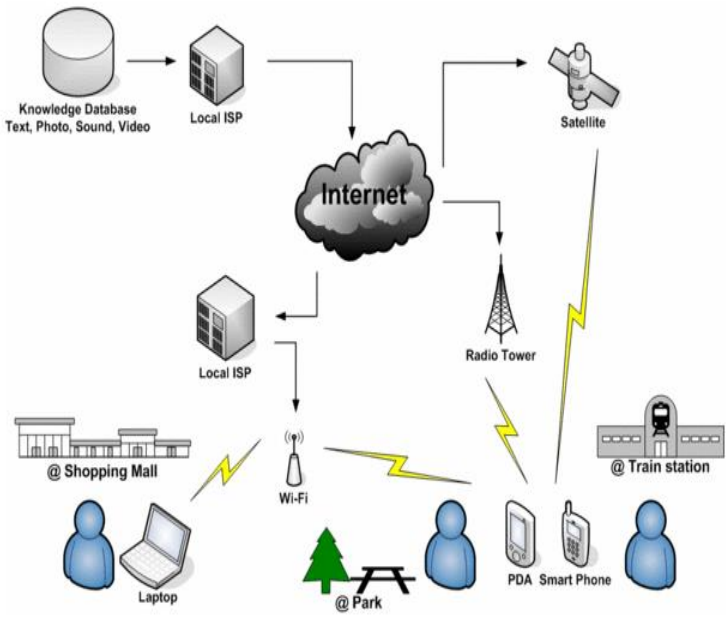



Figure. 1. The scheme of mobile learning

Over the last decade, mobile learning has become an object of study for a variety of projects in schools, workplaces, museums, cities and other regions.

The use of mobile computing devices (such as netbooks, tablet PCs, PDAs and smartphones) with wireless networks can talk about mobile learning. Because of them, there is an opportunity to teach and learn in any place outside the traditional audience. It is any

activity that allows people to be more productive by interaction, or create information using digital handheld devices that can be stored in a pocket or purse.

Hardware devices for mobile learning can be called [2]:

- Phone: ordinary mobile phones, smartphones, communicators;
- Portable computers: laptops, netbooks, Internet tablets;
- Storage and reproduction of information: electronic "readers» (Pocket Book, Amazom Kindle), MP3 / MP4 players.

In order to determine the relevance of implementing mobile learning in the learning process, we have made a meaningful analysis of the works of scholars who have studied the issue.

General view of the editor applications for mobile learning is presented in Fig. 2.

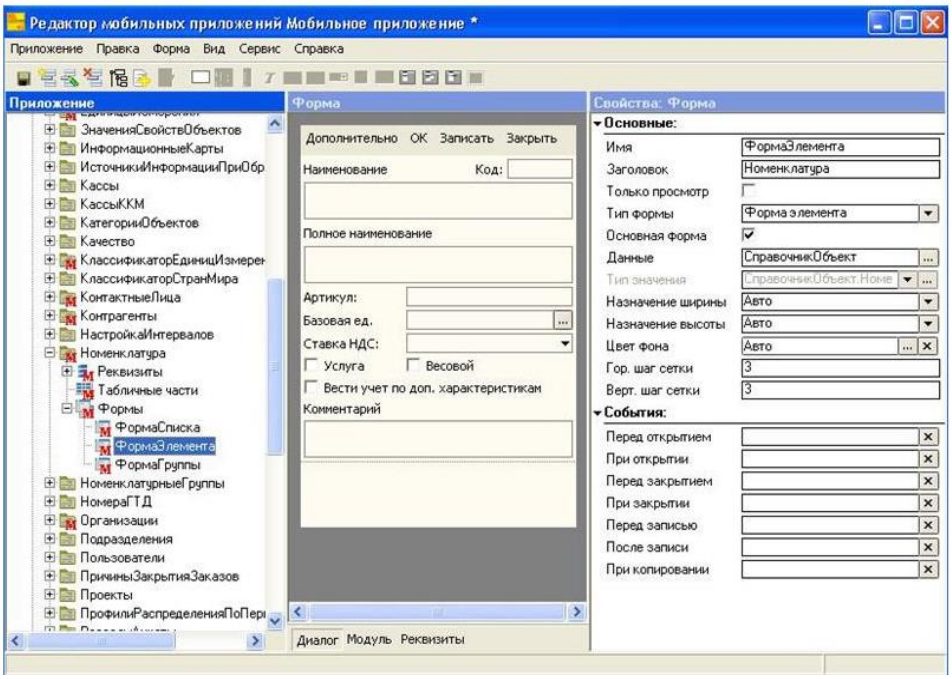



Figure. 2. General view of the editor applications for mobile learning (print screen, original image)

Mobile learning has its advantages and disadvantages.

K. Bugaichuk identifies the following advantages of mobile learning [3]:

1) Convenience and flexibility. The educational process with the use of mobile devices can be carried out regardless of time and location of students and teachers.

2) Mobile learning allows the use of "dead time". This is generally, travel time, queuing, waiting for meeting, etc.

3) Mobile devices can be used to quickly create and further transfer of information (photo, video, voice recorder, SMS, MMS, etc.).

4) Overcoming technological barriers. Use of the devices allows the listener further experience and encourages him to seek new ways to use their own device, including to meet training needs.

5) Personalization of learning. All information is only displayed to the user device to the same mobile device configured according to individual needs of the listener.

6) Mobile learning is suitable for different styles of learning: reading, viewing photos and videos, listening to podcasts, participate in discussion (forum or SMS), search for information online, testing, and participate in contests and more.

In turn, T. Brabason identifies the following advantages of mobile learning [5]:

1) Attracting potential students with unreached people groups.

2) Educational materials are available to a wider audience through podcasts, mobile applications, blogs and e-books that get the attention of potential students.

3) Help to the some social groups for which mobile learning represents an opportunity to improve their standard of living.

4) Improve the quality of teaching.

5) Renewal of the study programme, teaching methods and rethinking create better feedback to students.

6) Support for communication with the school pupils, their academic progress and transition between schools.

M. Sharpless also identified several advantages of mobile learning, such as [4]:

1) Means of voluntary and active learning.

Mobile learning environment is about access to content, colleagues and experts. It can be activated using a smartphone or tablet laptop.

2) Transparency. Transparency is a natural by-product of the connection. Students get in touch with local and global communities through social media - platforms of Twitter in Facebook, LinkedIn to Pinterest.

3) Training in practice. The game is one of the main characteristics of progressive education. The mobile learning environment, students are faced with a dynamic and often unplanned data set by changing the tone of academic lessons with elements of the game, acquiring skills becomes easier.

4) Asynchronous learning. Among most powerful principles of mobile learning is asynchronous access. This removes the educational environment of the school's surface and allows it to move in any place at any time.

Benefits of mobile learning by S. Semerikovym [6]:

1) Updating of learning through "fashion" on the device, includes more than the number of students in mobile learning.

2) Increased interactivity training: basic operations are performed online - mode, mode interactive cognitive activity.

3) Concentrated content: learning mobile learning objects more closely related to each other at micro-learning.

4) Developed means of cooperation. Quality education is rarely obtained in alone, and one of the best ways to successful training is teamwork is important for the exchange of ideas.

5) Ease of use of mobile learning at any time and in any place.

Implicitly mobile learning has many advantages, but along with the benefits of serving a number of shortcomings.

K. Bugaichuk out such shortcomings mobile learning [3]:

1) Lack of high-quality, full-featured educational content for mobile and tools of project development.

2) Low "technical" training of trainers in the field of mobile content.

3) There is a perfectly developed technique and evaluation results support mobile learning.

4) limited battery life of mobile devices.

5) Lack of, sometimes, high-quality Internet coverage. In Ukraine, only one operator is licensed in touch 3G, other networks that operate in the standard CDMA 2000, covering a small area (usually big cities). In addition, almost no schools support the development of wireless network.

6) The presence of "price barrier" for users. In addition to the cost of the device, it is necessary to mention the cost of the mobile Internet. Ukrainian mobile operators offer Internet packages cost from 30 USD per month, which should also add to the cost of education;

7) The presence of "price barrier" for schools.

L. Naismith identified the following shortcomings mobile learning [7]:

1) Battery. Educational programs on mobile devices "kill" batteries much faster than the standard "gadgets".

2) Confidentiality and security. Use of any gadget requires prior instruction on building safe and ethical relationships with others. Unfortunately, not invented universal method that would help prevent identity theft, harassment and intimidation.

3) Screen. Small touch screens are pretty much standard for tablets and smartphones, but are more problems in school. In some cases, they are inaccurate and blurred respond to user requests, while others, on the contrary - too sensitive to any stimuli.

4) Loss and theft. Because of its compact mobile devices are often lost or exposed to theft.

5) Compatibility. Compatibility - or rather incompatibility of mobile devices - one of the issues over which now scratching their heads hundreds, even thousands of programmers. Currently there is no universal platform for mobile gadgets, as synchronization projects seems incredibly challenging.

6) Storage. Storage is closely related to privacy, security, and other things.

Disadvantages of mobile learning by S. Semerikovym [6]:

1) The problems of security devices, and the data in it.

2) The small screen size and difficulty of access.

3) Lack of well-developed skills students objective knowledge (the ability to understand and control their own learning process).

4) The high cost of the initial deposits in training.

So, analyzing the benefits and disadvantages of mobile learning by different authors, can distinguish the following main advantages of mobile learning: user-friendliness; quick access to information; portability; the ability to attract large numbers of people; rapid transfer of information; and the main disadvantages of mobile learning: the high cost of mobile small screen size of the mobile device and vulnerability information from unauthorized access, limited battery life of mobile devices.

## 4.Support for mobile learning

Most personal technologies can support mobile learning, including:

• Personal digital assistant in the classroom and outdoors

• Tablet PC UMPC Mobile Phones, Mobile Phone, and Smartphone

• Learning Mobile Author, e.g. for publishing WAP, J2me and Smartphone
• Personal audio player, e.g. for listening to audio recordings of lectures
• Portable audio and multimedia guides in museums and galleries
• Portable consoles for gaming

Support for mobile learning is implemented in the form:

• 3GP for compression and delivery method of audiovisual content associated with mobile learning
• Wi-Fi provides access to teachers and resources on the Internet
• GPRS service provides high-speed connectivity and normal data.

Recently in higher education can be seen increased activity in the implementation of new methodologies to enhance the learning process, based on information and communication technologies [4, p. 45].

This is the impetus for change in all areas of the educational process:

• modernization of working methods;
• the introduction of new solutions lectures, practical laboratory classes, study time;
• changes in the selection, presentation and methods of distribution educational material;
• changes in methods and organization of teachers and students;
• changes in the organization and calculation of working hours;
• Development of individual working methods;
• Personalization of user study (adaptation to the needs of students);
• changes in the working methods of the group;
• Changing the type of assessment of student learning;
• Changing the type of individual assistance to students.

The need for the introduction of these changes came naturally, students share observations with each other that they derive from different cells, watching with interest the development of teaching methods that support the technology innovations.

Traditional teaching structure enriched m-learning - a way to a new quality of learning. Students are in a situation in which learning method always change (during the day using several different forms of study) [5]. Dynamics study is based on the alternate use of different ways of learning: student permanently present in the classroom, some of which are included in the mobile communicators, the other classes designed to deepen self-knowledge, presented by the Internet and taken with mobile agents. In the further course of the process, the student can annotate materials, projects or tasks and participate in different forms of the group. Teachers plan organization working with non-rhetorical hand, while technically monitoring its science [6]. You must pay attention to the fact that the younger generation accustomed to change and be more dynamic and intellectual stimulation than was a few years ago. All these conditions do methodological structures that combine various forms of work.

Various activities are combined into a system of training students and contribute to the overall implementation of the learning process.

Let us consider the specific forms and methods of implementation of mobile technology in the learning process.

1. Mobile phone provides Internet access to sites with educational information - is used as a form of distance learning.

The first (and most common) way is to use a mobile phone as a means of access to the global network. The organization of access to specialized sites that contain e-learning courses, tests, practical exercises and additional learning materials (drawings, photographs,

audio and video files). You can also e-mail exchanges for educational purposes and instant messaging program ICQ, QIP, the versions that exist for mobile phones. Therefore, at all stages of education, there are many opportunities for the transfer of information materials, and control the whole process of training and assistance in solving problems.

An example of the use of mobile phones is a project of M-Ubuntu [3], developed by the Swedish organization coarse Learning Academy Worldwide in the second half of 2007. The project was been presented by distance learning platform that allows you to create the conditions for new knowledge and active use of new information technologies, even in remote areas and third world countries. Particular attention M-Ubuntu developers paid training through mobile phones, and use such a system cannot only students. Especially for teachers were developed programs for training and testing program and control students. Any teacher, regardless of location, platform using M-Ubuntu, able to consult with professors largest universities.

2. Mobile Phone - means playing audio, text, video and image files containing educational information.

The use of mobile phones for learning - the use of special software platforms for mobile phones that are able to open and view files office applications, such as Office Word, Power point, Excel. Thus, having your mobile phone files that contain educational information, you can view their version, specially adapted for the screen phone with easy scroll bars corresponding type and user friendly interface.

In addition, a source of information can serve video and audio files, programs, players that are in each phone last year's release. Especially valuable this opportunity is for those who want to learn a foreign language - a huge number of available audio course and audiobooks, including files of various formats and lengths [6].

Example user interface m-learning Android given in Fig. 3.


Figure 3. The user interface of m-learning Android

An example of the successful application of this method of study has a number of educational programs at universities in Japan and China. Considering mobile technology, teachers of universities consider them very promising in terms of informatization of modern society. National Cyber Institute in Japan that specializes in distance learning via the Internet, in 2008, proposed an innovative training system - using a mobile phone that allows

you to learn any discipline, both at home and in cafes or in the subway. If the computer during class at the centre of the screen shows the text of the lecture, and all the necessary drawings, and in the corner is a video broadcast of the lectures, the version for mobile phones based on the technology of streaming video, all texts and drawings pumping extra. Students were asked to learn about 100 different items, including ancient Chinese culture, journalism and English literature [7].

In China, the company Nokia is developing a program Mobiledu, which began in 2007 and includes English-language training materials and other educational content from a huge number of content providers directly to mobile phones. Get access to this information is possible through Nokia mobile phones or via the website of the program. During the program Mobiledu more than 20 million people became its subscribers.

3. Mobile phone and its functionality can organize learning using adapted electronic textbooks, courses and specialized file types with educational information - tutorials developed specifically for mobile platforms.

Another way to use mobile phones for learning is the use of specialized books and courses adapted to view and execute on mobile phones students. Students are encouraged to download to their phone Java- applications that contain, for example, testing of certain items and information (electronic books, texts of lectures) required for their successful implementation. Modern technologies allow fairly easy to design and implement software such electronic aids. Maximum placement schemes, drawings and formulas makes writing e-learning courses for mobile phones universal and applicable to absolutely any subject being studied. It is also possible the implementation of educational programs in the shell game, using the graphics capabilities of phones, but the implementation of such applications - is complex and time-consuming process. As a result, writing e-books and programs substantive testing for mobile phones seems to be more promising direction. There are a number of special programs for mobile phones, such as calculators varying degrees of complexity (simple, scientific), office applications for mobile phones, applications that contain various tests with answers (e.g., psychologists), etc. [5].

To the process of mobile learning should include:

- presentation and learning educational material;
- mobile communications;
- mobile control knowledge;
- the skills and abilities;
- support for mobile learning.

It is necessary to organize the process of mobile learning in open distance education to students given the opportunity to: a) obtain the necessary fundamental knowledge, interpreting them in a way to use them to solve specific cognitive or practical problems; b) discuss with the teacher and (or) other members of the educational process of mobile learning occur in the learning of the problem; c) have access to additional sources of information needed to address the cognitive tasks; d) to observe, perform independent research and experimental work using wireless networking technologies to understand the acquired knowledge, development of skills and resolve problems; e) be able to reflection own cognitive efforts succeed adjusting its training activities.

The main components of the organization - educational software that improves the efficiency of mobile learning are: a) the adoption of mobile learning teaching staff in educational institutions; forming their motivation, operational and reflexive willingness to achieve the objectives of mobile learning; b) market research to identify niche mobile learning in open distance education; c) orientation methodological support mobile learning to

ensure that the content and results of mobile learning requirements of state educational standards; d) adaptive systems providing mobile learning methods; e) the cyclical process of methodological support mobile learning; e) the willingness of teachers to mobile learning; i) logistics educational process of mobile learning research, training and technical literature, software for educational purposes; c) monitoring of teaching status and outcomes of the educational process of mobile learning; i) comprehensive informatization system of mobile learning.

Actually pedagogical conditions of implementation of mobile learning can be: a) operation of educational communication in a forum, chat, blog, multimedia files; b) management of the educational process of mobile learning by the teacher; c) the presence of the media access to learning resources; d) the use of games, computer simulations, services Web 2.0; e) support collection development of digital educational resources to mobile devices; e) understanding of the meaning of educational content pupil; g) the safety of the educational process.

The result of the introduction of mobile learning is at the intersection of pedagogy, modern information and communication technologies and their applications, not every teacher is able to work at the intersection of these three areas and be both in their professional conductor innovation.

Thus, broad technical and functional capabilities of mobile phones are used for educational purposes as follows:

• Use the opportunity SMS-mailing or instant messaging with the teacher for advice;

• access to a global network lets you access sites needed, share via email, send the necessary information files;

• testing the mobile phone allows the student to independently control the level of knowledge of the subject;

• Electronic tutorials for mobile phones make it possible to obtain new information regardless of time and location of the student;

• Ability to play audio, image and video provides advanced capabilities, especially for language learning objects and creative fields, allows the use of different sources and methods of obtaining knowledge, interest unusual teaching methods of teaching;

• Mobile analogues language dictionaries and reference books, various types of mathematical calculators are easy to use and can contain a full and promptly updated information.

## 5.Conclusion

Most of today's students technically and psychologically prepared for the use of mobile technology in education, and should be considered new opportunities for more effective use of the potential of mobile learning. Meeting this challenge requires effort on the part of leaders of education, research and technical work of scientists and teachers with implementation strategies, forms and methods of mobile learning in the educational process of the university.

## References

1. Attewell J., Savill-Smith C., Douch R., Parker G. Modernising Education and Training: Mobilising Technology For Learning. LSN, London. 2010. URL: https://crm.lsnlearning.org.uk/user/order.aspx?code=100103.
2. Bugaichuk K. Mobile learning in higher education // Mobile learning. – 2008. – № 2. – P. 48-49 [In Russian].
3. Sharples, M. (2007). Big issues in mobile learning: Report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative. Nottingham, UK: University of Nottingham, Learning Sciences Research Institute.
4. Brabazon T. Mobile Learning: the iPodification of Universities / T. Brabazon, 2007. – 7 p.
5. Semerikov C. Foundation of informative training courses in higher education: Monograph / Science Academy of Pedagogical Sciences of Ukraine editor, prof. M. Zhaldak. - 2009 - 340 p. [in Ukrainian].
6. Kumar S. Blackboards to Blackberries: Mobile Learning Buzzes across Schools and Universities. 2010. URL: http://www.learningsolutionsmag.com/authors/315/sesh-kumar.

# The Economic Essence of the Financial Security Definition of an Enterprise

**Iryna Bylomistna[1], Oleksii Bylomistnuy[2], Alina Kovalchuk[3], Rekhletskyy Evgen[4]**

[1] PhD, Kharkiv institute of banking UB NBU, Peremogy av.,55, 61174, Kharkiv, Ukraine
belomestnay@khibs.edu.ua
[2] PhD, Kharkiv institute of banking UB NBU, Peremogy av.,55, 61174, Kharkiv, Ukraine
belomestny@khibs.edu.ua
[3] PhD student, Kharkiv institute of banking UB NBU, Peremogy av.,55, 61174, Kharkiv, Ukraine
alya_kovalchuk@mail.ru
[4] PhD, Lviv Academy of Commerce, Tugan-Baranovskogo str. 10, 79005, Lviv, Ukraine
RekhletskyyEvgen@gmail.com

**Abstract.** The article considers approaches to the definition of the financial security of a company, its components and economic substance. We propose the classification of threats to financial security, these negative consequences from internal threats that have the greatest impact on the state of development of the enterprise and its profitability.

**Keywords.** financial security, threats, stability, economic substance.

## 1.Introduction

The driving force of any process is the desire to get something better and more perfect. Therefore, it is not surprise that the companies are trying to increase their profits. However, more enterprises are facing the problem of the bankruptcy lately.

It may occur because insufficient amount of the available cash resources and reduction the liquidity and solvency of the enterprise or its debtors. It is important to find an approach to the organization of the activity that is what will prevent the enterprise from potential risks and threats. These functions perform the financial security of the enterprise. They provide the avoidance of various threats or help to avoid and prevent them.

## 2.Definition of the Financial Security

Ukrainian scientists like Baranovsky O., Blank I., Horiacheva K., Arefyeva O., Donets L., Papehin R. considered in their researchers financial security. They give many definitions of financial security and two of them are most complete. Blank I. says that financial security of an enterprise determines quantitatively and qualitatively the level of its financial position, which provides a stable defense of his priorities of balanced financial interests from the identified real and potential threats of external and internal nature, it's parameters are determined on the basis of his financial philosophy and create the necessary preconditions for financial support sustainable growth in the current and future period. While the Papehin R. considers that financial security as the company's ability to independently development and implement financial strategy in accordance with the purpose of the overall corporate strategy under uncertainty and competitive environment.

For more detailed analysis of the essence of the enterprise financial security's definition one should analyze the definition and interpretation of it as an economic category. In general, we may define the following approaches to the definition of the concepts that are presented in Table 1.

Table 1

Systematics interpretations of the enterprise financial security's concept

| Symptom | Solution | Approach |
|---|---|---|
| Component of economic security | Pokropyvnyy S.<br><br>Sudakova O. | 1. Financial security is a component of economic security, which uses most effectively the corporate resources.<br>2. Financial security is an important component of economic security, based on the independence, efficiency and competitiveness of business finance. |
| State security interests of the company | Zahorodniy A. and Voznyuk G.<br><br>Blank I. | 1. Financial security is a security of financial interests of businesses at all levels of financial relations, security of households, businesses, organizations and institutions, regions, industries and sectors of the state in general by financial resources to meet their needs and financial obligations.<br>2. Financial security - a quantitatively and qualitatively determined level of its financial position, which provides a stable defense of his priorities for balanced financial interests against real and potential identified threats of external and internal nature, it's parameters are determined on the basis of his financial philosophy and create the necessary preconditions for financial support sustainable growth in the current and future period |
| Will be an effective use of enterprise resources | Arefyeva O. and Kuzenko T. | 1. Financial security of a company is a state while there are used most efficient the corporate resources, expressed in the best values of financial profitability, management quality and the use of fixed and circulating funds, capital structure, payment standards on securities and the market value of its securities as a synthetic indicator of the current financial and economic situation and the prospects of its technological and financial development. |

The analysis presented by scientists' interpretations of the essence of financial security concept leads to the conclusion that financial security is an integral component of

the economic security of an enterprise, so the effectively functioning company as a business entity has to use its turnover sufficient funds to protect against internal and external threats.

Consequently, it is possible to characterize the notion of a company's financial security as a state of open socio-economic system, which is characterized by the absence of various dangers and threats, or opportunities for their prevention, protect their interests and prevent losses below the critical limits, and the ability of this system to ensure the effective functioning in the current and future period. This allows to consider the financial management of enterprise security as a set of measures, which makes possible to neutralize the destabilizing factors on the cash flows of the economic agent and return them to a state of equilibrium (a state of normal system functioning) in the oscillations of the enterprise functioning in the framework of the basic parameters of its stability.

Under the condition of normal operation of the system implies that the system state in which there is continuous between the time distribution of financial resources from investors to savers; goal is saved and the system parameters during the period of its functioning. Financial risks are proactively identified and predicted, and the system can absorb without changing the relations of its elements.

The concept of "financial security" includes many elements, such as financial security, objects, subjects, threats, principles and criteria for evaluating the system effectiveness of this.

You can define such factors of financial security of a business entity:

- the level of financial resources provision;
- the stability and resilience of financial condition;
- the balance of financial flows and settlement relations;
- the level of control over external and internal factors.

The level of company financial security depends on how well its leaders and professionals (managers) will be able to avoid potential threats and eliminate the harmful effects of certain negative elements of external and internal environment.

What’s about the economic substance of the "financial security" concept, we can distinguish such main characteristics:

- Ensure equilibrium and stable financial condition;
- To promote the efficient operation of the enterprise;
- Early warning of enterprise threats;
- Crisis neutralization and avoid bankruptcy.

Threat is the events development as a result of which increases possibility or appears probability disruption of an enterprise and incurs losses. Threats according to economic literature are dividing into internal and external.

The internal dangers and threats include intentional or accidental error management in the financial management of the company associated with the choice of business strategy, management and optimization of assets and liabilities of the company. The effects of internal threats is difficult to predict, but the following main effects are indicated in Figure 1.

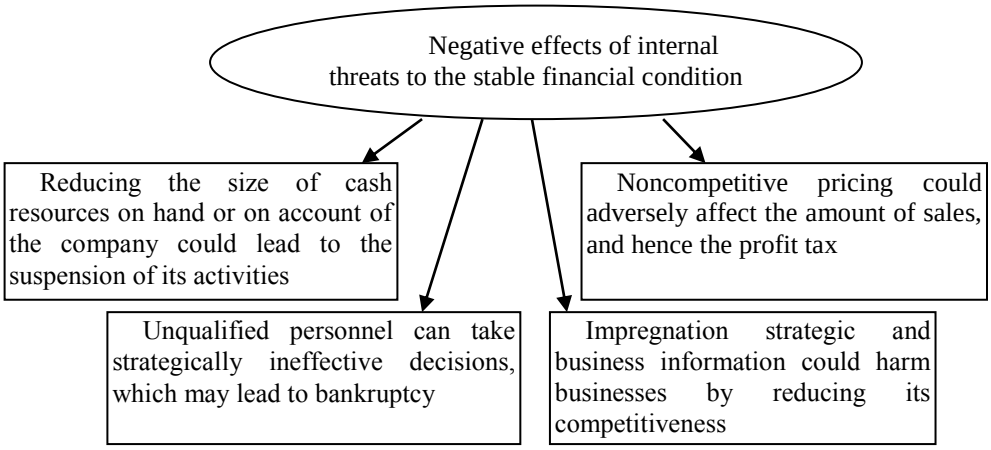



Figure 1. The effects of internal threats to the financial security of an enterprise

External threats include:
- Buy up of shares of company debts by undesirable partners;
- Presence of significant financial obligations of the company;
- Underdeveloped capital markets and their infrastructure;
- Underdeveloped legal system of investor protection and enforcement;
- Monetary and financial crisis and credit systems;
- Instability of the economy.

Today, the methodology of financial security formation of businesses, unlike, the financial security of banks, insurance companies, are not well developed. Therefore, it is necessary to pay more attention to the implementation of mechanisms to ensure the financial security of business entities, which associated with a significant amount of risk and volatility, both internal and external environment, as well as threats of acquisitions, in particular, through the bankruptcy procedure.

The financial security of the enterprise as a complex economic system implies a differentiated level of quantitative and qualitative parameters of the protection of the financial interests, adequate financial level. Consequently, the company's financial security is a system, which ensures the stability of important financial ratios the company's development, which form the protection of its financial interests.

However, considering the financial security of the enterprise as a stable system, attention has to be paid to the following main points:

1. the stability of the financial system business security is a dynamic characteristic, that depends on the speed of enterprise development, its independence may change the evaluation criteria, methods and management tools for financial security;

2. the stability of the financial safety of the company does not have an absolute character. It means that the system of financial security, for two similar companies, may have completely different components and the way protection from threats, so enterprises can be in different frames, with the level of training and it will be different directly affect an the state of security and governance financial security, and various levels of technical tools, etc.;

3. at some stages of its development, the system of company's financial security may vary, reaching the new level settings to its stability, and therefore, depending on the strategy of the company, its mission and goals and development, you need to keep up with changed factors and criteria for its effective operation all the time, which were selected by the enterprise, and depends on it and implement their own methods of protection against real and potential threats.

4. An important goal of financial security of an enterprise is the creation of the necessary financial requirements for sustainable growth in the current and future periods.

Financial security is a universal category that characterizes the subject protection of socio-economic relations at all levels it starts with the state and every citizen. Financial security of the enterprise reflects the security of its activities from the negative influences of the external environment, as well as the ability to quickly eliminate various threats or to adapt to existing conditions that do not adversely affect its operations. The content of this concept includes a system of tools that ensure competitiveness and financial stability of the enterprise, as well as contribute to improving the welfare of workers.

Research financial security of business entities today is quite problematic. There are very few scientific developments in this field. Much more attention is paid to studying the economic security of the whole enterprise. The need demand for continuous compliance with economic security is determined objectively existing for each entity the task of ensuring the stable functioning and achievement of the main objectives of its activities. The level of company financial security depends on how effectively its leadership and it will be able to avoid potential threats and eliminate the harmful effects of some negative components of the external and internal environments.

## 3.Conclusion

Thus, the businesses financial security is an important component and represents the ability of the enterprise to conduct its business, including financial activities, efficiently and stabilization for an indefinite period of time, through the use of a set of interrelated diagnostic, instrumentation and control measures of a financial nature, which should optimize the use of financial resources, to ensure their proper level and to minimize the risks of internal and external environment.

Ensuring stable operation of the enterprise depends on the condition of financial security on it, so you need to give her a lot of attention and use all necessary measures and methods to maintain it at the proper level. It should be remembered that financial security is one of the stages of increasing solvency and competitiveness of the enterprise, so if the company has problems with these aspects of its activities, it should be arranged with the security company, or created a post for the organization so to ensure enterprise financial security.

## References

1. Arefyeva O. (2010). Nature and characteristics of financial security of the enterprise as an economic category. *Journal of transport Economics and industry*, 32, 167-170 [in Ukrainian].
2. Zagorinsky T. (2009). Direction to ensure the financial security of the company. Retrieved from http://archive.nbuv.gov.ua/portal/soc_gum/prvs/2009_1/0346.
3. Orekhova K. (2012). Improvement of the classification factors of the financial security of the enterprise. *Bulletin of the University of banking of the National Bank of Ukraine*, 1(13), 79-86 [in Ukrainian].
4. Matviychuk L. (2009) Conceptual aspects of company's financial security. *Economic Visnyk of the Donbas*, 4 (18), 133-136 [in Ukrainian].
5. Payda-Nosyk H. (2011) The nature of the financial security of businesses and its role in securing national economic security. *Vestnik EDTU*, 1 (55), 340-342 [in Ukrainian].

# Practical Application of Indicators of Social Responsibility and Effectiveness of Activity of the Enterprises

**Ganna Zhosan**[1]

[1] Kherson national technical university, 73000 Kherson, Ukraine
ennzhosan@gmail.com, enn89@ukr.net

**Abstract.** It is established that the calculated indicators of social responsibility and productivity of the activity of the enterprise are dependent in the article. It is defined that it is possible to use them for definition of the best or worst enterprises of branch with a definite purpose: calculation of part of the enterprises in total, of the income which are sent to the budget; calculation of the number of workers who are subject to reduction on each of the enterprises; calculation of the volume of the allocated investments for each enterprise.

**Keywords.** social responsibility of the enterprises, effectiveness of activity of the enterprises.

## 1.Introduction

Indicator of transition of the Ukrainian economy on a new level of economic development is implementation of national projects. On the one hand, strengthening of national economy is a guarantee of development of the enterprises, emergence of new prospects and opportunities of development of business relations. On the other hand, the enterprises are direct participants of social and economic processes. Thus, in relationship to the state, society and the enterprise harmonization of interests and consolidation of efforts of the enterprise for achievement of common goals has to be traced.

Thus main for the enterprise is a problem how to reach the sustainable, balanced, multi-purpose and productive development. It can become possible thanks to social responsibility which will help with regulation of the relations of the enterprise with all participants of external and internal business processes.

## 2.Interrelation of social responsibility and effectiveness of activity of the enterprise

Transition to conducting socially responsible activity of the enterprise includes direct and full submission of the economy to the social purposes. The economic crisis showed the vulnerability of business, its dependence on the state and public opinion, which, according to forecasts of experts, will grow further. In the modern economic space of the enterprise have to be not only profit generators, but also to satisfy material, social and intellectual requirements of society. The social sphere from minor of which it is in industrial society, turns into the priority direction of development of post-industrial economy.

Strengthening of a social orientation of economic growth allows an increase of effectiveness of activity of the enterprises on the basis of creation of social values for all

participants of external and internal business processes. Thus, the mission of the enterprise in the market which consists in need of satisfaction not only material, but also the spiritual needs of the market by producing and today's realization of socially significant production which meets the international standards of quality essentially changes.

Due to the innovative orientation of the economy determined by social and economic policy of the state, all aspects of management of activity of the modern enterprise gain a certain social importance.

Unfortunately, in Ukraine yet not enough керівництв the enterprises realized the need of the doghouse socially responsible and that their image, a place in the market and effectiveness of activity depends on it.

We will consider more in detail versions of the practical application, offered by the author in the previous researches, method of calculation of a complex indicator of social responsibility of the enterprise and the generalizing indicator of effectiveness of activity of the enterprise [2, 3, 4, 5, 6].

## 3.Calculation of part of the enterprises in total, of the income, which are sent to the budget

Additional data, which will be necessary this total of the enterprises of the relevant branch in this region. Therefore, according to statistical data, the number of the enterprises of the food industry in the Kherson area makes 122 [1].

Ranging (Fig. 1) on at step 1 and a step 2 is carried out considering that 1 Rank is awarded to the enterprise with the greatest indicator, and 6 Rank (by number of the enterprises) - to the enterprise with the smallest.

Step 3. Definition of the general resultant rank is carried out in 2 stages.

1 . The indicator is defined $S_i$

$$S_i = \frac{P_{c_i} + P_{P_i}}{2}, \quad (1)$$

where, $Pc_i$ is the rank of a complex indicator of social responsibility i[th] the enterprises;

$Pp_i$ is the rank of the generalizing indicator of effectiveness of activity i[th] the enterprises.

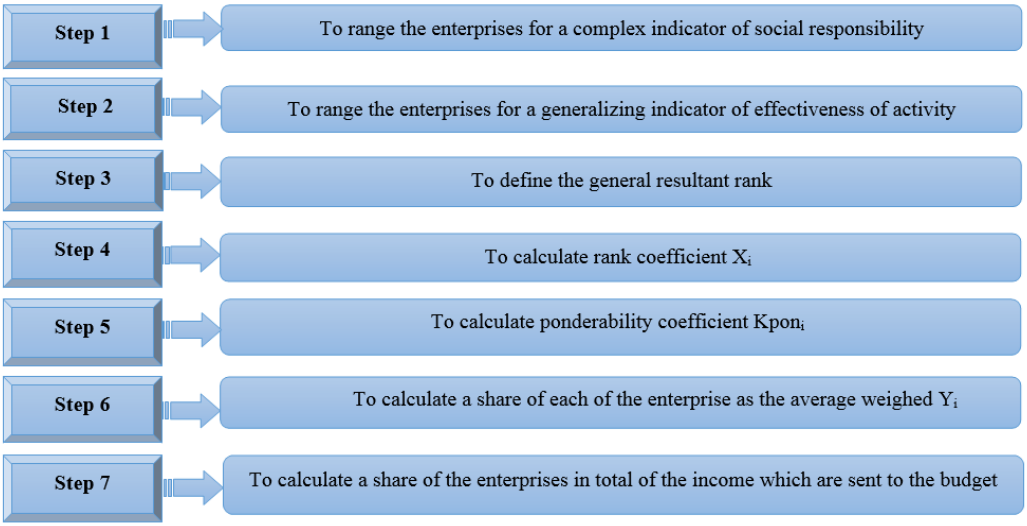



Fig. 1. Calculation of a share of the enterprises in total, of the income, which are sent to the budget

2. The general resultant rank is defined. Thus 1 rank is appropriated to the enterprise, at which $S_i$ indicator smaller.

Step 4. Calculation of coefficient of a rank $X_i$

$$X_i = 1 - \frac{Рз_i}{\sum_{j=1}^{m} r_{ij}}, \quad (2)$$

where, $Рз_i$ is the general resultant rank of the enterprise;

$\sum_{j=1}^{m} r_{ij}$ is the sum of serial numbers of ranks (depending on number of the enterprises). In our case of 6 enterprises, therefore:

$$\sum_{j=1}^{m} r_{ij} = 1+2+3+4+5+6=21. \quad (3)$$

Step 5. To calculate ponderability coefficient $Kpon_i$

Step 6. To calculate a share of each of the enterprise $Y_i$

$$Y_i = X_i \times Kpon_i, \quad (4)$$

where, $X_i$ is the coefficient of a rank of the enterprise;

$Kpon_i$ is the coefficient of ponderability of the enterprise.

Step 7. To calculate a share of each enterprise as the average is weighed $Y_i$.

For an example calculation of a share of the enterprises in total of the income, which are sent to the budget for 2013 is presented (Table 1).

Step 8. Calculation of part of the enterprises in total of the income, which are sent to the budget is carried out in some stages.

1. We define the general share of the studied enterprises among all enterprises of the food industry of the Kherson area:

$$Ч_з = \frac{\text{Number of the studied enterprises}}{\text{Total of the enterprises of the industry}} \times 100. \quad (5)$$

Therefore, the author investigates 6 enterprises of the food industry, total 122. Thus $Ч_з$ = 4,62%.

Calculation of part of the enterprises in total of the income, which are sent to the budget $Ч_i$

$$Ч_i = \frac{Part \times Ч_з}{100}. \quad (6)$$

Table 1

Calculation of part of the enterprises in total of the income, which are sent to the budget for 2013

| Enterprises | Indicators | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $SR_i$ | $P_i$ | $Psr_i$ | $Pp_i$ | $S_i$ | $Pз_i$ | $X_i$ | $1/P$ | $Kpon_i$ | $V_i$ | $Part$ | $Ч_i$ |
| Enterprise 1 | 0,9092 | 0,5980 | 2 | 3 | 2,5 | 1,5 | 0,93 | 0,67 | 0,29 | 0,27 | 30,99 | 1,52 |
| Enterprise 2 | 0,5636 | -0,8658 | 4 | 6 | 5 | 6 | 0,71 | 0,17 | 0,07 | 0,05 | 5,96 | 0,29 |
| Enterprise 3 | 0,5486 | 0,4938 | 5 | 4 | 4,5 | 5 | 0,76 | 0,2 | 0,09 | 0,07 | 7,63 | 0,38 |
| Enterprise 4 | 0,7078 | 0,6108 | 3 | 2 | 2,5 | 1,5 | 0,93 | 0,67 | 0,29 | 0,27 | 30,99 | 1,52 |
| Enterprise 5 | 0,3246 | 0,8767 | 6 | 1 | 3,5 | 4 | 0,81 | 0,25 | 0,11 | 0,09 | 10,13 | 0,45 |
| Enterprise 6 | 1,2437 | 0,2662 | 1 | 5 | 3 | 3 | 0,86 | 0,33 | 0,15 | 0,12 | 14,30 | 0,70 |
| Sum | - | - | - | - | - | - | - | 2,28 | 1 | 0,87 | 100 | - |

## 4.Calculation of the number of workers, who are subject to reduction on each of the enterprises

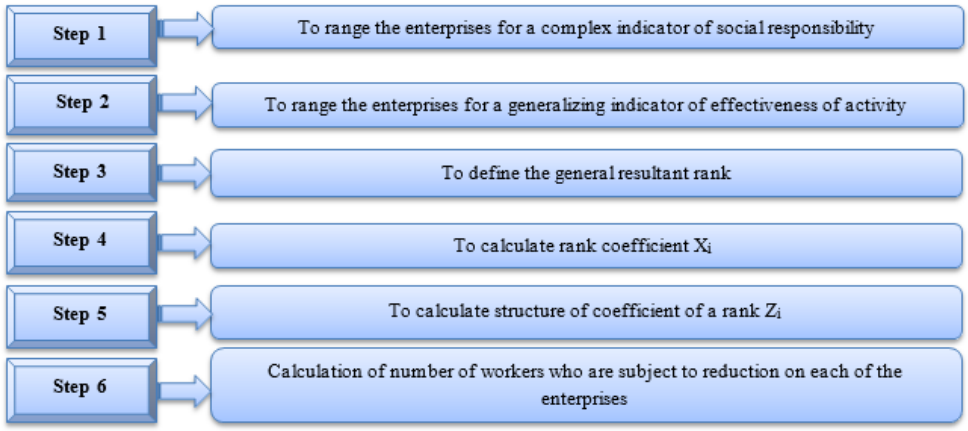



Fig. 2. Calculation of the number of workers, who are subject to reduction, according to the order of the local council

We will consider one more direction of application of the indicators calculated by the author. We will assume that for certain economic reasons the local council gave the order to reduce 500 people in the enterprises of the food industry of the Kherson area. It is necessary to find out how many people it is necessary to dismiss from each concrete enterprise. For this

purpose, at first, we will calculate how many people is necessary to reduce from all studied enterprises (Fig. 2). For this purpose, we find 4,62% of the general 500 people. It makes 25 people.

Steps 1-4 completely we repeat steps from the previous point.

Step 5. Calculation of structure of coefficient of a rank of $Z_i$

$$Z_i = \frac{X_i}{\sum X_i}. \quad (7)$$

Step 6. Calculation of the number of workers, who are subject to reduction on each of the enterprises of the admissible order of a local community is carried out behind a formula:

$$K_i = Z_i \times 25. \quad (8)$$

In Table 2, for an example, calculation of the number of workers, who are subject to reduction on each of the enterprises, to the order of local council for 2013 is presented. It is assumed that is offered to dismiss 500 people from the enterprises of the food industry of the Kherson area, then from the studied enterprises it is necessary to reduce 25 people.

Table 2

Calculation of the number of workers, who are subject to reduction on each of the enterprises for 2013

| Enterprises | Indicators | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $C_i$ | $P_i$ | $P_{ci}$ | $Pp_i$ | $S_i$ | $Pэ_i$ | $X_i$ | $Z_i$ | $K_i$ |
| Enterprise 1 | 0,9092 | 0,5980 | 2 | 3 | 2,5 | 1,5 | 0,93 | 0,19 | 5 |
| Enterprise 2 | 0,5636 | -0,8658 | 4 | 6 | 5 | 6 | 0,71 | 0,14 | 4 |
| Enterprise 3 | 0,5486 | 0,4938 | 5 | 4 | 4,5 | 5 | 0,76 | 0,15 | 4 |
| Enterprise 4 | 0,7078 | 0,6108 | 3 | 2 | 2,5 | 1,5 | 0,93 | 0,19 | 5 |
| Enterprise 5 | 0,3246 | 0,8767 | 6 | 1 | 3,5 | 4 | 0,81 | 0,16 | 4 |
| Enterprise 6 | 1,2437 | 0,2662 | 1 | 5 | 3 | 3 | 0,86 | 0,17 | 4 |
| Sum | - | - | | | | | 5 | 1 | 25 |

## 5.Scoping of investments for each of the enterprises

Following the direction of application of the calculated indicators of social responsibility and effectiveness of activity of the enterprise is scoping of investments for each of the enterprises (Fig. 3). Let us say that for the enterprises of the food industry of the Kherson area allocation of investments of 50 million hryvnias is planned. The part for 6 studied enterprises makes 2 million UAH (4,62%) of them. Each enterprise gave a request for necessary quantity of means. The total amount of inquiry on all six enterprises exceeds the allocated 2 million UAH. It is necessary to define, how many means to allocate for each concrete enterprise, proceeding from their social responsibility and effectiveness of their activity.

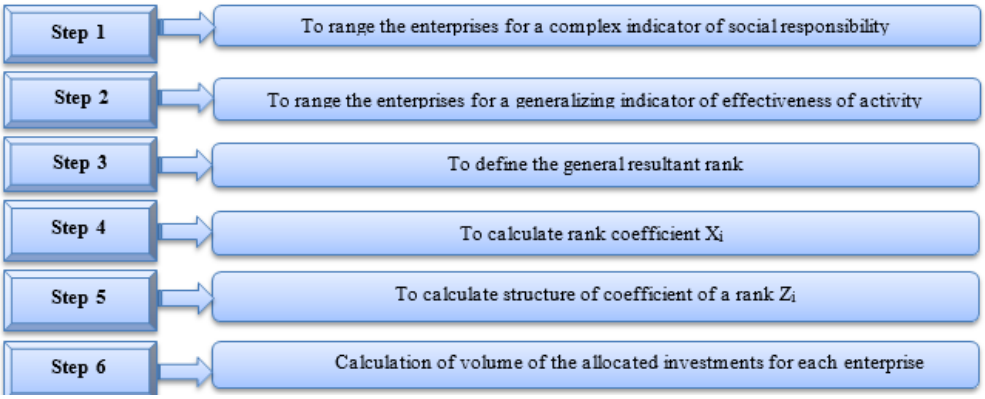



Fig. 3. Calculation of volume of the allocated investments for each enterprise

As a whole the technique completely repeats previous, except for a stage of definition of ranks. Here 1 Rank is appropriated to the enterprise with the greatest indicator of social responsibility and effectiveness of activity, and 6 Rank to the enterprise with the smallest indicator. This results from the fact that in this case it is necessary to define the best enterprise, after all to itself the most part of estimated investments will be allocated. Results of calculations for 2013 are presented in Table 3.

Table 3

Calculation of volume of the allocated investments for each enterprise for 2013

| Enterprises | Indicators | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $C_i$ | $P_i$ | $Pc_i$ | $Pp_i$ | $S_i$ | $Pз_i$ | $X_i$ | $Z_i$ | Request for investments | $I_i$ |
| Enterprise 1 | 0,9092 | 0,5980 | 2 | 3 | 2,5 | 1,5 | 0,93 | 0,19 | 120000 | 371428,5714 |
| Enterprise 2 | 0,5636 | -0,8658 | 4 | 6 | 5 | 6 | 0,71 | 0,14 | 255000 | 285714,2857 |
| Enterprise 3 | 0,5486 | 0,4938 | 5 | 4 | 4,5 | 5 | 0,76 | 0,15 | 550000 | 304761,9048 |
| Enterprise 4 | 0,7078 | 0,6108 | 3 | 2 | 2,5 | 1,5 | 0,93 | 0,19 | 350000 | 371428,5714 |
| Enterprise 5 | 0,3246 | 0,8767 | 6 | 1 | 3,5 | 4 | 0,81 | 0,16 | 750000 | 323809,5238 |
| Enterprise 6 | 1,2437 | 0,2662 | 1 | 5 | 3 | 3 | 0,86 | 0,17 | 100000 | 342857,1429 |
| Sum | - | - | | | | | 5 | 1 | 2125000 | 2000000 |

Therefore, analyzing results presented in Table 3, it is possible to draw a conclusion that the recommended volume of investment is $I_i$. This indicator is calculated on the basis of a complex indicator of social responsibility and a generalizing indicator of effectiveness. The bigger volume of investment should be transferred to the enterprise with the highest level of these indicators, and smaller – with the lowest. The indicator $I_i$ is only recommended and investors can distribute means, leaving also requests of the enterprises for the volume of necessary monetary resources.

## 6.Conclusion

Therefore, the calculated indicators of social responsibility and effectiveness of activity of the enterprise are dependent. It is possible to use them for definition of the best or worst enterprises of the branch with a definite purpose. The author reviewed three examples: calculation of part of the enterprises in total of the income, which are sent to the budget;

calculation of the number of workers, who are subject to reduction on each of the enterprises; calculation of the volume of the allocated investments for each enterprise.

However, it is possible to determine the best enterprises by the offered technique with the following purposes:

1. One of criteria of definition of the best enterprise during carrying out any trend.
2. For decrease in volume of taxes or granting "tax vacation".
3. Development of any infrastructure actual of this branch.
4. At reduction of number of the enterprises of this branch in the region.

For these reasons this is a promising area for future research.

## References

1. Ukraine today. Catalogue of leading enterprises (http://www.rada.com.ua/ukr/catalog/?words=&region=24&sector=17&title=1&body=1&mode=0). [in Ukrainian]
2. Tarasova T.F., Zhosan G.V., Bashkatova of V. Ya. (2014). Valuation of social responsibility of the enterprise: criteria and indicators, *Bulletin of the Belgorod university of cooperation, economy and right: International scientific-theoretical magazine*, Vol. 2, No. 50, PP. 120-127. [in Russian]
3. Korchevska Liliya, Ganna Zhosan, and Sergii Kavun. (2013). Social Responsibility as a Contextual Component of the Enterprise Economic Security, *Journal of Finance and Economics*, 1.4 (2013), PP. 95-104. Available from: http://pubs.sciepub.com/jfe/1/4/6/index.html#.
4. Zhosan G. V. (2014). Calculation of the general indicators of social responsibility of the enterprise for Components, *Zbornik scient. works "Financial and credit activity: theory and practice problems"*, Vol. 1(16)/2014, PP. 288-300. [in Ukrainian]
5. Zhosan G. V. (2013). Methodical aspects of the definition of the built indicator of social responsibility Enterprise, The state and prospects of development of social responsibility of economic subjects in the modern world: materials of the International scientific and practical conference (on November 21-22, 2013), KNU of V. N. Karazin, PP. 15-21. [in Ukrainian]
6. Zhosan G. V., Trukhachova K.V. (2013). Conceptual provisions of social responsibility of the enterprise as a theoretical basis balanced national Khmelnytsky's effectiveness, *Messenger of the university. Scientist magazine. Economic sciences*, Vol. 2, No. 4, PP. 247-254. [in Ukrainian]

# Brand Positioning through Marketig Enterprise Resources

**Olena Shvets**

Kherson national technical university, 73000 Kherson, Ukraine

elen.shvets@mail.ru

**Abstract.** For effective competitive advantage in the market of goods and services brand should be considered in terms of future prospects. The result of positioning is a clear expression in the consumer's mind a positive image of the brand. Clear position in the minds of consumers should be a prime purpose of marketing resources of the enterprise.

**Keywords**. Brand, positioning of a brand, marketing resources, enterprise.

## 1. Introduction

In a general view of positioning acts as a stage of creation of competitive strategy of the enterprises in the market. Based on research of the market, consumers, competitors and the internal corporate analysis, it provides formation of a competitive position of goods or the enterprise, therefore it is based on market research of segments of the market.

Positioning promotes creation of a complex of advance of the company, which provides difference from competitors, compliance to needs of buyers, and also stimulates and directs strategic initiatives, shows values and culture of the organization [3].

## 2. The value of the brand positioning for the company

Positioning is considered as the concept with many opportunities, which is used for the report of the main idea of a brand to consumers through continuous influence on their subconsciousness. From the point of view of marketing, positioning of a brand represents creation of accurate prospect in the target market in relation to competitors. In marketing literature of positioning, it is considered in association with marketing resources of the enterprise.

The vision of positioning is expressed by F. Kotler, who claims that "positioning is a process of creation by the enterprise of goods and services and their image which is based on perception by consumers of competitive goods and services"[5]. It defines positioning of goods as a way according to which consumers identify these or those goods on its major signs.

To create rather strong brand, it is necessary to build positioning of a brand extremely accurately. The main task of positioning to be reduced convincing the consumer of advantages of this trademark. Through positioning, it is possible to inform to wide audience of the purpose of the company, its philosophy and system of values.

In the past figures of the market learned to understand the consumers in the course of daily trade communication with them. However, growth of the sizes of firms and the markets

the manager of marketing of direct contacts with the clients deprived of many. It is necessary managing directors even more often I resorted to research of consumers. They spend more, than ever earlier, for studying of consumers, Trying to find out who exactly buys, how exactly buys when buys, where exactly buys and why buys.

## 3. Simple model of consumer behavior

The enterprise, which has rather understood how consumers react to various characteristics of goods, the prices, advertizing arguments, etc., will have huge advantage before competitors. For this reason, also the enterprises spend so many efforts to research of dependences between incentive factors of marketing and response of consumers. Starting point of all these efforts is the simple model presented in Fig. 1. It is shown that incentive factors of marketing and other irritants get into "a black box" consciousnesses of the buyer and cause certain responses.

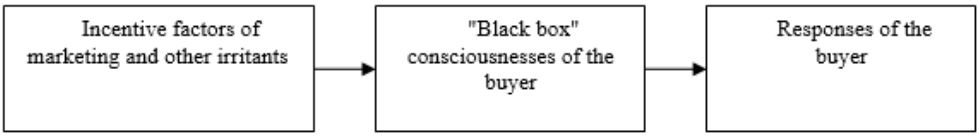



Fig. 1. Simple model of consumer behavior

Incentive factors of marketing include four moments: goods, price, methods of distribution and stimulation. Other irritants develop of the main forces and events from an environment of the buyer; economic, scientific and technical, political and cultural environment. Having passed through "a black box" consciousnesses of the buyer, all these irritants cause a number of the consumer reactions giving in to supervision presented in the right rectangle: goods choice, brand choice, choice of the dealer, choice of time of purchase, choice of volume of purchase.

## 4. Place of Marketing Resources among resource components of the enterprise

Stilvell F. [7] considers that to be global manufacturer in the class in the XXI century, It is necessary: to develop modern products and services quicker today; to introduce new production technologies and processes quicker today; to have possibility of selection, training of workers and heads. Therefore presently at any enterprise, is competitive in the market, a maximum of attention it has to be given to such aspect, as management of marketing resources.

Thus, it is possible to claim that marketing resources - means of any type of the enterprise, allowing to create effective system of creation of ideas and goods (services), their pricing, advance, distributions, and promote increase of image of the enterprise and loyalty of consumers. In other words, it is a certain complex of the interconnected and interdependent means, which are available, directed on production of those goods and/or the services having value for consumers, and are capable to satisfy quite their requirements and/or inquiries on the optimum price, at the same time leads to receiving profit to a producer.

Therefore, it is expedient and lawful to allocate marketing resources in a separate component of the general resource base of the enterprise (Fig. 2).

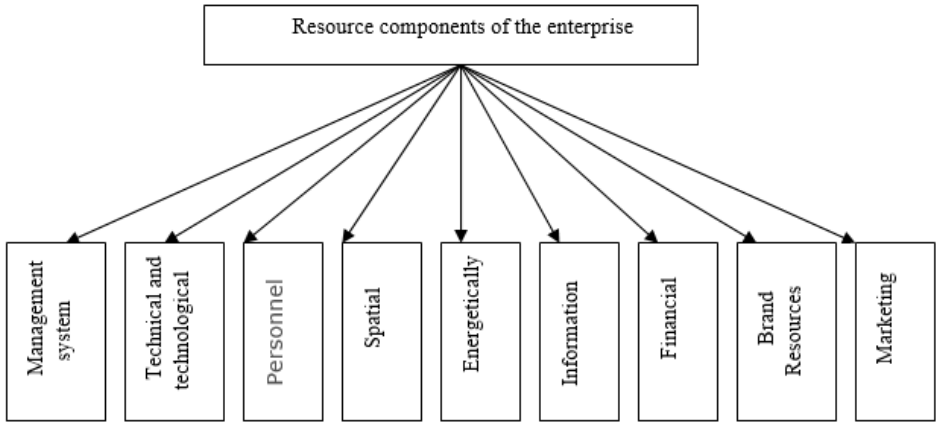



Fig. 2. Place of Marketing Resources among resource components of the enterprise

Resources of the enterprise are characterized by the following making characteristics:

- enterprise management system resources (character and flexibility of operating system, speed of passing of operating influences, etc.);
- technical and technological resources (production equipment, materials, methods of the organization of production and sale, technology, existence of scientific practices, etc.);
- personnel resources (qualification, demographic structure of workers, their ability to adaptation according to the purposes of capacity of the enterprise, enterprise ability, etc.);
- spatial resources (character of production rooms, territories of the enterprise, communications, expansion opportunities, etc.);
- energy resources (is a component of material and material elements of fixed assets: equipment, devices, etc.);
- information resources (information about the capacity of the enterprise and environment, possibility of its expansion and probability increase, etc.);
- financial resources (condition of assets, liquidity, existence of credit lines, etc.) [2];
- brand resources (strengthening of identity of a brand, increase of value of a brand due to unique usefulness and formation of an image for office of a certain complete characteristic).

In turn, the marketing resources include:

- marketing development strategy and program of implementation that meet the goals and objectives of the company;
- possession of a certain part (segment) of the market;
- the base of existing and potential buyers is created;
- existence of the existing and potential business relations (distribution system, the franchize);
- the created image of the enterprise, brand of goods, trade / commodity brand, etc.;
- existence and work with special marketing computer programs and etc.

## 5. Interrelation between marketing resources of the enterprise and brand positioning

Positioning has to be supported by all elements of marketing resources. In addition, definition of a position of a brand in the market requires understanding of how consumers to whom there is an address, determine for themselves the value and on what their choice is based.

Justification of need of implementation of positioning of a brand is based on risk of excessive improvement of the offer for the purpose of increase in its values, and insufficient improvement for the purpose of cut in expenditure.

Before beginning positioning of goods, it is necessary for the enterprise, considering there are marketing resources, to establish a number of the stages interconnected among themselves, namely:

1. Choice of parameters in which it will be carried out positioning. The choice of parameters of positioning is carried out based on the revealed motivation of consumers, elements, which motivate buyers in this target market.
2. Definition of the relevant enterprises, their goods and brands, which are present in this market. In this case all brands, which are present in the market, but only what really compete in this target market are selected not.
3. Drawing up schemes of positioning and perception cards. This stage provides data collection about characteristics of rival brands for the purpose of their further comparison.
4. Analysis of schemes of positioning. After completion of procedure of creation of schemes of positioning, pass to their analysis.
5. Positioning strategy definition. Based on the analysis of schemes of positioning of the company it is necessary to decide on positioning strategy.
6. Select the type of brand positioning. At this stage, it is necessary to decide on type of positioning of brand in the market that as the firm will take root with the brand in consciousness of target consumers. Realization of positioning demands overcoming of three echelons of defense in consciousness of the consumer: selectivity of attention, perception and storing.
7. Positioning realization in system of marketing strategy of the enterprise.

Modern marketing practice offers a wide choice of instruments of advance of production in the market. Powerful from them is an advertizing, (and the most expensive). The purpose of any kind of advertizing - to draw attention of the consumer to production, to show it all its advantages, to inform about the producer**.** Any production, even high quality, needs advertising. Nevertheless, depending on a look and quality of production the format of advertizing can be various: the advertisement in the newspaper or the magazine, a billboard, the booklet, a leaflet, the advertizing material on radio, television.

In system of marketing providing advertizing takes a special place. Its main objectives - demand formation, sales promotion and sale of goods. Advertizing is one of the most effective tools by means of which the enterprises try to regulate behavior of consumers, draw attention to the production and create to itself positive image.

Advertizing is a way of representation of goods in the market and acquaintance of the consumer with attractive qualities of goods, its advantages over similar products of competitors. Consumers will not buy goods of your firm if they do not know anything about it. Advertising - not the only way to inform the consumer on goods. It is a form of

communication with the buyer for his informing and belief. The purpose of any advertizing - to convince the consumer of acquisition of your production, and for this purpose it is necessary to prove what exactly your production, but not other producer, is such which will best of all satisfy his requirement. In order that expenses on an advertizing campaign of justified, the producer has to choose the most effective publicity.

Each aspect of behavior of the buyer is opportunity for advertizing. At a stage of definition, a main goal is drawing attention and providing the buyer with information.

In general, allocate such main stages of advertizing activity, as:

Sales promotion - activities for realization of the commercial and creative ideas stimulating sales of products or services of the advertiser, it is frequent in short terms. In particular, it is used by means of packing of goods on which various tools sales promotion (motivation are located to purchases, network functioning stimulation - a producer, for example, portraits of known people, heroes of the animated films, expensive brands of cars, and also by specialized actions on sale places.

Long-term goal – formation in perception of the consumer of bigger value of firm goods, the certain trademark; the short-term – creation of additional value of goods for the consumer (added value).

Public relations use of editorial part of means of distribution of mass information for the purpose of implementation of the prestigious advertizing directed on a gain of goodwill to commodity families or to firm assumes, them letting out.

Direct marketing – constantly to support the directed communications with certain consumers or the firms having obvious intentions to buy certain goods. Activity in the field of direct marketing is generally carried out by a direct mailing group (direct mail) or through tightly specialized advertizing means.

## 6. Conclusion

When positioning a brand it is necessary to consider perspective tendencies of marketing, characterized by a competition aggravation in the national and international markets, growth of the general awareness and insistence of consumers at a brand choice.

Development of effective competitive positioning is the planned complex of actions, is provided at the enterprise by means of marketing resources. By means of this complex, the analysis of information concerning goods, a sales market, competitors and target consumers who are capable to make the choice for your brand is realized.

Opportunity to be allocated from crowd and to separate from other brands in the market is the purpose of use of marketing resources in positioning of a brand. The competition it becomes constant more fiercely and heavier with the advent of each new brand in the market. Brands can be differentiated on marketing complex elements. The variety of communication channels is extremely important for achievement of consumers while a choice of the channel and its use has to answer with the communication purposes.

## References

1. Aaker D. Strategic market management. 7 edition / David Aaker. – John Wiley & Sons, Inc., 2005.
2. Dobykina DC The potential of the company: the formation and evaluation: teach. guidance. / OK Dobykina, VS Ryzhykov. - K. Centre textbooks, 2007 - 208 p.
3. Pashchuk A. Creating konkurentnoy pozytsyy enterprise of services and этарьı Development Strategy positioning / O. Pashchuk // Otdel marketing. - 2005. - № 4. - S. 44-48.
4. Primak T.O. Positioning strategy in marketing theory / T. Primak // marketing and management innovation. - 2012. - № 1. - S. 13-20.
5. Ries A. The Positioning Era Commeth / Al Ries, Jack Trout // Advertising Age – 1972.
6. Vind J. Positioning. Marketing. Encyclopedia/J. Vind; под.ред. M. Baker – SPb. : St. Petersburg, 2002. – 1196 p. – Pp. 369-376.
7. Stilwell F. Political Economy: The contest of economic ideas / F. Stilwell. – Oxford : University Press, 2006.

# Up-to-Date Global Stock Markets Trends

**Azarenkova Galina[1], Shkodina Iryna[2],**

[1] Kharkiv Institute of banking UB NBU, Kharkiv, Ukraine
azarenkova@khibs.edu.ua
[2] Kharkiv Institute of banking UB NBU, Kharkiv, Ukraine
shkodina@khibs.edu.ua

**Abstract.** The article has studied the dynamics of world trade market testifying, that not taking into consideration its regeneration, the market development is connected with a sharp increase of uncertainty both at developed markets and developing ones. The authors came to the conclusion that the dynamics of different segments of the world stock market depends not upon economic situation in certain countries but upon actions of central banks, in the first place of the FRB, that actively pursue the policy of the quantitative easing. Dissynchronisation of dynamics of different countries' stock markets is the demonstration of the increase of fluctuations at the global level, what leads to weakening of economics growth rate and also changes for the worse financial markets situation which become serious risks to the global economy.

**Keywords:** Stock Market, Quantitative Easing, Volatility, index MSCI.

## 1.Introduction

The present day condition of world trade market (and the whole world economic system) is characterized by strengthening of instability under the conditions of system uncertainty of the institutional dynamics of the global development process. A negative factor of development is a disproportion of the growth among developed countries and developing ones, which may deepen crisis phenomena in the budget, and financial sectors of these countries. Thus, consideration of the tendencies of development of different segments of world stock market is extremely topical.

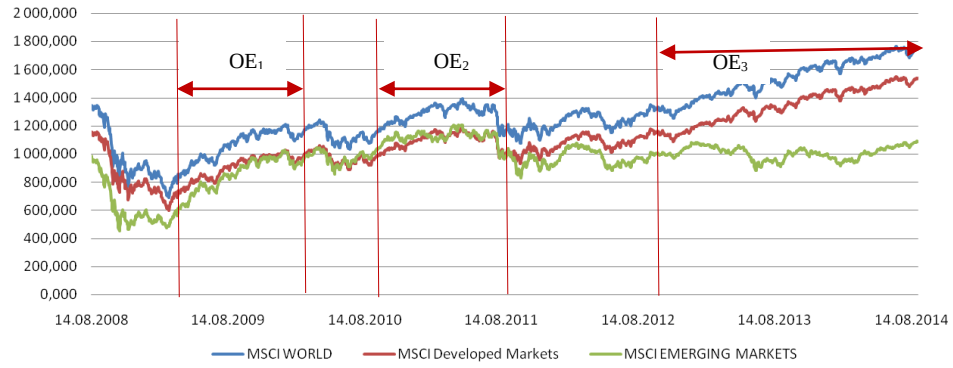
## 2.Current Trends

Notwithstanding the fact that the sharpest phase of the world crisis took place nearly 6 years ago, renewal of the economy rates in the majority of countries in the world are lagging behind the planned ones. As a result of financial pumping up economies black marketing is strengthening and main global disparities deepen, which is responsible for the cause of the crisis. The use of so-called «anticrisis methods» causes a sharp increase of the state debt levels already in a global scale. However, if in the previous decades debt problems were intrinsic in the developing countries today main debt risks are formed and come from developed countries.

Supporting insignificant economic growth of global economy of late is supported by different methods of nontraditional monetary-credit policy applied by the central banks of different countries (in the first place by the USA). One of the most significant is the use of

different programs of «quantitative easing» (QE). Just the same, despite infusing money into economy (money emission, debt monetization, lowering reservation rate, refinancing, etc.) and hoping that, at least a part of this money will be directed to the real sector of economy, economic growth is very moderate and the possibility to provide full employment is absent [1].

One of the main reasons for this is that the excess liquidity does not turn into credits for the real sector of economy and only raises the level of borrowed means usage and financial markets speculations (at the stock market in the first place). Considering the consequences of QE it may be asserted that stock markets grow actively in the time of easing and fall when the program of acquiring assert is over (drawing 1).



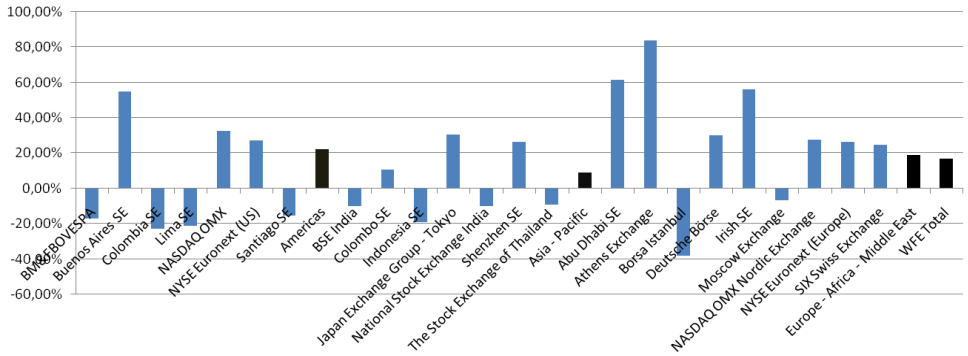

Composed according to MSCI data [5]

Drawing 1. Influence of quantitative easing policy upon the stock market.

Any kind of monetary stimulating introduces a disbalance into the development of some or other sectors of economy. In spite of the fact that main decisions in the policy of quantitative easing were adopted at the end of 2008 renewal of economy growth and productivity rates in the majority of developed countries proceed in lagging behind rates compared to the population growth. Thus, applying these instruments is not enough for providing stable high economic growth. Besides, bearing in mind that short-term interest rates remain close to naught, they have nowhere to fall in order to stimulate a raise of investment spending to provide full employment in the economy.

In addition, carrying out quantitative easing in developed countries leads to excessive capital outflow to the markets of developing countries, since interest rates for state bonds in the USA, Japan, Great Britain, and Germany are at a very low level. The monetary policy of developed countries (the USA in the first place) has a great influence upon South America, Eastern Europe, Middle East and Africa. Index MSCI analysis for developing countries corroborates this: during the first quantitative easing MSCI index for developing countries increased by 80%, during the second quantitative easing the increase was much smaller – 18,7% (that shows the ineffectiveness of solving the problems of stability of the financial system though the methods of financial infusion). In the period between the two infusious index fell by 4,4% (see drawing 1). Consequences of the third quantitative easing are even more pessimistic – if developed stock markets grow the markets of developing ones are characterized by negative tendencies [4; 5]. In 2013, the majority of developed markets

demonstrated profitability higher of average historical meanings. For instance, Japanese stock indices increased by 50 % American ones – by 30%. The developing markets quite the reverse fell down – MSCI Emerging market index decreased by 2 %.




Composed according to the World Federation of Exchanges data [4]
Drawing 2. Profitability of world stock exchanges in 2013

Regional disproportions of stock market development are becoming stronger. If stock markets of American region grew by 22% stock exchanges in Asian-Pacific ocean region only grew by 7% (drawing 2). In the North and South America the growth was brought about by the increase of American stock exchanges indices which grew by 29% at the same time Latin American stock exchange indices fell down by 13%. In the Asian-Pacific Ocean region the highest tempos of growth were observed at Shenzhen (26,3%) and Tokio (30,6%) stock exchanges, while Indonesian (-19%), Indian (-10%) and Thai (-19%) stock exchanges suffered a down fall. Dynamics with different directions is expected in the region Europe-Africa-Middle East: European stock exchanges grew by 2,3% on an average and African and Middle Eastern ones – by 13%. At the same time the most profitable stock exchanges of the world in 2013 happened to be stock exchanges of developing countries - Athenian (84%), Abu Dhabi(+61,4%), Buenos Aires(+55%), Irish (56,1%) [4].

Nevertheless, there exist risks, which the developing countries encounter. They are mainly connected with the capital flows ability to enlarge the problems of home politics. Capital influxes may stimulate inner credits boom and assets prices increase that later on turn into a quick abatement. In the report on international financial stability, the IMF studied capital flows to the developing markets and came to the conclusion that instead of concentrating on the power of a country the flows of «portfolio» investments ever more depend on the general financial situation in the world. Therefore, stock markets of the developing countries become more integrated into international financial expanse becoming one of the main channels of widening financial crises. Even now in a lot of developing countries the signs of overhearing of economy are already observed. External emission of corporative bonds of these countries is the highest during last years. These debt securities are more frequently considered as a substitute for highly profitable corporative USA bonds because they are characterized by a similar market capitalization, lower leverage and higher profit norm with the same credit ratings which makes them more attractive for a wide circle of investors. It may seem to be a positive dynamics as the companies that previously felt lack of credit would gain access to capital now. The risk is in the following if such a tendency goes on then a too large volume of capital may too quickly move to the countries with

forming markets and it will form a risk of fixing the wrong credit interests and an unexpected change of financial flows provided that negative events lead to a sharp reduction of risk inclination.

A quantitative easing consequence may become «the bubbles» creating in different sectors of financial market. Many economists mark forming of potential bubbles. Among them, Nobel prize-winner Robert Shiller who warns that there is an overheating observed at the market. In August of 2014 Shiller P / E ratio (average companies profit from index S & P 500, corrected according to the inflation during last 10 years) was on the level 25, and since 1881, it exceeded this indicator only three times – in 1929, 1999 and 2007.

Especially program $QE_3$ influences real estate market (MBS ransom) and, as is known, it was the poor control of mortgage derivatives that provoked the crisis of 2008. Besides the flow of "cheap" dollars and their considerable volume in circulation reduce risks control and provoke not always justified growth of assets. Janet Yellen notes these moments. The issue of painless bubbles liquidation is an issue not only of professionalism and qualitative building of economic and monetary decisions but also a human factor: greed forces to buy, fear provokes panics. The easiest way is shock and a quick but painful disbalance liquidation. A long way with an unknown end is a search of mutually bound solutions system that has many "mine fields". In her opinion, possibility of the first variant (crash in one or another form) is quite big [2]. According to Lawrence Summers (the former head of American Ministry of Finance) there is a long postcrisis regeneration that is a forerunner of secular stagnation in the majority of leading developed countries of the world. One of the main reasons of the raised stagnation risks is too soft monetary politics of world central banks [7].

At the present moment reduction of monetary stimulation goes on. In July 2014, Federal open market committee decided on reduction of volumes of monthly treasuries purchase (USA State bonds) from $20 milliard to $15 milliard, volumes of securities purchase provided with mortgage – backed securities from $15 milliard to $10 milliard [3].

Analysts of international rating agency Standard&Poor's believe that curtailing the program of quantitative easing is the most dangerous for the developing countries. Curtailing QE will have the strongest effect on China, Russia and Angola, extremely detrimental influence on Ukraine, Georgia, Turkey, Belorussia, Bulgaria, Macedonia and Croatia [6].

## 3.Conclusion

Thus, the analysis of the word stock market testifies that in spite of renovation its development is connected with a sharp increase of uncertainty at both developed markets and developing ones. The dynamics of different segments of world stock market depends not on economic situation in definite countries but on the actions of central banks, first place FRS, which actively pursue in the politics of QE. Financial markets instability is possible. It may occur in the case if US FRS begins to increase interest rates, as the liquidity excess at the markets will reduce.

Desynchronization of dynamics of stock markets of different countries testifies of fluctuation strengthening on the global level, that leads to the weakening of economic increase tempos and also deteriorating the situation at financial markets which become one of serious risks for global economy. Besides the tension connected with Ukraine and financial sanctions of Russia makes the situation worse.

## References

1. Azarenkova G., Shkodina I., Hoyhman M. (2013) Influence policy of quantitative easing in financial market. *Proceedings of the National Bank of Ukraine*, 12, 56 -60 [in Ukrainian].
2. International Monetary Fund. Retrieved from http://www.imf.org.
3. Monetary Policy Releases. July 30 (2014) Retrieved from http://www.federalreserve.gov/newsevents/press/monetary/20140730a.htm.
4. World Federation of Exchanges (WFE). Retrieved from http://www.world-exchanges.org.
5. MSCI Index. Retrieved from http://www.msci.com/products/indexes.
6. Standard & Poor's. Retrieved from http://www.standardandpoors.com.
7. Summers L. (2013) Why stagnation might prove to be the new normal. *Financial Times*, 15. Retrieved from http://www.ft.com/cms/s/2/87cb15ea-5d1a-11e3-a558-00144feabdc0.html#axzz2nik9jJEc.

# Optimization of Bank Investment Lending Innovation Activity in Ukraine on the Basis of the Positive Experience of the European Union Countries

**Snizhana Pashova[1], Inna Vyadrova[2]**

[1] Ph.D. in Economics, Kharkiv Institute of Banking of the University of Banking of the National Bank of Ukraine, Victoty Avenue 54, apt 205, 61202 Kharkiv, Ukraine
Pashova.snezhana@gmail.com

[2] Ph.D. in Economics, Kharkiv Institute of Banking of the University of Banking of the National Bank of Ukraine, Victoty Avenue 55, 61174 Kharkiv, Ukraine

**Abstract.** Is argued the necessary to develop and implement new business models of banking institutions. These models will be focused on new priorities of the economy and banking sector development in Ukraine with the help of practical experience of the European Union. Author grounds on the necessity of involving all commercial banks in Ukraine into investment lending innovation, based on the size of their loan portfolios. In our opinion, it will reveal the overall richness of methodical bases of calculation of innovation and investment component in the loan portfolio of banks. Appropriateness of the choice of innovative projects is substantiated for bank investment lending, given the limited size of the value of innovation and investment component and its size depending on the size of the loan portfolio of a particular bank.

**Keywords.** Investment crediting, non-financial criteria, the Ward method, cluster analysis, ranking projects, innovative development.

## 1.Introduction

Considering the market relationships between the agents of investment and innovation and basing on the continental European model of their organization, it should be emphasized the presence of stable relations between them and the banking sector of the economy. Therefore, both the use of various management decisions on the accumulation and redistribution of 'long money funds' and the adoption of measures to identify the most attractive for investment banking lending innovative projects and programs are defined as the needs of the subjects of innovative activity in the banking lending to the innovation projects.

One of the areas of scholars' researches who were engaged in the improvement of the investment bank lending is the introduction of the optimization of management, the essence of which is to provide a process of bank investment lending favorable characteristics that are in the ratio of the minimum price of borrowed funds, the minimum risk and maximum return.

On the one hand, the banking sector is a separate component of the funding sources of innovation, the development of which requires 'long' financial resources, the accumulation of which is achieved by the use of different instruments of bank accumulation and redistribution of funds between the subjects of market relations. On the other hand, it affects the implementation of state policy innovation and investment by achieving efficiency and

sustainable development of the subjects of innovation through investment lending of innovative projects. Hence, to achieve the identity of strategic guidelines of the state and trends of development of the banking sector of Ukraine in the sphere of support and stimulate innovation is very important task.

That is why the use of the optimization of management in the implementation of bank investment lending innovation activity in Ukraine is of great importance. The use of the management optimization is impossible without the development and implementation of measures to determine the most attractive projects and programs. One of the problems of the Ukrainian banking sector is the divergence of attractiveness criteria for an innovative project. This problem causes the banks to accept a variety of management decisions on the use of financial instruments not only for the implementation of measures to eliminate the problem, but also to meet the needs of the subjects of innovative activity in the banking lending to the innovative projects. As a result, this problem caused the relevance of the topic.

The purpose of the research is to improve the approach of the best size of innovative and investment component in the loan portfolio of banks in Ukraine, using the practice of the European Union; to identify priority characteristics of innovative projects and their ranking criteria to optimize the investment bank lending innovation activity in Ukraine.

## 2.Practical contribution of research

It is necessary to develop and implement new business models of banking institutions. These models will be focused on new priorities of the economy and banking sector development in Ukraine with the help of practical experience of the European Union.

Author grounds on the necessity of involving all commercial banks in Ukraine into investment lending innovation, based on the size of their loan portfolios. In our opinion, it will reveal the overall richness of methodical bases of calculation of innovation and investment component in the loan portfolio of banks, which should be equal to 0.77% of the loan portfolio. This size is not burdensome for commercial banks and adequate for stabilization of economic processes in the country and for the implementation of innovation and investment policy. We suggest that banks' use of this component will promote: partnerships between potential investors , the acquisition of useful experience in syndicated lending; risk diversification between investment agents; the creation of innovation and investment infrastructure; the organization of close business partnerships with innovation and investment agents and consequently - the activation of bank investment innovation lending.

Appropriateness of the choice of innovative projects is substantiated for bank investment lending, given the limited size of the value of innovation and investment component and its size depending on the size of the loan portfolio of a particular bank. Thus, we propose the scientific and methodical approach to the optimization of the investment bank lending innovation activity in Ukraine which will make it possible not only to strengthen the participation of banks in lending to innovation, but also will contribute to the economic and innovative potential of our country.

## 3.Hypothesis

The hypothesis of this research is to determine the component of the banks' loan portfolio, which can be spent on financing innovative projects and programs based on the positive experience of the European Union. It is necessary to identify the most important

financial and non-financial factors, which will determine the effectiveness of using a long part of the loan portfolio of banks in lending to innovative projects and programs.

## 4.Literature review

Many Ukrainian scientists as: A. Baranowski, T. Vasilieva, A. Vasyurenko, I. Vyadrova, N. Zinko, L. Kloba, A. Kuznetsova, K. Kuklik, S. Leonov, S. Onishko, B. Pshick, T. Smovzhenko and foreign scientists: P. Drucker, W. Sihurt were been investigated the banking investment lending and financing of innovation activity.

Government regulation has very important value for supporting innovation and investment, because innovative model of development is the most expensive among all of them and requires long-term investments. Traditional issues of state regulation of innovation and investment activities were described by such scientists as : A. Booth, A. Kalina, Lapko, M. Tugan-Baranovsky, A. Cherep and foreign scientists: P. Drucker, M. Kalecki, R. Landau, G. Mensch, B. Sihurt.

Problems of innovation and investment restructuring of Ukraine's economy, the development of organizational and economic model of the national innovation system, mechanisms for innovative transformation in the real sector of the economy devoted to the works of Ukrainian scientists V. Alexandrova, V. Heyets and foreign researchers: M. Begovic, N. Kondratiev, G. Mensch, J. Schumpeter.

T. Vasilyeva in "Banking Investment on innovation market" [1] assumes that the importance of investing in innovation consists of acceleration of the renewal of fixed and working capital, the production technology in order to achieve a positive effect in the form of profits. The reason is that innovations affecting on various aspects of national development, determine country's place in the global economy, internal socio-economic status and etc..

B. Rysin in "The formation and placement of long-term resources of banks in Ukraine" defines the investment credit as one of the "varieties of long-term debt associated with lending to direct investment activities or real investment (fixed capital formation and growth of the material - inventories)" [3, p. 46]. We assume that author does not take into account the high profitability of this type of lending.

A. Vovchak in his "Credit and Banking" treats loans in investment activity as loans for borrowers to meet their temporary needs for funds when making investments. However, investments are defined as a part of the financial resources in the form of their long-term investments [2, p. 275-278]. The drawback of these definitions shows that the author does not emphasize high profitability and risks. Agreeing with A. Vovchak "Investment" [2], we believe that today the development of the Ukrainian economy is connected with investing in innovation, with the timely and successful financing of strategic importance for the state and society projects and programs.

## 5.Methodology

Statistical database of the research consists of the European Innovation Scoreboard indicators and indicators of banking financing of innovation processes in the member countries of the Continental European model of development. In order to determine the distance of Ukraine from the general European rates, which are the average index of innovation of European group, we use the following formula (1) (Table 1):

$$S_n = \sum_{i=1}^{n} \sqrt{(Z_{ec} - \frac{x_i - \frac{1}{n}\sum_{i=1}^{n} x_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)^2}})^2}, \quad (1)$$

where $S_n$ – the distance of the country from the general European rate; $Z_{ec}$ – general European normalized value; $x_i$ – sample value; $\frac{1}{n}\sum_{i=1}^{n} x_i$ – arithmetical mean; $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i)^2}$ – standard deviation; $n$ – sample size.

Table 1

**Indicators of innovation development of Ukraine**

| № | Countries | | Distance of innovative development from the standard, % | | | Productivity increase of innovation activities, % | | | The bank financing share of enterprises in the development of innovations, % | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Short name | Full name | 2007 | 2010 | 2013 | 2007 | 2010 | 2013 | 2007 | 2010 | 2013 |
| 1 | UA | Ukraine | 6,37 | 5,5 | 1,49 | 0,45 | 0,44 | 2,85 | 0,48 | 0,46 | 0,6 |
| … | … | … | … | … | … | … | … | … | … | … | … |
| 33 | EU | The average European rate | 0–4 | 0–3,5 | 0–1,2 | 0,45 | 0,48 | 0,85 | 1,2 | 1,42 | 1,36 |

Calculated by the author with the use of data [6, c. 237 – 242; 7]

We perform cluster grouping of the data using the software Statgraphics Centurion XV. Thus, using the Ward method, we designed matrix of distances between clusters on which (Table 2) conducted the group observing similar characteristics.

Table 2

**Matrix of distances between clusters to perform grouping by the similar features of observations**

| Cluster | Centroid of the first magnitude | | | Centroid of the second magnitude | | | The number of members in each cluster | | | Centroid of the third magnitude | | | The number of members in each cluster | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2007 | 2010 | 2013 | 2007 | 2010 | 2013 | 2007 | 2010 | 2013 | 2007 | 2010 | 2013 | 2007 | 2010 | 2013 |
| 1 | 2,621 | 0,531 | 1,358 | 0,496 | 0,531 | 2,015 | 8 | 8 | 9 | 0,471 | 0,507 | 0,671 | 9 | 6 | 5 |
| 2 | 5,542 | 0,377 | 0,835 | 0,356 | 0,377 | 2,097 | 5 | 7 | 4 | 1,135 | 1,209 | 1,441 | 8 | 7 | 4 |
| 3 | 2,194 | 0,47 | 1,334 | 0,664 | 0,47 | 0,206 | 5 | 5 | 5 | 1,667 | 1,913 | 0,927 | 7 | 5 | 6 |
| 4 | 4,662 | 0,293 | 1,033 | 0,262 | 0,293 | 3,26 | 5 | 6 | 6 | 0,181 | 0,934 | 2,103 | 4 | 6 | 7 |
| 5 | 3,09 | 0,265 | 1,114 | 0,35 | 0,267 | 6,798 | 2 | 4 | 5 | 2,34 | 2,455 | 0,264 | 3 | 7 | 7 |
| 6 | 5,848 | 0,643 | 1,785 | 0,21 | 0,643 | 3,245 | 8 | 3 | 4 | 2,987 | 3,207 | 2,871 | 2 | 2 | 4 |

Calculated by the author.

Using the cluster analysis results and the previously calculated data of innovative development of the investigated countries, we graphically depict the place of Ukraine in the world innovative area of the investment lending in 2013 (with the use of Microsoft Excel software) (Figure 1).

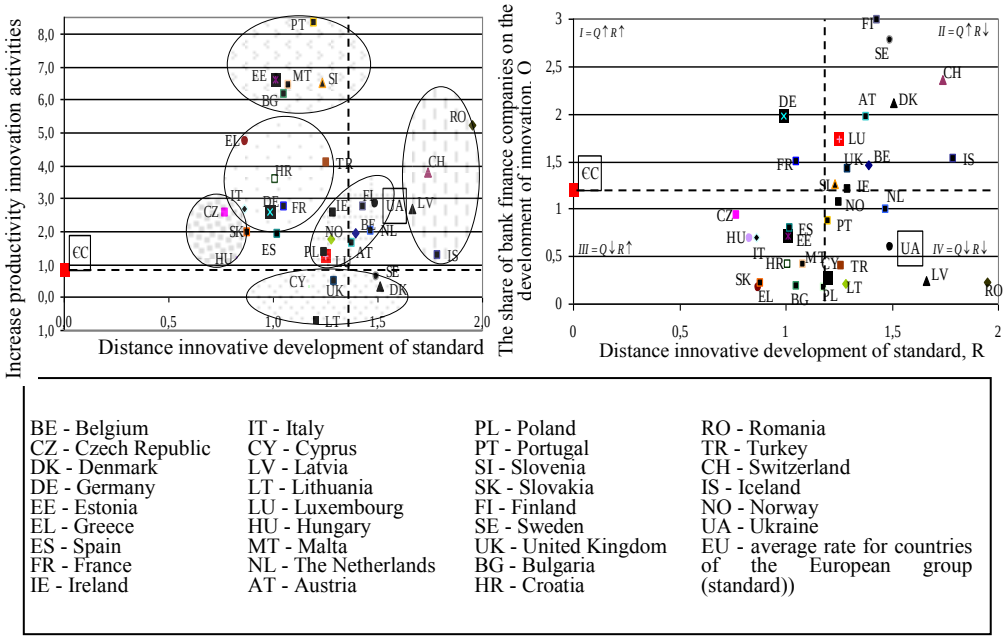



Figure 1. Characteristics of global investment lending practice of innovative development in 2013

The dotted horizontal lines show the overall rate of productivity increase of innovative activity in the European Union and the rate of the companies' bank funding for the development of innovations. According to the calculations, the rates were equal to 0,85 % in 2013 and 1.36% respectively. The vertical dotted line depicts standard value of innovative development in the European Union. Based on the conducted calculations it had limits from 0 to 1,2 % in 2013.

Concerning the standard, the countries are divided into four groups. It means that we are interested only in those countries which match to the standard values and are within their limits, namely the countries of sector I = Q ↑ R ↑ : Germany and France.

The Countries of Sector II = Q ↓ R ↑ - possess a bank innovation financing rate higher than the standard one, and the rate of innovation is outside the standard value, i.e. the distance of the investigated countries from the standard value is longer than the permissible limits specified.

The Countries of Sector III = Q ↑ R ↓ - possess a bank innovation financing rate lower than the standard one, and the rate of innovation is within the reference value.

The Countries of Sector IV = Q ↓ R ↓ - have as the lower measure of bank innovation

financing than the standard and the indicator of innovative development out of limits of the standard value, i.e. the distance of the surveyed countries from the standard value is longer than the acceptable limits defined. Unfortunately, Ukraine locates exactly in this sector.

Methodological approach has allowed setting up the countries, which occupy pioneering positions in bank innovative lending, and the place of Ukraine in the global innovative process. According to the trends of the practical continental European model of innovation implementation, we have reason to believe that the best loan and investment potential is applied by two countries: France and Germany.

To realize the economic and innovative capacity of our country, the effective model of innovation and investment lending must be launched. This model should correspond the European standards and will draw Ukraine closer to the average value. Relying on the above-mentioned foreign experience, we reasonably consider involving all Ukrainian banks to the innovation financing process. This could be achieved by the implementation of the innovative investment part into the bank loan portfolios. Schematic visualization is presented in Figure 2.

Estimated value of the innovative investment share size in the loan portfolio of the bank is 0.77 % (1 465 400 000 U.S. dollars). Today, the banking system of Ukraine takes part in providing financial innovation development only at the level of 0.34 % (646.5 million dollars). Thus, to cover the existing gap between the estimated and actual values, it is necessary to take measures for the implementation of innovative and investment share into the Bank's loan portfolio.

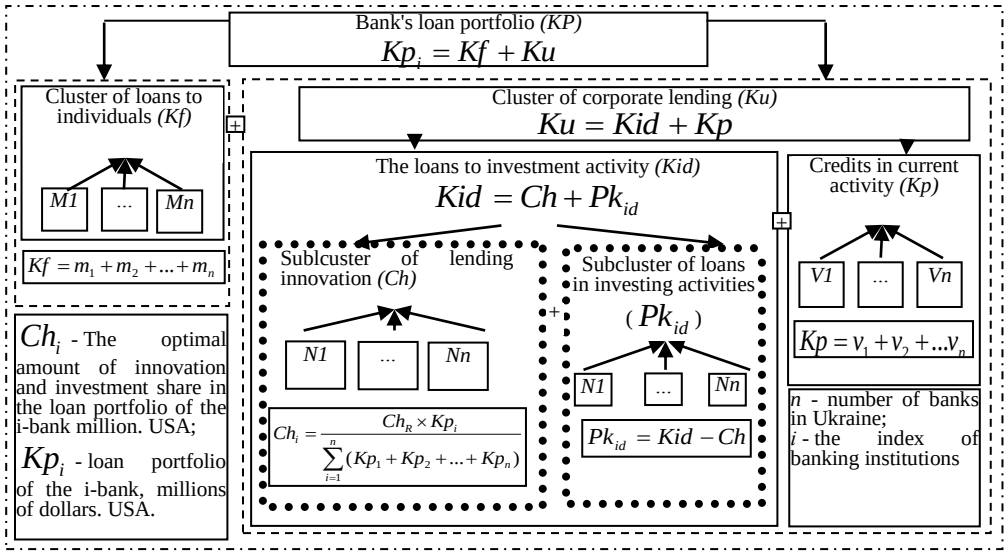



Figure 2. Formation process of the innovative investment component in the loan portfolio of the bank.

In our view, while implementing the investment bank lending innovation activities, the innovative projects of highest priority must include the following guidelines:

- Criteria for financial and non-financial project.

With the help of experts and by means of an expert survey, we have identified the most common financial and non-financial criteria, which are applied by banks when lending in Ukraine. The priority of the criteria was evaluated as a 10-point scale. The survey

involved Ukrainian banks that operate in the domestic market and are part of the 26 largest banks in terms of equity.

The survey results led to the conclusion that the financial indicators which are often used in decision-making on investment lending innovation and the most significant effect on the adoption of such a decision, are indicators for assessing the financial state of the borrower, according to the financial statements and efficiency indicators of the innovative project. The non-financial indicators are the credit store existence, the business proposal, the project risks, the markets of the innovative products (Table 3). The results of the ranking of projects in terms of quality.

Table 3

**Priority of the Financial and non-financial criteria when deciding on the possibility of investment lending innovation and the power of their influence on this decision**

| The evaluation of the financial criteria | | | The evaluation of the non-financial criteria | | |
|---|---|---|---|---|---|
| Criteria | Frequency* | Points** | Criteria | Frequency* | Points** |
| The financial state of borrowers according to the financial statements:<br>- Ability to pay;<br>- Financial stability;<br>- Revenues from sales of products | Compulsory<br>Compulsory<br>Compulsory<br>Compulsory | 10<br>10<br>10<br>10 | A credit history existence | Frequent | 8,4 |
| | | | Amount of previous payments | Frequent | 6,1 |
| | | | Business proposal and technical and economic base of the project | Compulsory | 8,6 |
| Gains and losses in the last 3 years | Compulsory | 8,8 | | | |
| Turnover on accounts | Frequent | 8,5 | The influence of geographical and sectoral factors:<br>- Residency;<br>- Risk area;<br>- Prospects of the industry;<br>- The seasonality of production or distribution of products | Frequent<br>Frequent<br>Frequent<br>Frequent<br>Frequent | 5,7<br>5,6<br>5,8<br>5,4<br>5,8 |
| State and dynamics of receivables and payables | Frequent | 8,5 | | | |
| The presence of outstanding accounts receivable and accounts payable | Frequent | 8,7 | | | |
| Profitability of manufacture | Frequent | 8,5 | The competitive environment | Frequent | 8,6 |
| The share of investment credit in the value of the innovative project | Frequent | 8,3 | The duration of the company in the market | Frequent | 7,0 |
| | | | Business quality management | Frequent | 6,6 |
| Performance indicators of investment and / or innovative projects:<br>- Internal rate of return<br>- Net present value ;<br>- The index of profitability;<br>- PBP | Compulsory<br>Compulsory<br>Compulsory<br>Compulsory | 9,2<br>9,6<br>9,3<br>9,4<br>9,6 | Experience of similar projects | Rare | 2,4 |
| | | | Image and popularity Company | Rare | 3,4 |
| | | | The presence of the brand (known or unknown) | Rare | 3,1 |
| | | | Markets | Compulsory | 8,7 |
| | | | The protection of rights to an innovative product (Patents) | Frequent | 5,6 |

| | | | | | |
|---|---|---|---|---|---|
| The scale of investment lending innovation project , the amount of interest payable with respect to equity | Frequent | 8,4 | Consistency of development strategy and product innovation | Frequent | 5,9 |
| | | | Target Audience | Frequent | 7,2 |
| | | | Project risks | Compulsory | 8,8 |
| Expected performance of innovative investment lending project profit and loss forecasts , cash flow , profitability and solvency | Frequent | 8,3 | Life cycle stages of the innovation project, the availability of the first industrial design | Frequent | 7,6 |
| | | | The growth rate of the target market | Frequent | 6,0 |
| | | | The value of the project to the market (stimulating or not) | Rare | 5,2 |
| | | | The ability of the project to create a new market | Rare | 2,2 |

* - determining the frequency of use criterion in evaluating a loan application; ** - Scoring potency of each criterion in decisions on investment lending innovation activities

We should point out the fact that applying the below methods of ranging projects with the use of quality characteristics allows to determine the maximum rating value and to choose the most attractive for investment banking lending innovative projects (Table 4).

Table 4

**Ranking projects for quality characteristics**

| Group performance | Innovative projects | | | | | |
|---|---|---|---|---|---|---|
| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | … | $X_n$ |
| $A_{X_1 \ldots X_n}$ | $X_{1a}$ | $X_{2a}$ | $X_{3a}$ | $X_{4a}$ | … | $X_{na}$ |
| $B_{X_1 \ldots X_n}$ | $X_{1b}$ | $X_{2b}$ | $X_{3b}$ | $X_{4b}$ | … | $X_{nb}$ |
| $C_{X_1 \ldots X_n}$ | $X_{1c}$ | $X_{2c}$ | $X_{3c}$ | $X_{4c}$ | … | $X_{nc}$ |
| … | … | … | … | … | … | … |
| $M_{X_1 \ldots X_n}$ | $X_{1m}$ | $X_{2m}$ | $X_{3m}$ | $X_{4m}$ | … | $X_{nm}$ |
| Number of points | $\sum_{i=1}^{m} X_{1_i}$ | $\sum_{i=1}^{m} X_{2_i}$ | $\sum_{i=1}^{m} X_{3_i}$ | $\sum_{i=1}^{m} X_{4_i}$ | … | $\sum_{i=1}^{m} X_{n_i}$ |

Source: own elaboration

In our opinion, the ranking of projects in terms of quality simplify the banks' choice of investment project. It helps to save time for decision-making, to enhance the rapid response to changes in the conditions of the project, to reduce bank-operating costs etc.

- Appeared problems in the ranking of alternative projects. Assessing projects, which are similar or alternative to its beneficial effect, the bank may face a situation when despite the ranking it is difficult to choose the right one.

In this case, we assume to apply the alternative assessment procedure. We believe that a properly chosen method of evaluation of the innovative project will allow the bank in the future to maximize the economic result of investment lending.

To draw the conclusion one can say that these guidelines allow the bank to justify the feasibility of investment lending innovation and make a rational allocation of scarce typically "long" investment resources and/or the existing innovation and investment component in the loan portfolio of the bank [6].

Due to the difficult political and economic situation in Ukraine, the need for innovation has increased rapidly. For this reason, we believe that it is highly important to examine and determine the influence of the innovative part in banking loan portfolio on financial independence and stability of Ukraine.

## 6.Conclusion

To draw the conclusion one can say that these guidelines allow the bank to justify the feasibility of investment lending innovation and make a rational allocation of scarce typically "long" investment resources and/or the existing innovation and investment component in the loan portfolio of the bank [6].

Due to the difficult political and economic situation in Ukraine, the need for innovation has increased rapidly. For this reason, we believe that it is highly important to examine and determine the influence of the innovative part in banking loan portfolio on financial independence and stability of Ukraine.

## References

1. Vasilieva T. “Banking investment in the innovation market”: monograph / T. Vasilieva. Sumy, 2007. – 513 p.
2. Vovchak O. “Credit and Banking”: textbook / O. Vovchak, N. Rushchyshyn, T. Andreikiv, Kyiv, 2008. – 564 p.
3. Voinovich A. “Support of innovation development in the regions of the USA: the experience of small companies and research centers” / A. Voinovich, V. Soloviev // Science problems. - 2000. - № 2 - P. 28.
4. Rysin V. “Formation and placement of long-term resources of banks in Ukraine”: monograph / V. Rysin, M. Rysin, Kyiv, 2008. – 234 p.
5. Pshik B. “Financial investment of banks”: teach. Guidances / B. Pshik, O. Sarahman, N. Drebot, Kyiv, 2008. – 194 p.
6. Peresada A. “Management of bank investments”: monograph / A. Peresada, T. Mayorova, Kyiv, 2005. – 338 p.
7. Analysis of innovation drivers and barriers in support of better policies // Economic and Market Intelligence on Innovation. - 2013 - 15 th November. - Vienna. - P. 247.
8. Central Intelligence Agency [electronic resource] : The World Factbook. - Mode of access: www.cia.gov.

# Methodological Approaches of Assessment and Statistical Analysis of the Energy Security of Ukraine

**Oleksandr Diachuk[1], Roman Podolets[2]**

[1] Institute for Economics and Forecasting Ukrainian Academy of Science, 01011 Kyiv, Ukraine
diachuk@ief.org.ua

[2] Institute for Economics and Forecasting Ukrainian Academy of Science, 01011 Kyiv, Ukraine
podolets@ief.org.ua

**Abstract.** The article describes the methodological approaches to assessment and analysis of energy security of Ukraine, which can be applied for other countries. This approaches based on the evaluation and analysis of the economy security introduced by the Ministry of economic development and trade of Ukraine and statistics databases of the International Energy Agency.

**Keywords.** energy security, indicators, emissions intensity, energy intensity, energy balance.

## 1.Introduction

National energy sector, its state, opportunities and development prospects, as well as efficiency of national energy policy implementation, significantly influence the economic state in the country and conditions of its development. General integration processes of national economy and energy into the world and European structures considerably raise the requirements for the national energy security (ENS) ensuring.

Among the urgent problems, which need to be solved in any country, there are problems of meeting the economy and population needs in energy resources, which have crucial meaning and are the main problems of ensuring energy security. Today these problems for the vast majority of the countries are not only unsettled, but on the contrary, acquire exceptional value importance and become one of the main arguments in the state decision making process.

While conducting research on national energy security improvement, it is necessary to clearly identify both: the definition of "energy security" term and its components, as well as methodological issues of their evaluation and analysis.

## 2.National energy security components identification

It is widely accepted that energy security is one of the main components of economic security. Consequently the concept of operations of the executive authorities in the process of energy security ensuring [1] states that energy security of Ukraine is an integral part of national and economic security and a necessary condition of state's existence and development. "The basic ways of national policy in the field of energy security of Ukraine" [2] states that "energy security is one of the key components of national security, a necessary condition of the national sustainable development implementation".

Similarly the role of energy security is defined in works and documents [3-7]. Historical analysis of the existing scientific-technical approaches to the understanding and

evaluation of the energy security level is presented in work [8], in general form all aspects of ENS ensuring are reflected in the definition of energy security in work [9], where the main components of the ENS are energy supply, energy independence, environmental acceptability and social stability in the area.

According to the definitions of the ENS given in the works and documents [1-9] energy security – is the ability of the state to ensure the most reliable, technically safe, environmentally acceptable and reasonably sufficient energy supply of economy and population, as well as guaranteed state capability to administrate the process of development and implementation of the national interests security policy in the energy field without excessive external and internal pressure in the current and projected conditions.

Furthermore the energy supply factor characterizes technical and economic aspect of the ENS ensuring. The energy independence factor characterizes the level of state capability to administrate the process of development and implementation of the national interests security policy, defined by the maximum possible level of self-sufficiency by the main energy resources (ER) (otherwise the state dependence in the external energy supply can be a factor of interference and pressure).

Ecological acceptability production factor characterizes acceptable level (on the part of state and world community) of the pollutant emissions in the process of manufacturing and consuming of the energy resources. Social stability factor characterizes the availability level of the ER to all social groups of population and industry sector and to a greater extent connected to the pricing characteristics of the main ER and their availability level. In the given definition of the ENS such an important aspect of safety ensuring, as the dominating role of the state in the process of development and implementation of the national interests security policy.

## 3.Overview of Ukraine's current energy security state

To evaluate the energy security state the indicative is used in most of works. Its main principle lies in the process of creating of the metrics and indicators, which reflect the state and dynamics of change of the system in case of influence of certain threats to the country.

According to the *Technique for evaluation and analysis of the economy security, introduced by the Ministry of economic development and trade of Ukraine* [10] calculation of the energy security level index, as one of the aspects of the national energy security is conducted using two methods. According to the first method indexes are being compared, which define the national economic security with optimal value. According to the second method indexes are being evaluated depending in which range (optimal, threshold or limit value). The integrated index of the energy security level is the average value, calculated using the 1st and 2nd methods (table 1).

*Table 1*

**Indicators and their threshold value of the energy security state**

| № | Indicator, computing unit | 1 method | 2 method* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | OPT | LL | LT | LR | UN | UT | UP | WC |
| 1 | Oil import ratio from one country (company) total value, % | < 30 | 15 | 20 | 25 | 30 | 40 | 50 | 0.17 |
| 2 | Fixed capital assets depreciation in FEC, % | < 50 | 15 | 25 | 40 | 50 | 55 | 70 | 0.11 |
| 3 | Own resources part at the state balance of FER, % | > 50 | 30 | 40 | 50 | 60 | 70 | 100 | 0.10 |
| 4 | Part of the dominant power resource in FER consumption, % | < 30 | 10 | 20 | 30 | 40 | 50 | 60 | 0.10 |
| 5 | GPD energy intensity, kce/uah | 0,2-0,5 | 0 | 0.1 | 0.2 | 0.5 | 0.7 | 1 | 0.10 |
| | Load ratio of the transit oil and gas pipeline system: | | | | | | | | |
| 6 | Oil transit mil tons | 55-65 | 27 | 40 | 54 | 58 | 63 | 68 | 0.09 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 7 | Gas transit, billion cubic meters | > 175 | 80 | 110 | 154 | 175 | 180 | 190 | 0.09 |
| 8 | Degree of coverage FER, % | > 100 | 70 | 90 | 100 | 100 | 140 | 150 | 0.08 |
| 9 | Volume of the coal extraction, mil tons | 70-100 | 40 | 50 | 70 | 100 | 110 | 120 | 0.08 |
| 10 | Ratio of the investments to FEC companies to GDP, % | 3-4 | 2 | 2.5 | 3 | 4 | 6 | 10 | 0.08 |

**Used following abbreviations: OPT – optimal value, LL — lower limit, LT — lower threshold, LR — lower rate, UN — upper norm, UT — upper threshold, UL — upper limit, WC — weight coefficient.*

Main positive aspects of this method include based on it quantitative calculations of economic and energy security level for the last years and determination of weight coefficient (table 2) as separate safety metrics and indicators as well as separate blocks of the economic security (to which ENS is included).

*Table 2*

**Dynamics of the energy security indicators**

| Indicator | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|---|---|
| Own source interest in the FER balance of the state, % | 49.5 | 53.2 | 50.8 | 50.7 | 66.6 | 57.9 | 64.3 |
| Dominating part of power source consumption of the FER, % | 40.0 | 42.4 | 41.5 | 41.1 | 37.9 | 42.6 | 41.5 |
| Fuel import ratio from one country (company) total value of import , % | 68.5 | 54.6 | 49.3 | 38.0 | 56.0 | 75.5 | 68.9 |
| Ratio of the investments to FEC companies to GDP, % | 1.0 | 1.0 | 1.0 | 1.1 | 0.9 | 0.8 | 1.6 |
| GPD energy intensity, kg of fuel equivalent/uah | 0.730 | 0.710 | 0.670 | 0.635 | 0.650 | 0.640 | 0.633 |
| Volume of the coil extraction, mil t. | 60.4 | 61.6 | 58.8 | 59.3 | 54.8 | 54.4 | 61.8 |
| Gas transit, billion. м$^3$ | 133.6 | 128.5 | 115.2 | 119.6 | 95.2 | 98.6 | 104.2 |
| Degree of coverage FER, % | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Fixed capital assets depreciation in FEC, % | 60.6 | 60.9 | 62.2 | 62.0 | 62.2 | 62.2 | 62.2 |
| Oil transit, mil t. | 31.4 | 33.2 | 39.8 | 32.8 | 29.1 | 20.1 | 17.8 |

Thus the authors' opinion of the Method beginning from the 2003 the ENS level was within the interval limits of 50-80% keeping within the risk area. The ENS index level showed the stable growth during 2003-2008 with historical peak fixed in 2008 (figure 1).

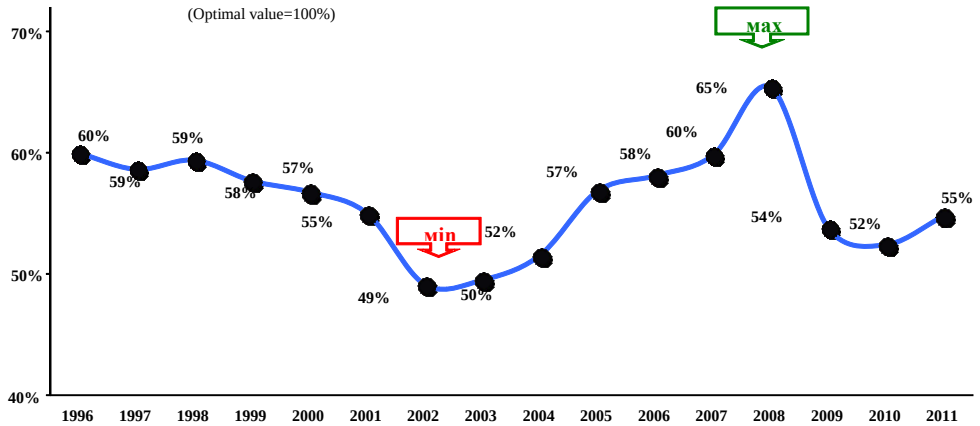



**Figure 1.** Dynamics of the energy security indicator according to the calculation of the Ministry of Economic Development and Trade of Ukraine

According to this Method the frequency of indicator system revision and their proper value due to the fact of the national and global economy changes is done if necessary, but not less than once every two years. Therefore, currently the possibility of introducing changes to

the Method is being discussed, particularly it is being planned to update the indicator list of the energy security with the given figures in the table 3 below.

*Table 3*

**Addition to the indicators list of the energy security**

| Indexes | Optimal value |
|---|---|
| Ratio of the energy import in general import structure | < 20% |
| Capacity utilization level of the crude oil distillation | 90-100% |
| The ratio of the gas export volume of Russian Federation to gas transit through the territory of Ukraine | 50% |
| Average annual cost of the imported gas | $250 per t cub. m |
| Average cost of the imported oil | $80 per barrel |
| Fuel and energy resources supply per person | < 100% of world average value |
| Domestic household customer share in general structure of the consumption use of the fuel and energy resources | less than 20% |

*Statistical base and assessment methodology of the energy security by the International Energy Agency (IEA).* To overcome the energy crisis of the early 70's of the last century the energy statistic base of IEA was created, which is constantly being improved and with the help of which, particularly, basic energy indicators of the world, regions and separate countries are being formed. Thus, in paper [11] it is noted that the ENS indicators, which are calculated based on the IEA statistic data, are the analytic tool of the national energy security evaluation, and give the possibility to analyze and understand the reasons of changes, that took place in sectors and types of economic activity.

The indicative analysis methodology of the IEA is recognized all over the world, which allows carry out proper comparisons between countries and is consistent argument during international climate change negotiations.

The analysis of the energy security condition of the economically developed and other countries of the world using the IEA methodology is reviewed in works [11]. Reviews of the Ukraine's energy policy, in which being carried out a detailed analysis of the energy development and its separate sectors based on statistic data of the IEA, are given in works [12, 13], however they are without certain separation and most important, without quantitative evaluation of the ENS condition of Ukraine. Therefore authors of this paper give below analysis of the main indicators of the current energy security state of Ukraine according to the IEA data.

One of the key indicators of the efficiency level of the ER consumption, which considerably determines the amount of the FER needed for guaranteeing of a certain acceptable and potentially possible economic development level is the national energy intensity of GDP (for more objective comparison of the given indicator is used national level value of the GDP taking into account purchasing power parity (PPP), which is determined by the actual price level.

According to the statistic data of the IEA [14] Ukrainian level of energy intensity of GDP (2005 PPP) in 2010 was 0,47 toe/1000$ (2005 PPP), while corresponding index for the world – 0,19, and for the countries of the OECD – 0,15 (table 4).

The dynamics of change for the last years in the energy intensity of GDP (PPP) of Ukraine in comparison with other similar indexes of major countries and regions of the world and limit levels is given in figure 2.

*Table 4*

**Indexes of efficiency level of the energy resources consumption, toe/1000$ (2000, 2005 PPP) ****

| Country/region | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| World | 0.22 | 0.22 | 0.22 | 0.21 | 0.21 | 0.21 | 0.20 | 0.20 | 0.19 | 0.19 | 0.19 |
| OECD | 0.20 | 0.20 | 0.19 | 0.19 | 0.19 | 0.18 | 0.18 | 0.17 | 0.16 | 0.16 | 0.15 |
| Germany | 0.16 | 0.17 | 0.17 | 0.16 | 0.16 | 0.16 | 0.15 | 0.14 | 0.14 | 0.14 | 0.12 |
| Poland | 0.22 | 0.24 | 0.22 | 0.21 | 0.20 | 0.20 | 0.20 | 0.18 | 0.17 | 0.16 | 0.15 |
| Russia | n/a | 0.62 | 0.54 | 0.51 | 0.49 | 0.47 | 0.46 | 0.42 | 0.42 | 0.42 | 0.35 |
| **Ukraine** | **0.74** | **0.66** | **0.57** | **0.53** | **0.50** | **0.50** | **0.45** | **0.41** | **0.40** | **0.40** | **0.47** |
| Great Britain | 0.16 | 0.17 | 0.15 | 0.15 | 0.14 | 0.14 | 0.13 | 0.12 | 0.11 | 0.11 | 0.10 |
| USA | 0.24 | 0.23 | 0.23 | 0.22 | 0.22 | 0.21 | 0.21 | 0.20 | 0.19 | 0.19 | 0.17 |

*** For 2000-2009 the used data set is taking into account PPP of 2000, as for 2010 – PPP of 2005.*

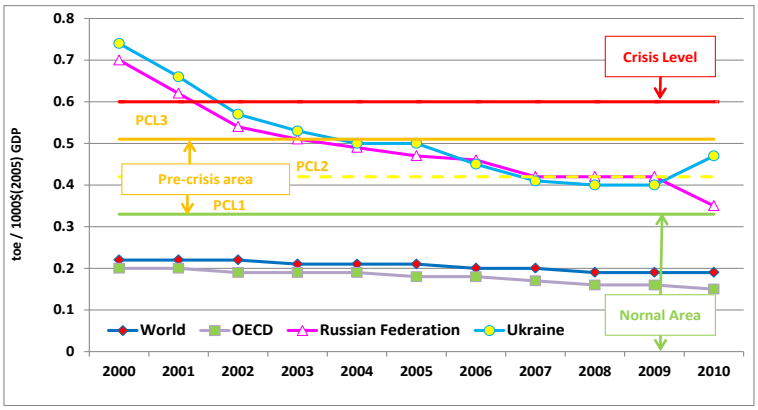



**Figure 2.** Dynamics of GDP (PPP) energy intensity of Ukraine, Russia, world and OECD countries and the crisis levels of given index

As it can be seen from the table 4 and figure 2, the world energy resources consumption efficiency index during 2008-2010 has not changed and remained at the level of 0.19 toe/1000$ (2005 PPP), however the economically developed countries index has decreased at the beginning of the XXI century. At the beginning of the XXI century Ukraine also had positive dynamic of reducing GDP power intensity, which greatly outran the rate of reduction in the OECD countries, however in 2010 this index rose by 17.5% in comparison to the previous year, which is almost 2.5 times higher than world index, while Russia and Belarus managed to decrease it by 16.5 and 20.5% accordingly.

In the work [15] ranking of the GDP energy intensity in three ranges is suggested. In case if GDP energy intensity lies in the range from 0.09 toe/1000$ to 0.33 toe/1000$ it can be characterized as normal area, in the range from 0.33 toe/1000$ to 0.6 toe/1000$ — pre-crisis area, which in turn is divided into three stages: pre-crisis level 1 (threshold stage) — 0.33-0.42 toe/1000$ ; pre-crisis level 2 (development stage) — 0.42-0.51 toe/1000$ ; pre-crisis level 3 (critical stage) — 0,51–0,60 toe/1000$ . Energy intensity greater than 0.6 toe/1000$ is considered to be the crisis level area of the GDP power intensity.

According to such ranking it turns out that in 2004-2010 Ukraine was at the pre-crisis stage. Moreover from 2007 to 2009 at the pre-crisis stage 1, but in 2010 got back to the pre-crisis stage 2. Significant lag of Ukraine from the world and OECD countries energy intensity index depends significantly on the high cost of the import ER (gas, oil and nuclear fuel), however in to a greater degree depends on the reforms effectiveness of the energy

sector. Taking into account the great importance of this index for national ENS ensuring and its general sustainable development, it requires effort concentration and activation of the appropriate efficiency improvement policy of the energy consumption and saving.

To evaluate national energy supply level it is not only the efficiency level of ER consumption which is important, but also their used amount – the sufficiency level of energy supply. For Ukraine the index of the ER specific consumption is in the normal condition (higher than global level), however often stand down to the relevant one of the OECD countries, USA, and Russia. This indicates a quite great potential to raise the level of the specific energy consumption, although on condition of significant efficiency increase.

Also important characteristic of the national energy supply level is the total amount of used electricity per person. Index of the electricity consumption in Ukraine is in normal state (higher of the global level), it is much less than equivalent level of the economically developed world countries, pointing out quite low level of the Ukrainian life quality level, compared with OECD countries. It is impossible to overcome current lag from other developed world countries without far-reaching reforms in the Ukrainian power sector.

*Energy balance* characterizes current ER and their consumption and reflects interrelation between energy production (extraction) and consumption of different types of FER [16-18]. Indexes and indicators of the ENS, which determine key features of the national energy balance (part and basic levels of consumption of certain types of fuel, share and levels of ER which are used for power and heat production etc.) are strategically important, and characterize available and potential level of the energy consumption.

*Table 7*

**Basic indexes of the energy balance in the world, Ukraine and OECD**

| *Type of the FER* | *World* | | *Ukraine* | | *OECD* | |
|---|---|---|---|---|---|---|
| | **ktoe** | **%** | **ktoe** | **%** | **ktoe** | **%** |
| Coal | 3 299 514 | 27,2 | 35 582 | 30,8 | 1 033 297 | 19,7 |
| Oil | 4 095 594 | 33,7 | 11 654 | 10,1 | 2 064 722 | 39,4 |
| Oil products*** | -108 182 | | 1 365 | 1,2 | -106 552 | |
| Gas | 2 540 218 | 20,9 | 43 810 | 37,9 | 1 248 379 | 23,8 |
| Nuclear power | 703 312 | 5,8 | 21 682 | 18,8 | 584 522 | 11,2 |
| Hydro power | 279 644 | 2,3 | 1 019 | 0,9 | 113 584 | 2,2 |
| Renewable energy resources | 1 338 146 | 11,0 | 882 | 0,8 | 298 475 | 5,7 |
| **TOTAL** | **12 149 845** | **100** | **115 472** | **100** | **5 237 724** | **100** |

**** Negative values are received when export exceed the sum of import and domestic production*

Analyzing table 7 it can be noted that the coal consumption in Ukraine in general matches the world tendencies and oil consumption in Ukraine is significantly less, than in the rest of the world and especially OECD countries. Considering the fact that main part of the oil products are consumed by transport, it may be used as national economy growth level characteristic and indicate the economical lag of Ukraine.

As for the gas consumption in Ukraine – it far exceeds the international, as well as the OECD countries level. Moreover taking into account of level of own gas production and needs of import, this fact can be considered as a great threat to the ENS of Ukraine. It should also be noted that the part of used renewable energy resources in Ukraine is significantly less than the world and OECD countries level and requires set of measures to increase this index and to change the strategy of the national energy supply.

Generally, energy balance of Ukraine does not match the global development trend of the energy supply, is not at optimal level and requires a set of measures for its improvement both in terms of ENS, and economic and ecologic feasibility of the ER consumption.

*Stocks and reserves indexes.* Generally, without taking into consideration seasonal natural gas and coal reserves, there are no strategic reserves of the main ER in Ukraine (oil

and oil products, coal and gas), which doesn't meet the demands of the ENS ensuring and global trends.

*Fixed capital assets depreciation in energy sector.* Many works have mentioned that the fixed capital assets depreciation of the energy sector exceeds 60% and may be considered as critical [19].

*Indexes of general and monopolistic country dependence.* Certainly the fact that Ukraine is experiencing lack of energy negatively influences its ENS, but many other countries have the similar or even worse level of the energy self-supplying.

*Efficient indicators of environmental pollution.* In spite of positive dynamics of the index decrease of the GDP carbon dioxide capacity (PPP) of Ukraine till 2009 value in 2010 (0,96 kg eq. CO2/$) more than by 20% exceed the Russian level and is more than twice greater than world level and three times greater than OECD countries. Such situation is a threat to ENS of Ukraine, especially considering the world tendency of global climate changes prevention actions.

*Indices of ecological emissions intensity.* Despite the implementation of different measures aimed at struggle with climate change, all countries in 2010 increased $CO_2$ emissions per person.

It should be mentioned that there is definite link between intensive indexes and efficiency of the ecological emissions, which displays certain country development orientation, namely: achievement of a certain level of the development (GDP per person) may be reached through of widespread energy usage and thus high level of emission intensity, or through reaching high national energy consumption efficiency (low level of the GDP carbon intensity) and consequently lower emission rate level.

According to the given ENS index Ukraine takes intermediate position between economically developed countries (OECD countries) and world level of this index, which on the one hand shows the availability of a quite considerable industrial and economical potential, on the other hand – poor ecological state.

## 4.Conclusion

Current state analysis of the Ukrainian energy security implies that the national general level of energy dependency in 2003-2010 lies within the limits of safe range, but needs its improvement. Coal energy dependence of Ukraine is normal, however oil and oil products is considered to be pre-critical and gas dependency is critical. The general monopolistic dependency level as well as oil and gas monopolistic dependency is critical.

According to the main metrics and indicators of the ENS, Ukraine significantly concedes not only to the economically developed OECD countries, but to the global level as well. First of all it refers to energy intensity indexes, GDP carbon intensity, renewable energy sources usage, monopolistic dependency level on strategically crucial energy resources etc. In spite of certain positive trends in the dynamics of change of certain indexes of the ENS, which indicates the correspondence between strategic energy development of Ukraine to the world trends, however rate and sustainability of such changes are quite unstable. In general the ENS level of Ukraine can be considered as close to critical, which implies development of strategic and tactical series of measures in order for its improvement and bringing it to a normal state.

Main ways to improve ENS of Ukraine are large-scale improvement of energy efficiency of the technological and social processes in national economy, under the condition of proper national quality of goods and services ensuring; economically appropriate and ecologically acceptable replacement of the imported energy resources with national

production or with their substitutes of home-production; evaluating of rational standards and creating of the proper tactical energy resources supplies in the country in order to prevent the negative influence of unforeseen situations, either natural, economical or political; massive improvement of investment activity in the energy sector through reforming and transparency improvement of the energy markets with state supervision and regulation, aimed at supporting normal level of competition, monopolization prevention and unjustified price fro energy resources growth.

## References

1. The concept of executive bodies in ensuring the energy security of Ukraine (approved by the Cabinet of Ministers of Ukraine of 19 January 1998 p. №48), http://zakon.rada.gov.ua.
2. Main directions of state policy in the sphere of energy security, approved by the Decree of the President of Ukraine on December 27, 2005 №1868 / 2005.
3. Abalkin L.A. Russia's economic security: threats and reflection // Economic Issues. - 1994. - №12, pp 4 - 13.
4. Green Papers / The European Union. http://europa.eu/documentation/official-docs/green-papers/index_en.htm.
5. The Energy Charter Treaty. http://www.encharter.org.
6. The Law of Ukraine "On Electricity" of 16 October 1997 p. N 575/97 VR. http://zakon0.rada.gov.ua/laws/show/575/97-%D0%B2%D1%80.
7. Ensuring energy security of Ukraine / National Security and Defense Council of Ukraine, Nat. in. T-international security problems. - K .: NIPMB, 2003 - 264 p.
8. Barannik V.A. Energy security: an analysis of the formation of the modern paradigm // Journal of Volodymyr Dahl East Ukrainian National University. - № 1 (172), 2012 - P.127-132.
9. Shevtsov A.I., Zemlyany M.G., Barannik V.A. and others. Energy Security of Ukraine: strategy and mechanisms to ensure / Ed. A.I. Shevtsova. - Dnepropetrovsk: Podogy, 2002 - 264 p.
10. Methods of calculating the level of economic security of Ukraine (approved by the Ministry of Economy 02.03.2007 №60) / http://me.kmu.gov.ua/control/uk/publish/.
11. Energy Policies of IEA Countries / International Energy Agency, http://www.iea.org/publications.
12. Ukraine: Energy Policy Review – 2006 / International Energy Agency, http://www.iea.org.
13. Energy Policy outside of IEA member countries: UKRAINE 2012 / International Energy Agency, http://www.iea.org.
14. Key World Energy Statistics / International Energy Agency. http://www.iea.org
15. Barannik V.A. Energy security of the BSEC: quantitative comparative analysis // Problems of development of foreign economic relations and attract foreign investment: a regional perspective. Collection of scientific papers. Part II. - Donetsk, 2011 - P.9-15.
16. Podolets R.Z. The energy balance of Ukraine: modeling and forecasting. K.: Institute for Economics and Forecasting, Ukrainian National Academy of Sciences, 2007 - 174 p.
17. Lir V.E. Organizational and methodological aspects of development, economic analysis and forecasting total energy balance of Ukraine // Economics and Forecasting. - 2004. - № 2. - P. 101-115.
18. Lir V.E., Podolets R.Z. Structural changes in energy markets and long-term energy balance // Ukraine Ukraine's economy: long-term strategy and policy development / ed. Academician V.M. Heets. - K .: Phoenix, 2003 - P. 771-793.
19. State Statistic Service of Ukraine. http://www.ukrstat.gov.ua/

# Information Technologies of Internet Resource Promotion at Enterprise

**O.V. Pugachova**
Gomel Fr. Skaryna State University, Belorussia
opugacheva@gsu.by

**Abstract.** Basic methods of the increase of an enterprise's internet resource attendance, their advantages and disadvantages have been considered in the article. Ways and means of practical realization of these methods for a particular enterprise and positive effect from its use have been described.

At the global level, today the Internet technologies increase global GDP by 20%. More and more companies, including in Belarus, attract them to develop their business, transforming various business processes, including sales, logistics, marketing. According to experts, in the first half of 2014, 18% of Belarusians bought goods over the Internet. So many entrepreneurs are aiming to turn their visitors into customers. One of the key tools here is advertising. Progressing this sector, first of all, thanks to the specialists in the field of Internet marketing.

The tendency to transference of commercial activity into the internet is being observed nowadays in the modern conditions of business dealing. More and more enterprises began to use actively internet technologies to organize their activity. It has been the main reason for beginning of active competitive struggle between the main sources of information and advertisement on the internet, which have an influence on public image and give the first and strong enough impression about it, specifically between enterprises' websites. Not insignificant factor, which gives incontestable advantage, is high attendance of this source.

As an example we will analyze the website of one of Belarusian enterprises, which produces and realizes household chemical goods. This digital resource, which is also the main advertising resource, can't be considered effective because of its low attendance. The conclusion is based on the statistic facts about the enterprise's website attendance. So for the considered period of time (month) in 2014 year the number of visitors was 141 persons. To be noted that only unique visitors are calculated. Among the main cities, from which access put into practice, Gomel, Minsk, Moscow, Grodno and others can be marked out. In order to understand how low this number is, we will use statistic information from the searching system Yandex, which official resource is www.wordstat.yandex.ru. So for the considered month in 2014 year the request "household chemical goods" was shown 94 013 times by searching system Yandex. On the basis of the information about the number of requests "household chemical goods" on regions we can calculate the enterprise's attendance percent of the digital resource in each city: Gomel – 10,8; Minsk – 1,5; Grodno – 7,1; Moscow – 0,09.

As a result, it's actual for an enterprise to carry out the work on the increasing of website attendance, which will raise the number of consumers of household chemical goods who are conusant of enterprise's activity, and therefore will influence on demand for the goods produced by the enterprise.

There is a large variety of methods to increase internet resource attendance, they are:

- search optimization (search engine optimization, SEO) — package of measures for raising the website positions in the results of search system issue on defined users'

requests with the purpose of site promotion. When entering request «household chemical goods, Gomel» into search line Yandex, the digital resource of superadded liability company, consider the enterprise can be met on page 13 of 14 ones. This says about the significance of SEO of the enterprise's site. As a matter of fact, SEO touches upon all technical moments of site promotion and also optimizing and sense acceptability. However another methods of promotion and means of raising of attendance shouldn't be forgotten[1];

- the placement of information in social networks. This way of advertising and informational promotion of the production of the enterprise and the enterprise itself can exist without main digital resource, but at the same time when having your own site, it's effective as far as the increase of attendance of the site is concerned, due to the establishment of the click-through to the internet-resource and its advertisement on the page in a social network;
- the publication of links on pages of thematic forums. To attract audience of thematic forums to the site is more difficult than to increase attendance by means of social networks, because this method demands the publication of messages, which contain references. And as is well-known, far not every comment, left at a forum and containing the reference, has the chance to be approved. It is necessary also to get some authority at such forums, otherwise constant participants of forums will ignore sent messages and references in them;
- Online advertising of the site and organization. As the considered enterprise realizes the production through internet-resource, this way can be used both in the increase of site attendance and in the increase of sales volumes.

Online advertising and SEO consider maximum expense. Internet advertising value depends on its chosen type and advertising purpose. In this case Online advertising purpose is the increase of the number of visitors of an enterprise's site. The main types of Online advertising are contextual advertising, media advertising, media contextual advertising and geo-contextual advertising. To be noted that geo-contextual advertising cannot be used for the accomplishment of the mentioned purpose of internet promotion, because it informs about the location of a store or an enterprise and is appropriate more for retail outlets. The first three types of advertising can be regarded as the most appropriate ones.

When all the pros and cons of each type of advertising were analyzed, it was concluded that the most appropriate type of advertising for the accomplishment of the planned purpose is contextual advertising. First of all its advantage is attributable to the fact that this type of advertising considers payment for clicks-through, which is payment only for ad transfers but not just for impression as for instance in banner advertising.

The best of all is to deliver development, placement and analytics of contextual advertising into the charge of experts in this field. Generally fees of the agencies which specialize in contextual advertising depend on the number of keywords in an advertisement and advertising budget. Payment standard can be established as a fixed monthly sum or as percent of the monthly budget of an advertising campaign. In order to define which method of establishing of payment standard for the service of an unaffiliated company is more profitable for an enterprise, it's necessary to become aware of the sum of advertising budget, which depends on cost per click.

When creating an advertising campaign with tariffication model "CPC" ("cost per click"), it's necessary to establish the maximum sum, which an enterprise is ready to pay for a user's click on an advertisement.

Maximum price for a click is the maximum sum that an enterprise is ready to pay for pressing on an advertisement. As a rule, advertising providers pay less. Final amount that is

paid by an enterprise is an actual price for a click. Cost of pressing never exceeds the maximum price for a click fixed by an enterprise.

Rates can be established automatically and manually. In the first case they are determined by a system depending on the specified budget of an enterprise. As far as an advertising budget is unknown, we will analyze the establishment of cost per click manually [2].

In order to find a necessary cost of click-through it's necessary to define an average cost of goods or service. Then to calculate a forecasting offtake from an expecting number of clicks-through. An average conversion from clicks-through to calls or applications works out approximately 10 percent, and a conversion from applications to purchases is about 15-20 percent. However these numbers are not accurate. An actual result will depend on the quality of a site and service. After matching of costs with an average sum which can be received, we have to make assertions about reasonableness of the price for a click [3, p. 90]. When accounting on this pattern was performed, the optimum price for a click for an enterprise was established at the rate of 900 Br.

Then it's necessary to determine a possible number of clicks on an advertisement per month. Good contextual advertising with configured time and geo targeting produces more than 350 clicks daily (more than 10500 clicks per 30 days) [4]. As far as advertising campaign will be at initial level, an expecting number of clicks-through from one placement of contextual advertising on the digital resource of an enterprise can be determined at the rate of 4000 clicks per month.

With cost per click at the rate of 900 Br an advertising budget will be 900*4000=3 600 000 Br. This advertising budget can be concerned as low, therefore the optimal method of payment to an unaffiliated company is percent of the monthly budget. Then expenses on the organization of contextual advertising can be calculated as the sum of an advertising budget and payment for the service of an advertising agency.

On the basis of providing of all mentioned calculations the cost advantage from the increase of the number of visitors by means of Online advertising was determined at the rate of 50,7 million Br per year.

In the course of SEO work of qualified specialists in this field is also necessary. As far as an enterprise doesn't have specialists in SEO of digital resources, there are two possible methods of their involvement:

–the organization of training to SEO of a digital resource of enterprise's personnel;

–the search of an unaffiliated company for the accomplishment of SEO of a digital resource.

The first method considers completing of the special course of SEO by a specialist of superadded liability company consider the enterprise. It has to guide the specialist who is an advanced user of computer and computer technologies to such course. In most instances such specialist is the programmer of an enterprise. However one-time completing of the course doesn't suggest that the specialist of the enterprise will learn all aspects of current type of work, and most probably valuable time and resources will be required for the effective organization of SEO by means of training of an enterprise's specialist.

The second method of involvement of specialists in SEO considers the search of unaffiliated companies in this field. On the basis of learning of various suggestions on the part of these companies and reviews about them, it was concluded that the most profitable condition is the monthly payment for site promotion on the actual result. It considers fees only for those months when site is situated on the first page in the results of search systems

issue. The optimum price per month is 1974 thousand Br. On the mentioned conditions for instance organization Aktok works.

To be noted that search engine optimization is one of the most effective ways to increase site attendance. The value of search engine optimization is that the visitor, when entering into a search system and finding out some sequence of sites in the answer to his request, suggests that these sites are in the top issue due to the fact that at the current moment their contents are the most valuable (relevant) and correspond maximally to his request. While advertising surrounding the issue is nevertheless the manifestation of someone's will.

Advantages of Search Engine Optimization:

−very accurate targeting (similar to contextual advertising);

−an appeal to actively seeking audience;

−the highest "clickability" (compared to other sources of search traffic);

−long-term effect (some time after the cessation of promotion the site is still in search results, which leads to the residual effect of free advertising);

−no negative "advertising" effect (a user thinks that he came to the site only due to his own effort);

−relatively low budgets;

−high conversion of visitors into contacts, and then converting into purchases (at the level of contextual advertising).

Disadvantages of Search Engine Optimization:

−the necessity to refine the website for the requirements of search systems;

−a significant period required for achieving of sustainable results.

It should be noted that search engine optimization is not well suited for the sudden marketing tasks (short-term promotions) because of a long period of preparation of campaign and is appropriate little in the conditions of a new product on the market [3, p. 116-117].

To determine the possible number of visits, when the site company is in the first ranks issued by search system results, it's necessary to view statistics of other sites that are on the first page of search system results. When introducing the request "household chemical goods" in search system Yandex, the top position is the online shop of Mary Poppins. His attendance during the analyzed month was 8407 unique visitors, which exceeds the number of visitors of digital resource of superadded liability company consider the enterprise in the same period in 59,6 times.

Trade portal www.shop.by, located on the first result page of search issue on request "household chemical goods", also has a high attendance at the rate of 7263 unique visitors. Therefore promoting the site on a more prominent position in the search systems, we can provide a significant increase in site traffic to the level of more than 8000 visitors. It can be suggested that as a result of a placement of the digital resource of the enterprise in the top ten sites, which is issued on request "household chemicals" and other similar inquiries, site attendance will increase to at least 5000 visits per month, compared to 141 visits in a given time period. The economic effect from search engine optimization with involving of unaffiliated companies per year will be about 99,7 million Br.

Summing up, it should be noted that the providing of SEO and the placement of contextual advertising should be carried out so as to embrace as many search combinations as possible and provide the high level of "clickability". Otherwise, it can't be said that the economic effect of these measures will be equal to the sum of the economic effects of each event separately. In fact, if contextual advertising will appear in the search system Yandex when typing request "household chemical goods", and, at the same time, in the top line the

site of the company will be located, the combined effect will be equal to the effect of holding one of these events. That is it will not allow to increase the number of visitors to 9000 visitors, their mark will be at the level of 4000-5000. At the same time, the expenses will be equal to the costs of both SEO and the placement of contextual advertising.

If the proposed activities are carried out properly, it will allow to achieve the economic effect in the amount of: 50,7 + 99,7 = 150,4 million Br per year.

Thus, at present, widely used technology, Internet marketing, each of which has its own features and tools. Thus, contextual ads (ads in search engines for the keyword set) aimed at increasing sales and attracting new customers through the Internet. Contextual ad see users who are looking for a Network that offers the advertiser. The advertising message is perceived by the user as the answer to your question and helps you find the right product or service. On the advertising market of Russia in 2013, revenues for Internet advertising 71.7 billion USD, including contextual advertising - to 51.6 billion USD. In Belarus, the volume of revenues from online advertising increased from 7.8 million USD in 2010 to 15.6 million in 2013) this year, experts predict the achievement of 20.9 million, and in 2018 - 32.2 million USD, and about half of these amounts will be contextual advertising.

Another type of online advertising - SEO - website promotion in search engines. This tool is a set of measures that provides a "hit" on the top of the search results due to work on the content of the site. Its main advantage is the high profitability of attracting traffic. If contextual advertising requires payment for each visitor to the site, for SEO unplanned visitors - free. Therefore, this tool costs the advertiser cheaper. The SEO activities aimed at ensuring the leading position of the company's website in search engines for specific key words. This is achieved by optimizing the structure of the site, generating tags and meta tags, the increase in the external thematic links, register in directory and so on.

To secure the attention of site visitors and turn them into clients called remarketing. It allows you to access these visitors to other web sites using promotional messages and special offers that will encourage them to return to the original resource and to make a purchase.

If the company has sufficient funds on advertising and it is interested in increasing the demand for their goods and services, it can be widely applied banners and video ads. In the formation of demand for innovative products employ the services of copywriters, including opening and blogging on the same topic on different sites to attract visitors to the news. Here are also widely used social network.

## References

1. Карпенко В. Анализируем конкурентов и просчитываем стоимость SEO // seoprofy.net: ежедн. интер-нет-изд. – 2010. – 26 фев. – URL: http://seoprofy.net/2010/02/analiziruem-konkurentov-i-proschityvaem-stoimost-seo/
2. Официальный сервис Google [Электронный ресурс] / Официальный сервис контекстной рекламы. – М., 2005. – URL: https://support.google.com/adwords .
3. Интернет маркетинг на сто процентов / С.В. Сухов [и др.]; под ред. С.В. Сухова. – СПб.: Питер, 2009. – 240 с.
4. Официальный интернет-ресурс организации click-i [Электронный ресурс]. – Минск, 2010. – URL: http://www.click-i.by.

# Evaluating the Effectiveness of Information Security

**Oleg Solonenco**

ASEM, Chisinau, Moldova.
mcsa05@mail.ru

**Abstract.** The paper reviews actual questions of economics aspects of information security.

**Keywords.** Information security, return on security investment.

## 1.Introduction

Karl Marks in his times has defined three the most important for the society resources: labor, land and capital. To these resources were added- time, that each day is not enough in any type of the human activity and the informational resource, the importance of which is difficult to underestimate, because of the fact, the present of this resource can cover the absence of the other resources. Informational resources, like the material resources, have quality and quantity, have their first cost and their price. The first cost of the information is determined by the quantity, of the energy (mental efforts), that was spend for the it's production, financial and material expenditures for its documentation, saving, processing and transferring through the connections' channels. The price of the information constitutes from: first cost and profit from its distribution. It is clear, that the value of the information (real and potential) is defined firstly by the profits, that provides this information. In competitive struggle are widely used different activities that are directed on the receiving of the confidential information by the different methods, even by using the direct industrial spying and the utilization of the modern technical spying means.

JPMorgan Chase & Co, the nation's largest bank by assets, has revealed the scope of the cyber-attack that compromised its data. And while the number of households affected doesn't surpass the 110 million accounts that were compromised in the data breach in late 2013, it does comprise more than half of all U.S. households. Compromised information at JPMorgan consisted primarily of customer contact information — names, addresses, phone numbers and email addresses, as well as internal JPMorgan information about customers [1].

The question about the defense of the information against the illegal capture of it is very important for today. Because the main rule of the XXI century will be the rule: who possess the information, those posses the world.

## 2.Relevance of the topic

The development of the informational technologies during the last years has a very high speed. The modernization and the changes of computer facilities transfer to the new operational software, refreshing of the new versions of the applicative programs and so on, is the characteristic feature of the modern computer application. At present anywhere is implemented the computer networks (local, regional and global), thanks to which the remote access to the public resources is assured (to the data basis, computers, printers and so on). The quick increase of the informational flows for the electronic exchange of the data, the

development of the e-post and the internet have simplified the procedures of the trade transactions, that are performed on the national and international levels. The informational flows are very big, their speed increases each day, they practically overflow the people. The volume of the scientific knowledge, without the other information, the speed of which is significantly higher, on the scientists' estimation is doubled every 5 years. It is necessary to mention about so called compliance of the different laws, acts, the number of which during the last time have grown. In the year 2009 by the European Parliament was approved the directive Solvency II ( analog of the Sarbanes-Oxley Act),the main purpose of which is to protect the holders of the policies and the beneficiaries in the European Union, by defining the spectrum of claims to the capital and the reports of the insurance companies. The creation of the trade mark and the business reputation takes time, while the only one incident of the informational security can spread to them a big damage or even lead to the unrecoverable consequences.

In the result of the changes of the external environment of the business processes, the company is damaged. Then the informational risk - this is the possibility of the appearance of the random event, that lead to the break of the functioning and the decrease of the of the information's quality in the informational system of the company, and also to the direct utilization or distribution in the external environment, in the result of which the company is damaged. The informational risk is caused by the internal and external reasons. The factors of the informational risks, in less degree are connected with the concrete sources of the risk, then the reasons of the risk. They reflect basically the state of the informational system of the company in whole, and specially the state of the sub-system of the counteraction to the informational risks. Factors of the risk are very close to the concept of the vulnerability of the system, that is used by the specialists of the informational defense. For the coming of the risk event, there is a necessity of the existence of the reason and the factor of the risk. Damage to the company as the result of the realization of the informational risk can be the consequence of the utilization in the business-process of the managerial information, the quality of which in the result of the action of the informational risk has decreased till the unaccepted level. Also the companies have great losses because of the direct influence of the informational risks on the objects of the informational system, in the result of which the objects come to the invalid state. Such risks are called the direct informational risks. For the reconstruction of their availability the company has to expend the resources. The examples of such risks are the deletion of the technical means in the result of the emergencies or the force- majors, losses of the program technologies, informational data basis and so on.

## 3.Economic aspect of information security

During the evaluation by means of the classical methods of the economical efficiency of the offered measures of the informational resources defense, appear difficulties with the definition of Return On Investment (ROI), because during the calculation of the waited losses from the concrete threatening for the certain period of time - Annual Loss Exposure (ALE), is necessary the price of the resource - Asset Value (AV) and the measure of the vulnerability of the resource to the threat and vulnerability Exposure Factor (EF). During the calculation of the Asset Value (AV) appear the complications with the evaluation of this or that resource or the asset of the informational system, intellectual property, brand, knowledge and skills of the staff, efficiency in staff- management, correct distribution of functions, tasks and resources of the company among its subdivisions for the more effective achievement of the goals of the development of the organization and at the end, its effective cooperation with

its clients. In relation with this in modern approaches to the strategically management of the organization's development if very important the management of the intangible assets- human capital and business- processes of the organization. In such a way the Organization for Economic Co-operation and Development (OECD) defines the intellectual capital as an economical value of two intangible assets of the company – human capital and the organizational capital. A number of researches consider, that this type of capital is formed not only with human and company's but also with social capital. Other researches focus their attention not so much on the components, that create the intellectual capital, but on the processes, that determine the success of the formation of the all component parties of this type of assets. To this firstly refer the processes of the formation of the "company's knowledge", that is the processes of the definition of the hidden (individual) knowledge of the managers and the generation on the basis of these knowledge the team knowledge. The link that consists from the mentioned two processes provides the company a possibility to expropriate the management's knowledge from their carriers and to form on the basis of the already detected successful decisions for different business problems the library of the efficient decisions.

During the calculation of the value of the threat and vulnerability Exposure Factor (EF) appear the complications with the calculation of the probability of the threat realization, because there are no necessary statistical data per domains and per countries and the expert's evaluation cannot guarantee 100% of the result. It is difficult to guarantee the results for the reporting to the financial management, because of the fact, that the economical efficiency appears during the appearance of the forecasted event. In other words, the economical efficiency appears during the successful realization of the threat.

By the utilization of the calculation of the costs methods evaluation - the definition of the Total Cost of Ownership (TCO) it is necessary the comparison of the certain indicator TCO with the analogical indicator TCO per branch (with the analogical companies) and with the "best in group" - that is not real because of the fact of absence of such data in our country. The infrastructural project can be evaluated firstly of all by using the Total Cost of Ownership (TCO), Real Option Valuation (ROV) and by the method of the Applied Information Economics (AIE). The first methodology provides a possibility to estimate the influence of the project on the costs; two other add the risk's evaluation and the flexibility that are offered by this infrastructural decision. The evaluation of the business-projects – this is first of all the evaluation of the return for the business, and it is better to perform it by means of the Return On Investment (ROI), Economic Value Added (EVA) or Applied Information Economics (AIE). For the financial evaluation of the outsourcing projects methodologies are chosen in dependence on the character and the scale of the project. Taking the decision about outsourcing, the company firstly comes from the non financial profits. If the outsourcing can resolve the problem more quickly, efficiently and chipper, then there should be used the outsourcing. If the business tends to receive from the outsourcing the financial profits, then these profits and related to them risks should be taken into the consideration. In this case pass ROI (including the TCO like a method of the costs registration), ROV or AIE, in dependence on the project's scale and the qualification of the staff IT specialists. If it is accepted the program of the creation of the new business in the outsourcing's sphere of the IT services, there will be a need in the EVA evaluation, the forced evaluation of the risks and the possibilities on the basis of the ROV and AIE [2,3].

## 4.Conclusion

Economical efficiency of the process of the informational security's management in many cases depends namely on the recognition of the fact, what it is necessary to defense and which efforts are necessary for this. The risks management provides a possibility to structure the activity of the management of the informational security, to find the common language with the top management of the organization, to estimate the efficiency of the work and to prove the decisions related to the choose of the concrete technological and organizational steps for the defense in front of the top management.

To resolve this problem is impossible without the attraction of the middle and top managers of the main direction of the organization's activity. Using the different approaches for the measuring and improving of the degree of the informational defense in the organization, the evaluation of their objectivity, represents the principal factor, that enable the degree of their efficiency and the basis for the performing of necessary improvements in the domain of the informational security of organization.

It is supposes the receiving of the methodology, that will provide a possibility on the projecting phase of the measures of the informational security to receive the results about the economical efficiency and the economical effect that we can obtained from the implementation of the suggested measures and that will provide a possibility to connect the interested subdivisions through their participation in the construction of the economical model, informational model and the model of the informational security.

## References

1. JP Morgan Says 76 Million Households Affected By Data Breach (http://www.forbes.com/sites/maggiemcgrath/2014/10/02/jp-morgan-says-76-million-households-affected-by-data-breach/)
2. Solonenco O. "Evaluation of the Economic Efficiency of Information Security": Papers of the 6th International Information Security Conference " Securitatea Informaţională - 2009" (May, 20-21, 2009), Academy of Economic Studies of Moldova
3. Solonenco O. "Methods of Calculation of Economic Efficiency of Information Security": Papers of the 7th International Information Security Conference "Securitatea Informaţională – 2010" (April, 15-16, 2010), Academy of Economic Studies of Moldova
4. Solonenco O. "Information risks in the Business Intelligence systems": Papers of International Scientific Conference on the theme:" Business management systems for small and medium enterprises" (April, 24-25, 2010), D. A.Tsenov Academy of Economics. The Republic of Bulgaria
5. Measuring effectiveness in Information Security Controls (http://www.sans.org/reading-room/whitepapers/basics/measuring-effectiveness-information-security-controls-33398)

# Aspects Regarding the Implementation of International Financial Reporting Standards in the Republic of Moldova

**Lica Erhan,**
Academy of Economic Studies of Moldova
e-mail: lica.erhan@mail.ru

**Abstract.** This study aims to analyze the process of implementation of International Financial Reporting Standards (IFRS) in the Republic of Moldova, which has become mandatory for all public interest entities from 1 January 2012. The paper includes a practical analysis of the impact of transition to IFRS on the financial results of a public interest entity - Moldova Agroindbank, which is the largest commercial bank, with the highest market share in the banking sector of the Republic of Moldova. By the light of the analysis of the impact of transition to IFRS can be concluded that the connection to the new standards had a beneficial impact on the bank, maximizing its market value, increasing investors, customers, business partners confidence and contributing to the reduction of risks.

**Keywords:** International Financial Reporting Standards, Public Interest Entity, Financial Statements.

## 1.Introduction

International Financial Reporting Standards are a set of accounting standards issued by the International Accounting Standards Board (IASB). IASB was established in 1973 by professional accountancy bodies from nine countries (Germany, Australia, Canada, USA, France, Japan, Mexico, the Netherlands and the UK) for the purpose of issuing accounting standards to be used for the preparation and presentation of financial statements and promote their acceptance and implementation worldwide. IFRS are considered to be a set of standards that establish general rules, but also require some specific treatments. International Standards in their evolutionary process, involve two types of standards:

- International Financial Reporting Standards (IFRS) - standards issued after 2001;
- International Accounting Standards (IAS) - standards issued before 2001.

Application of IFRS around the world involves common accounting principles valid in all the countries of the world, facilitating the development of international relations between states and ensuring comparability of accounting data, minimizing costs of information systematization and ensuring the transparency of local business environment.

Currently, IFRS is applied in many countries of the world, like: EU Member States, Hong Kong, Australia, Russia, South Africa, Singapore, Pakistan, Brazil, Chile, Costa Rica, Guatemala, Mongolia, Namibia, Panama, Venezuela, Kazakhstan, Tajikistan, Georgia, Belarus, the Baltic States and other 100 countries promote a policy of convergence to them.

The banking sector of the Ukraine adopted IFRS from 1998. IFRS is applied in Uzbekistan since 2002, and in Azerbaijan - since 2006. Russian banks have implemented IFRS from 1 January 2004. IFRS are longer applied in Armenia, Kazakhstan , Kirghizstan, Tajikistan, Georgia, Belarus, the Baltic States, etc.

According to European Directive from 2002, listed companies on European stock exchanges were required to present consolidated financial statements according to IFRS since 1 January 2005. All the neighboring countries of Western Europe, including Romania, began the transition to international standards since 2005, these changes being perceived as a part of the process of European integration.

## 2.Definition and general conditions of implementation the IFRS in the Republic of Moldova

Since the entry into force on 1 January 2008 of the Accounting Law no.113-XVI, in the Republic of Moldova many changes have occurred on the connection of the national accounting standards to the international ones. A special place in this law, lies on the subject related on the application of International Financial Reporting Standards - IFRS, according to EU Directives. According to the new law, all public interest entities (financial institutions, investment funds, insurance companies, non-state pension funds, joint stock companies that are listed on the Stock Exchange of the Republic of Moldova) were obligated to apply international financial reporting standards IFRS since 1 January 2012. In this way, about one hundred of Moldovan companies have complied with this law.

The need to meet the requirements of the global economy dictates the passing of the Moldavian entities to a financial reporting in accordance with International Financial Reporting Standards (IFRS). Implementation of IFRS in the country was a difficult process, but also a necessary one, its main objective being both the increase of financial reports transparency and reliability, within the country for shareholders, state authorities, as well as externally, among investors and international financial organizations.

According to art. 3 paragraph. (1) of the Law on Accounting, IFRS are standards and interpretations issued by the International Accounting Standards Board (IASB), which are valid in Moldova after their acceptance by the Government. These standards include International Financial Reporting Standards (IFRS), International Accounting Standards (IAS) and Interpretations on the application of these standards. On September 1st, 2013, in the Republic of Moldova were available 13 IFRS (table 1), 28 IAS, 16 IFRIC Interpretations and 8 IAS Interpretations[1]. Thus, in 2013, in the country were accepted additionally 5 IFRS: IFRS 9 "Financial Instruments", IFRS 10 "Consolidated Financial Statements", IFRS 11 "Joint Arrangements", IFRS 12 "Disclosure of interests in other entities" and IFRS 13 "Fair Value Measurement"[2]. There were also withdrawn IAS 31 "Interests in Joint Ventures", which was replaced by IFRS 11 "Joint Arrangements" and 5 interpretations (IFRIC 8, IFRIC 11, IAS 12, IAS 13, IAS 21).

IFRS 13 "Fair Value Measurement" is a relatively new standard, which started its applicability from 1 January 2013. The purpose of this standard is to strengthen guidance on the use of the fair value term. It does not introduce new requirements regarding its necessity, but it only provides a single source of fair value measurement, that must be used every time when the fair value is required or permitted in other IFRSs.

According to IFRS, the fair value is the activity of the evaluators, who believe that the fair value is nothing else than the market value of assets. Fair value is the amount for which an asset could be exchanged or a liability can be settled between stakeholders. Therefore, the concept of fair value is a challenge, but also a premise of value in the context of transition to international standards.

In some cases, however, the consideration given or received (or face value) may not necessarily be the fair value of the financial instrument. For example, the fair value of long-term receivables bearing no interest is not equal to its nominal value and, therefore, a part of the consideration received is different than its fair value. Whereas the receivable should be initially registered at the fair value, this value must be estimated. The fair value of the instrument can be highlighted in comparison with other transactions that may be currently

observed on the market in connection with the same instrument or based on a valuation technique that includes only data from observable markets. For a loan or receivable free of declared interest, the fair value is identified using the discounted cash flow technique. However, the standard allows that short-term receivables and debts to be assessed at original invoice amount, when the effect of updating /discounting is insignificant.

**Table 1 The list of IFRS operating in the Republic of Moldova**

| IFRS applied in the Republic of Moldova | |
|---|---|
| *1* | *2* |
| 1. IFRS 1 | First-time Adoption of International Financial Reporting Standards |
| 2. IFRS 2 | Share-based Payment |
| 3. IFRS 3 | Business Combinations |
| 4. IFRS 4 | Insurance Contracts |
| 5. IFRS 5 | Non-current Assets Held for Sale and Discontinued Operations |
| 6. IFRS 6 | Exploration for and Evaluation of Mineral Resources |
| 7. IFRS 7 | Financial Instruments: Disclosures |
| 8. IFRS 8 | Operating Segments |
| 9. IFRS 9 | Financial Instruments |
| 10. IFRS 10 | Consolidated Financial Statements |
| 11. IFRS 11 | Joint Arrangements |
| 12. IFRS 12 | Disclosure of Interests in Other Entities |
| 13. IFRS 13 | Fair Value Measurement |

Source: developed by authors based on data available at www.ifrs.org.

## 3. The threats

The adherence of the Republic of Moldova to the International Standards is quite difficult but also very necessary.

**The main difficulties** faced by the Republic of Moldova in the context of transition from the national accounting standards to international ones are:

- Insufficient staff qualification;
- High costs related to the implementation of IFRS (IT, training, etc.);
- Lack of transparency in the local business environment;
- Fiscal instability caused by frequent legislative changes;
- The difficulty to perceive the essence of IFRS at the initial stage of implementation;
- The limits imposed by the existence of outdated IT systems.

**The main problems** in the implementation of IFRS in the Republic of Moldova are as follows:

- About 98% of the Republic of Moldova entities are SMEs for which the implementation of IFRS is expensive;
- Insufficient financial resources deriving from the high costs related to the transition to IFRS and accounting reform;
- Educational materials do not comply with IFRS, and the curriculum do not provide adequate training of future specialists in accounting field;

- Low level of training in accounting field, of the public officials, that belong to the tax and legal system;
- Contradictions between Romanian and Russian translation version of IFRS.

## 4. Structure

However, the connection to international standards ensures to the Republic of Moldova several **advantages**, including:

- Increasing transparency of local business environment, of foreign investments;
- Ensuring the comparability of accounting information worldwide;
- Reduction of corruption and accounting or fiscal fraud;
- Decrease financial reporting costs;
- Development of internal and external control system;
- Increase the number of specialists;
- Development of international relations between Moldova and other countries;
- Increasing the competitiveness and investment attractiveness of the Republic of Moldova companies, that would be internationally listed;
- Increasing the confidence in the financial reporting in the Republic of Moldova;
- Modernizing the educational system and its adaptation to international requirements;
- Capital market development in the Republic of Moldova;
- Easier way to identify the entity's financial problems and to take measures to combat them;
- Attraction of external funding for developing accounting regulatory framework.

The harmonization of the accounting system regulatory framework in accordance with the requirements of IFRS will create a unique accounting system available internationally. As a result, it will be possible to strengthen the financial statements for multinational companies and the confidence of potential foreign investors. IFRS ensure a greater transparency in financial reporting, but also a better understanding of the financial performance of the reporting entities, increasing the national entities access to international financial markets.

With the adoption of a common accounting system of terms accepted globally, that will allow to perceive correctly the reports, there will be promoted and active developed the capital markets. However, implementation of IFRS will reduce costs of financial reporting and external audit, eliminating the need for multiple sets of preparation of financial reports. And last but not least, the implementation of IFRS facilitates the economic integration of our country in the European Union.

IFRS is not just a pretty innovative approach of financial reporting, balancing between the desire to optimize costs used in your own business and the desire to maximize profits by dishonest optimization, but also a way to check their good implementation. Respectively, with the implementation of new standards by public interest entities (since 2012), several problems were identified both in terms of financial accounting and tax record level, which need to be resolved.

According to the *Report on the Observance of Standards and Codes on Accounting and Auditing in Moldova (A&A ROSC)*, report implemented by the World Bank and the

International Monetary Fund, where is presented the evaluation of the accounting and auditing practices, currently, in the Republic of Moldova there are 45 public interest entities and approximately 50,000 active entities. A total of 569 entities are required to audit their financial statements. This figure consists of 524 joint stock companies and all the public interest entities.

Moldova Agroindbank is an example of a lending financial institution from the Republic of Moldova that annually prepares and submits consolidated financial statements according to the provisions of International Financial Reporting Standards (IFRS) issued by the International Accounting Standards Board (IASB).

The financial statements are prepared in Moldovan Lei (MDL). These financial statements are prepared under the historical cost convention, excepting the securities held for trading which were estimated at fair value, as well as buildings and grounds.

Moldova Agroindbank prepares the financial statements according to the following basic principles: continuity of activity, accrual accounting, assets and liabilities separation, consistency. The consolidated financial statements include the financial statements of Moldova Agroindbank and subsidiaries - MAIB Leasing and MoldMediaCard LLC – on 31 December every year. The consolidated financial statements include the following types of reports:

- Consolidated statement of the financial position;
- Consolidated statement of the global outcome;
- Consolidated statement of changes in equity;
- Consolidated statement of cash flows.

Additionally, financial statements prepared in accordance with International Financial Reporting Standards include notes to the financial statements and independent auditor's report, that is mandatory. Below, a model of financial reporting in Moldova Agroindbank is presented in accordance with the provisions of both IFRS and National Accounting Standards (NAS), by realizing a comparative analysis of the registered outcome (table 2).

**Table 2 The balance sheet indicators determined according to IFRS and NAS requirements at Moldova Agroindbank, 31.12.13**

| NAS | | IFRS | |
|---|---|---|---|
| Indicators | Amount, thousand lei | Indicators | Amount, thousand lei |
| *1* | *2* | *3* | *4* |
| Total Assets | 9201531 | Total Assets | 9617713 |
| Total Debts | 7439194 | Total Debts | 7624404 |
| Total Equity | 1762337 | Total Equity | 1993309 |
| Total Debts and Equity | 9201531 | Total Debts and Equity | 9617713 |

Source: developed by authors based on data from the balance sheet of commercial bank Moldova Agroindbank, from 31.12.11.

In the process of analyzing the results from the table above, there is a tendency to increase the total value of bank assets, as well as the total banking liabilities (total debts and equity) if reporting under IFRS provisions. In this case, when reporting under NAS provisions their value is 9.201.531 thousand lei but when reporting under IFRS – 9.617.713 thousand lei, that is 416.182.000 lei more.

To study deeper, the incidence of reporting under IFRS provisions, in the table below, there is presented the liabilities weight and, respectively, the equity weight in the total bank liabilities (table 3).

**Table 3 The weight of debts and equity in the total banking liabilities**

| NAS | | IFRS | |
|---|---|---|---|
| Indicator | Weight, % | Indicator | Weight, % |
| *1* | *2* | *3* | *4* |
| Total Debts | 80,85 | Total Debts | 79,27 |
| Total Equity | 19,15 | Total Equity | 20,73 |
| Total Liabilities | 100 | Total Liabilities | 100 |

Source: developed by authors based on data from the balance sheet of commercial bank Moldova Agroindbank, from 31.12.11 [4].

The analysis of data presented in the table above reveals that when reporting according to IFRS, equity's weight in total liabilities (about 21%) is higher than when reporting under NAS, when its weight in total bank liabilities constitutes about 19%. In the same context, the weight of debts when reporting under IFRS tends to be lower than under NAS, constituting, respectively, 79% and 81%.

Therefore, it can be easily noted that, under IFRS provisions, there is a tendency to increase the capital's weight in total bank liabilities and to decrease the weight of debts, unlike the reporting under NAS.

Analyzing the data presented in the table 4 it can be noted that the net profit of Moldova Agroindbank at 31.12.12 is higher in the case of preparation the financial statements according to IFRS requirements, representing 290.802.000 lei. At the same time, the net profit obtained by the bank when reporting under NAS provisions constituted 282.362.000 lei, that is 8.440.000 lei less.

Also, when reporting as required by IFRS, the net income before tax is higher than that obtained when reporting under NAS requirements. Therefore, the adoption of the international standards by the bank led to the increase of the income before tax. Nevertheless, increasing the tax base, even if the income tax grows up, the net profit increases too, contributing to increase the market share of the institution as a whole.

**Table 4 Banking net income determined under NAS and IFRS provisions, 31.12.11.**

| Indicators | NAS | IFRS |
|---|---|---|
| | Amount, thousand lei | Amount, thousand lei |
| *1* | *2* | *3* |
| Net income before tax | 282417 | 290857 |
| Net profit | 282362 | 290802 |

Source: developed by author based on data from the Profit and Loss statement of commercial bank Moldova Agroindbank, from 31.12.11.[5]

The same positive situation under reporting according to IFRS can be mentioned in the case of cash flow statement, presented in table 5.

**Table 5 Cash flow analysis under NAS and IFRS provisions, 31.12.11**

| Indicators | NAS | IFRS |
|---|---|---|
| | Amount, thousand lei | Amount, thousand lei |
| *1* | *2* | *3* |
| Net cash flow from the operational activity | (653211) | (622886) |
| Net cash flow from the investment activity | (24914) | (37371) |
| Net cash flow from the financial activity | 191969 | 175293 |
| Net flow before extraordinary posts | (486155) | ------- |
| Net flow after extraordinary posts | (486155) | ------- |
| Total net flow | (494353) | (494095) |
| Cash and cash equivalents at the end of the period | 1703631 | 1706618 |

Source: developed by author based on data from the Cash-flow statement of commercial bank Moldova Agroindbank, from 31.12.11.

The analysis of data in the table 6, denotes that the value of each of the indicators presented in the table is greater by reporting under IFRS compared to the NAS.

**Table 6 The relative impact of applying IFRS in Moldova Agroindbank, 31.12.11**

| No. | Indicators | NAS | IFRS | Relative impact, % |
|---|---|---|---|---|
| *1* | *2* | *3* | *4* | *5* |
| 1. | Total assets, thousand lei | 9201531 | 9617713 | +4.52 |
| 2. | Equity, thousand lei | 1762337 | 1993309 | +13.11 |
| 3. | Debts, thousand lei | 7439194 | 7624404 | +2.49 |
| 4. | Total liabilities, thousand lei | 9201531 | 9617713 | +4.52 |
| 5. | Net profit, thousand lei | 282362 | 290802 | +2.99 |
| 6. | Return on Equity, % *(rd.5/rd.2)* | 16.02 | 14.59 | -8.93 |
| 7. | Solvency, coef. *(rd.1/rd.3)* | 1.23 | 1.26 | +2.44 |
| 8. | Global financial autonomy rate, coef. *(rd.2/rd.4)* | 0.19 | 0.21 | +10.53 |
| 9. | The leverage,% *(rd.3/rd.4)* | 80.85 | 79.27 | -1.95 |

Source: developed by author based on data from the financial statements of commercial bank Moldova Agroindbank, from 31.12.11.

Therefore, reporting in accordance with international standards and requirements generated at Moldova Agroindbank, an increase of 4.52% of total bank assets, of 2.49% of total liabilities and a significant increase in equity of 13.11 %. As a consequence, higher financial results are recorded, so the net profit grows up by 2.99% when reporting under IFRS and, the same time, it generates a higher taxation related to various taxes and fees payments. It also increases the coefficient of the solvency rate and global financial autonomy, which can be positively appreciated as it increases confidence in economic and

financial stability of the institution as a whole. Simultaneously, return on equity decreased by 8.93% due to considerable increase equity value, and together with this, have increased capital cost of ownership, which led to a lower return on equity value. Another factor that can be positively appreciated is the decrease of the bank leverage of 1.95%. As a conclusion, besides the fact that reporting under IFRS in Moldova Agroindbank tends to increase financial results and, at the same time, the income tax increases, there are also advantages of applying IFRS, such as that it creates a solid image of the bank on market, provides financial stability of the bank in the banking sector, and greater credibility to customers and business partners.

## 5. Conclusion

The transition to IFRS is a very important step in the economic development of every country. In the Republic of Moldova the process of adaptation and reforming the financial regulatory framework was a difficult and lengthy one. There existed and still persist problems of vocabulary, practical applicability because of the insufficient number of trained specialists, perception, and assimilation regarding the essence of the new standards. Nevertheless, the advantages of implementing international standards are much higher compared with the transition costs, because reporting according to the international standards will contribute to a substantial reduction in the cost of systematization, processing and presentation of information. As a result, there will be established a common language of financial reporting and will be harmonized the economic relations between states. This will increase the transparency, comparability of information between companies in the same field of activity, but on different markets, will increase the number of foreign investors and their confidence in national financial reporting quality. As a benefit of international standards implementation, may be that, on medium and long term, it ensures an increase in liquidity of all entities applying IFRS and, as a result, it reduces the risks related to every entity. Most entities have adopted IFRS because it prepared consolidated financial statements, the process being a very expensive and complex one, causing problems, problems that are substantially reduced by the use of IFRS. The intensification of globalization process is the main premise for the implementation of IFRS, and at the same time, the global adoption of uniform standards tends to reduce costs, increase efficiency and significantly increase investments in capital markets both on national and international level. The companies hope to increase their financial results through the implementation of IFRS. Therefore, the transition to IFRS is the only way for an entity or a state to progress, development, economic ascension, transparency and safety.

**References:**

1. Law on Accounting nr.113-XVI din 27.04.2007.
2. Report on the Observance of Standards and Codes on Accounting and Auditing in Moldova (A&A ROSC), June 2013.
3. www.ifrs.org
4. www.mf.gov.md
5. www.cnpf.md.

# Data Integration in the University Information System

**Constantin Sclifos**
Academy of Economic Studies of Moldova
sclifcon@vle.ase.md

**Abstract**. In this article are discussed some implementation aspects of the unique identifier of the person in the university information system, and aspects to the integration of resources.

**Keywords**: information systems, e-learning, LDAP, SAML, anti-plagiarism

Among the many problems associated with the operation of university information systems (IS), in our opinion, one can identify the organizational, technical, software-related ones and others.

The issue of information system and data integration has been studied since the emergence of the first database and management system. So far, the problem of integration has still been explored insufficiently, since along with the complexity and great difficulty of this task, due to the rapid development of information technology, the requirements for the information systems and consequently for their integration change. Particular attention is paid to the integration of resources with a focus on the following problems:

- Classification of information resources;
- Development of a single classifier of information and processes for university environment;
- Development of technological processes for processing, storing and transmitting information, and many others.

Close attention is paid to the last two tasks in the operation of IS. Aspects of information resources classification, presented in the works [1,6], served as the basis of this study. This is due to many circumstances and, above all, to a variety of formats of information used.

One of the major trends in the application of information technology in educational institutions is connected with the formation of a unified information environment of the university. Requirements for IS are set based on the goals and objectives and can be summarized as follows:

- All applications should support the main activities and implement comprehensively the necessary functions, ranging from collection, processing and storing to analysis, planning and decision making support;
- Access to the information services of IS should be provided only to authorized users according to their role and performance of duties regardless of their location;
- High level of data integration is designed to ensure the processes of information updating and consistency;
- Connection is required of scalable hardware and technical platforms and specialized technologies for different purposes (for example, database management systems, Internet technologies, electronic document management systems, distributed computing, virtual networks, and others);
- various technologies require the integration of application on the basis of developed model and architecture;

- To manage effectively the information environment, it is necessary to develop and use a set of indicators combining technical and economic indicators that can objectively reflect the demand for applications, promptness of environment, and many others;

- A prerequisite is the process of documenting operations related to backing up, archiving, and restoring.

For interaction between systems, an identification and authentication method suitable for use on all systems must be proposed and tested in practice. In this case, some limitations have outlined the choice of technologies used:

- Technologies must be standardized or generally accepted to avoid problems in the exchange of information between the IS;
- The selected technologies must be sufficiently flexible and easy to use due to the complexity of the university information environment;
- They must not complicate the exchange of information among universities, or participation in projects as EDUGAIN and EDUROAM [7,8];

In AESM systems based on the platform, CMS / VLE MOODLE have been used since 2005 [9]. Since 2010, this system has been utilized as a platform for the development of IS components of AESM. The first of the developed components was the Anti-Plagiarism system. As a starting point, the project - module Anti-Plagiarism CROT, the only available at the time, was used [10].

However, the use of this system as a platform for development has its own advantages and disadvantages. The benefits of utilizing the system CMS / VLE MOODLE as a platform for development are as follows:

- A well-developed course management system with the possibility of using add-ons;
- A large community of users;
- Open License that allows code supplement, if necessary;
- Documented API, etc.

The disadvantages of utilizing this system as a platform for development are:

- Requirements for the competence of the developer due to the complexity of the system CMS / VLE MOODLE for development;
- Dependence of the development on the system evolution;
- When integrating data from other systems, it is necessary to consider the peculiarities of both developed subsystem (module) and the platform / system MOODLE itself;
- The performance of developed subsystem (module) depends directly on the platform itself.

In [2] are shown the basic methods of implementing data integration, which have been used in designing IS:

- Exchange based on data;
- Data replication;
- Web-services technologies;
- Service-Oriented Architecture (SOA);
- Integration servers.

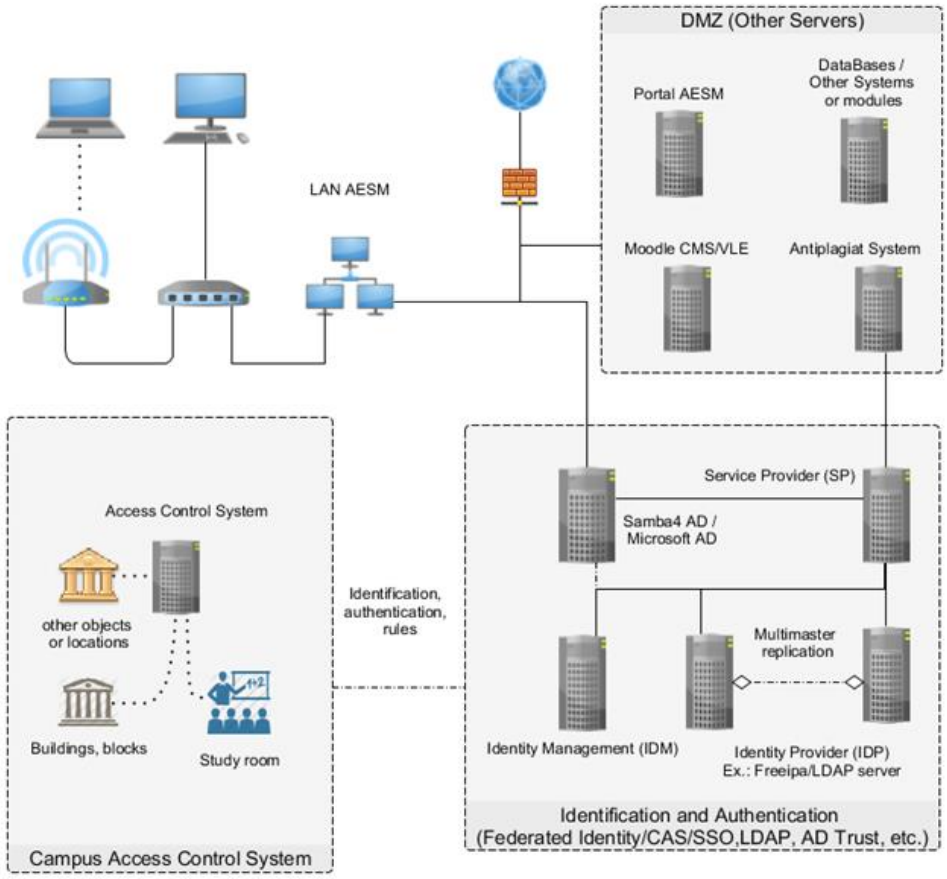



Figure 1: Model of IS used in AESM.

The method of semantic data integration described in this paper has not been used since it requires the organization of both the semantic architecture (semantic server and related services), as well as its support. Based on [3,4,5] and own practical experience a modular organization of IS was selected, fully depending on the direction of work.

This approach allows organizing the flow of information only between certain units of IS without the use of intermediate services, as illustrated by the example of integration servers in fig.1. At the same time, the dependence on a particular server and the risk of compromising itself as well as the dependency on errors in the implementation of the server / service decrease.

The server LDAP was originally selected for identification and authentication, allowing the use of a single identification system and its development in conformity with the requirements given in [7,8] using [11].

All work on the integration of data, and certainly the IS that use the data must be carried out not only according to the internal preferences as the opinions and preferences of developers, users, administration, but also compulsorily to the recommendations and

standards that are used to exchange information between universities and other scientific or educational centers. In conformity with the concept that we implemented, we have tried to use and develop available IS, and, given the opportunity, rely on the open-access developments and foresee the possibility of exchanging data with other systems, in strict accordance with existing laws, regulations, instructions, certain requirements (for example, security).

## References:

1. Гаевская Е.Г. „Теоретические аспекты классификации электронных учебных ресурсов”
2. Шибанов С.В., Яровая М.В, Шашков Б.Д., Кочегаров И.И., Трусов В.А. „Обзор современных методов интеграции данных в информационных системах”
3. Вичугова А.А., Вичугов В.Н., Дмитриева Е.А., Цапко Г.П. „Методы и средства интеграции информационных систем в рамках единого информационного пространства проектирования”
4. Комаров С.Н. „Интеграция информационных систем НОУ на основе персональных данных”
5. Бубарева О.А., Попов Ф.А., Ануфриева Н.Ю. „Решение проблемы интеграции данных при построении интегрированной автоматизированной информационной системы вуза”
6. Хорошилов А.В., Селетков С.Н., Днепровская Н.В. „Управление информационными - ресурсами”, Москва „Финансы и статистика”, 2006г., ISBN 5-279-03168-2
7. EduGain http://www.geant.net/service/eduGAIN/Pages/home.aspx
8. EduRoam https://www.eduroam.org/
9. CMS/VLE MOODLE http://moodle.org
10. Crot Plagiarism Checker Plugin. https://moodle.org/plugins/view.php?plugin=plagiarism_crot
11. Security Assertion Markup Language (SAML) http://xml.coverpages.org/saml.html

# Special Aspects of Enterprise Resource Planning Solutions in Ukraine

**Vasyl Grygorkiv[1], Andrii Verstiak[2], Mariia Grygorkiv[3], Svyatoslav Ishchenko[4]**

[1] Chernivtsi National University, 2 Kotsjubynskyi Str. Chernivtsi 58012, Ukraine
vasyl.hryhorkiv@gmail.com

[2] Chernivtsi National University, 2 Kotsjubynskyi Str. Chernivtsi 58012, Ukraine
a.verstyak@chnu.edu.ua

[3] Chernivtsi National University, 2 Kotsjubynskyi Str. Chernivtsi 58012, Ukraine
gmvmarichka@gmail.com

[4] Chernivtsi National University, 2 Kotsjubynskyi Str. Chernivtsi 58012, Ukraine
isv.emm@gmail.com

**Abstract.** The paper reviews peculiarities and current trends in Ukrainian ERP market. The IT market structure is investigated. There are suggested strategies in accessing the Ukrainian IT market with a view to minimize potential risks and delays. Basic problems of ERP solutions are tracked. The article considers the special aspects of ERP implementation at the Ukrainian enterprises based on IS KASKAD. Basic principles of KASKAD solutions are underlined. The own developed technologies "ObjectCore" and "Kaskad API" are presented. Besides it is outlined important tasks for future pilot projects.

**Keywords.** ERP solutions, ERP implementation, IT market structure, Enterprise Information Systems, Personalized Information Systems.

## 1.Introduction

Transformational conversion of the Ukrainian enterprises caused by the market economy reforms and globalization processes is of paramount importance due to singing the Association Agreement with EU. A great number of the Ukrainian enterprises has been owned by transnational companies or have drawn actively foreign capital investments. It stipulates a certain change in enterprise management approaches which are alike those of Western managers'. It concerns, largely, the implementation and use of modern information systems and technologies.

In real problems of IT-management and ERP implemenetation in UKRAINE are explored only by ERP-vendors and consulting companies such as: SAP, Microsoft, GAAP, IDC, CONIM etc [1].

The answer as for expediency or non-expediency of enterprise automation has been known for a long time. However the problems of the choice of the kind of informational system, the way of its designing, programming and implementation, the degree of automation of business processes and information system integration in Ukraine are still open for the discussion.

## 2.Current trends in Ukrainian ERP market

Actually all national enterprises have started automation with the implementation of accounting informational systems, using typical, the so-called "box" solutions which were very useful at the initial stage partially meeting the demands of management. Thus, it has been successfully completed the computerization of the financial accounting including the automated formation of bookkeeping and calculation of taxes. More perspective seems to be with managerial accounting which under the conditions of automated informational systems is based on practically unlimited analysis and the possibility of operative formation of accounting which, in its turn, is calculated on the basis of absolute and reliable bookkeeping information.

But the problems of automation of supplies, manufacture, distribution, logistics, personnel management, basically the problems constituting the core of enterprise activities and profit are still remain open. Sometimes certain spheres are automated. Such an approach allows some advantages for a certain employee which enable to reduce the number of mechanical errors, the number of routine operations, accelerate their performance. If it is possible, the attempts are made to export/import the data among different informational systems which are to automate different tasks by different technologies. For example, there are often the situations when the department of planning and financing imports bookkeeping of accounting systems in a tabular redactor Microsoft Excel and employing its standard potentialities conducts a certain analysis and calculates the massif of these or those indices. The use of the local approach to the automation problem does not permit to consider comprehensively all the processes that take place at an enterprise and to analyse them from the point of view of systemic analysis.

For Ukraine nowadays, the tasks of the economic growth, increasing of production quality and competitiveness, becoming a member of the world market are most topical. The solution of these tasks without the increase of the quality management and automation of all spheres of economy (state bodies of management, different enterprises and small businesses) is impossible.

The major problem is non-stability of running business in Ukraine which often results in the absence of real long term strategic plans of the enterprise development. This is a real problem for the use of informational technologies. The fact is that while the informational technology is in use only to solve local tasks and improve some business processes, it does not require the strategic approaches and fits completely the common pattern of national enterprise management when its activities are planned maximum for a year period. When an enterprise wants to implement the ERP system in full scale and not for solving local tasks, then may arise some problems. And the implementation of such a system is a constituent of the business development strategy. If this factor is not taken into consideration by the company the negative consequences are evident. The only way out for this is to agree the strategic goal of the company with the purposes of implementation of the informational system. In any other case the purposes ERP system implementation will have a declarative character, thus the project is doomed to be a failure.

IT sector is a multibillion dollar market and one of the most important export categories of Ukraine. IT industry is highly competitive at the global stage and will continue being a major driver behind Ukrainian economic growth over the next 10 years.

The IT market of Ukraine has been characterized by the consulted Ukrainian experts [2] as the one market sector in Ukraine that has been developing with highest capacities and speed compared to international standards over the last years. Within the IT sector the telecommunication market has been described to cover the biggest segment with 7% of the

annual turnover followed by the hardware market with 25%. The third position is being taken by services with 9% including software developers and outsourcing. A considerable part of the market with 5% of the annual turnover is generated by game developing enterprises in Ukraine [3].

General overview of Ukrainian IT is presented at fig. 1.


**Fig 1. IT industry of Ukraine: overview [4]**

IT-sphere growths very dynamically in Ukraine. The next figure shows part of IT to GDP over last 6 years. Based on the calculations that projected growth in domestic IT industry (35-40% per year) by 2015 we will have 8% of IT to GDP.

The speed of growth of IT-market outperforms GDP growth. This means that the tendency of saturation of the global market, are not typical for Ukraine (Fig. 2).

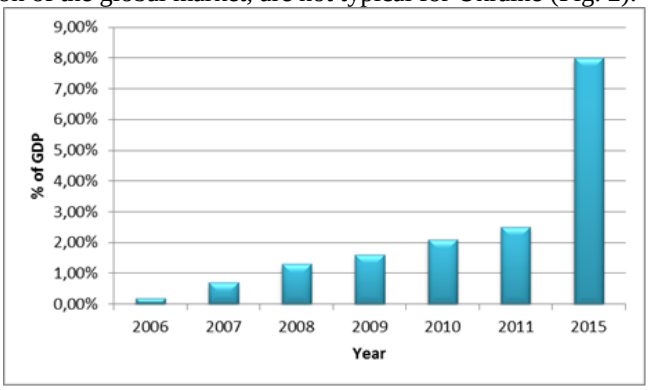



**Fig. 2 IT-sphere in Ukraine [5]**

The difference between the issues solved by ERP in the West and in Ukraine, the same as between the goals of a crowded passenger transport: those who are inside want to get a seat as comfortable as possible, while those outside - to squeeze the least at the first stair. From the beginning of 2008 to the end of 2010 Ukrainian IT-market has experienced one of the most difficult periods. Vendors and consulting companies gave an unprecedented discount, offered attractive options of purchase and payment in installments, but in most cases it was extremely difficult to find new customers for implementing ERP-systems. As a result, the sector in terms of money last year decreased by 60-80%. Now it is time to

recovery industry. However, it is not the same as before. Segment of theenterprise management system has undergone many changes in quality.

The market of ERP-systems is the part of the information and communication industry, which, in turn, brings together the two main sectors: Communications and Information Technology.

In the early 2000s, the domestic IT market, contrary to global trends, grew at an accelerated rate. In this period ERP-systems were one of the most attractive segments of IT-market. The market of corporate software in Ukraine is also influenced by general economic situation.

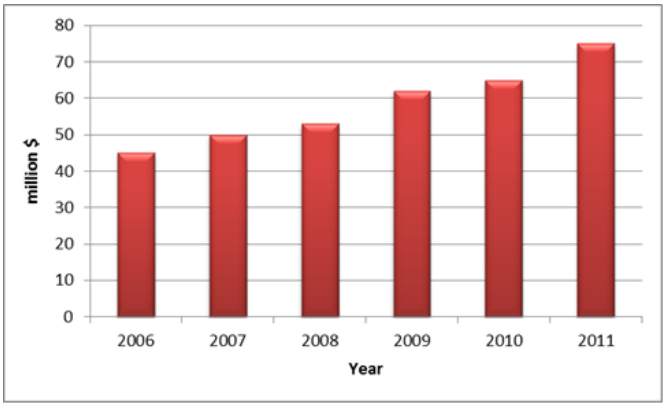



**Fig. 3 Ukrainian market of ERP-systems, [5]**

A considerable part of incomes within the ERP market is achieved by direct sales. At the same time a significant growth of income from license sales, involved systems and IT consulting support has been. The license sales contributed significantly to the dynamic growth of 2010 and the increase of incomes from technical support of constantly growing clients base was especially pointed out as noticeable for those companies that have been long established within the Ukrainian market, and were able to form a substantial number of customers.

The high demand originating from Ukrainian companies for ERP systems will be continued all through 2014 and even beyond. It was stated that the important national economic growth in the majority of sectors combined with the estimated political stabilization could stimulate the interest of foreign and native investors for long-term investments into production in Ukraine. It was highlighted that one important factor of increasing ERP system demand is the currently ongoing process of consolidation inside of several economic branches. Newly groups of companies and holdings need sufficient software and licenses in order to allow them to integrate different information systems into one management control system. According to estimations of the Ukrainian IT consulting enterprise IDC a high potential of growth have such ERP systems in industrial sectors within sales and distribution, banking sector and governmental and public administration. The speed of growth of these sectors will be higher than Ukrainian ERP market growth in general, which has been estimated to reach 26% per year until 2015 [6]. The governmental and public administration sector has been considered to be the main customer of ERP systems in terms of license volume. Governmental institutions show a lot of interest to systems of corporate management. The realization of large IT projects in customs and rating authorities is expected within a short timeframe [6].

We predict a rise of native enterprise management software with additional functionality (CRM function, supply-chain management, workflow-management etc). Similar to the ERP systems market we predict this development to come up in the very near future. Some Ukrainian companies are already focusing on this IT market segment such as "Parus" and "Galaktika". The Ukrainian experts highlighted the continuing change of demand from universal management systems towards very specialized industry solutions that take into account peculiarities of the economic fields and involve typical resolving applications of industry specific tasks.

## 3.Ukrainian ERP: IS KASKAD

Due to the high cost of ERP systems the Ukrainian enterprises use the similar systems, the IS KASKAD being the one, for the automation of big and middle enterprises to increase the effectiveness of their work.

The following basic principles underlie the IS KASKAD system:

• the use of means of effective processing and data protection;

• module principle of making the program complex to achieve the staging of implementation;

• decision scaling for further system development due to the increasing demands of the enterprise;

• flexible system of the data organization in combination with the own development system and business logic programming;

• possibility of integration with the systems and improvements of other manufacturers.

IS KASKAD is realized on the basis of modern technologies with the use of RAD, WEB, OLAP, principles of ERP and CRM. The functional completeness of the system, flexibility of setting and adaptation permit to obtain a peculiar solution for each enterprise taking into account in full its peculiarities of organization and functioning. The core of AIS KASKAD is a centralized data base based on DBMS ORACLE which is built on customer-service architecture with the usage of modern technologies of information processing and preparation of electronic documents. IS KASKAD employs its own object model of storing data (IntelligentCore), which has a built-in system of fast making of business attachments (RAD), modern programming languages being in use (C, Java, Pascal + PL/SQL) as well as API KASKAD. The data base and the system of program modules which ensure the access to the data for the user together with the program interface constitute the core of AIS KASKAD. All the applied systems are on the basis of the core that ensures the information completeness and protection.

As a result we have got the system having some key advantages.

Fast adaptation of the system to the needs of the client: due to the own developed technologies "ObjectCore™" and "Kaskad API" we managed to remove the basic contradiction of modern automation systems:on the one hand, the usage of object-oriented approach while compiling the program, on the other hand, the usage of the relational data bases for information storage these programs work with. The IS KASKAD™ system completely uses object approach both in the programs and the data these programs operate with. In other words "ObjectCore™" technology gives a possibility for the developers and the users to deal with completely objective data including the hierarchical data structures, imitation, etc. Such an approach enables the developers and the users to speed up two, three times the speed of the development of a new functional system and ten times more the speed of the modification of the system functionality (for example the change of the hierarchy of

object structures with the preservation of their data and so on) in comparison with standard methods of development..

The simple and understandable interface ensures quick learning of the users. At the first stage it is not necessary for the user to possess the knowledge and skills in modern basic software to work with the system. In some days of work the user can easily cope with the system in his/her area. In some cases a complete simulation of an interface is done for a specific work place. These technologies enable to considerably accelerate the system implementation, quick change and addition of other automated sectors for the user.

Preservation of confidentiality. It is an important property of the system where the information of the enterprise is stored with which many users work. The conception of distributing of powers of the users and roles underlie the basis of the AISC «KASKAD™»system. The system regulates the access to all the information and functionality. To be precise, the system enables also the possibility of blocking/unblocking of objects besides the standard rights which gives the complete control of the sector under activity. For example, after the document printing it is automatically blocked of changes, or the user has the access to the document instead of its deleting. The system also has a complete audit as by the system on the whole so by the users.

The possibility of the expansion of functionality is an imperative property of the system which is aimed at the effective durable usage. The IS «KASKAD™» system is built on the basis of open standards, all business logic is provided by the complete output code. The employment of such high level programming languages as PL/SQL, Visual Basic for Applications, JavaScript, Kaskad, FastScript permits easily, quickly, in short, and reliably realize all the necessary functionality of the system. With the help of the program utilities there is a possibility of automatically transfer of the functionality and the usage of its potentialities in the system compiled by other manufacturers. At the same time in contrast to other systems due to the available applied program interface access to the core (Kaskad API), the uniform data structure for all the clients is preserved which enables to update regularly and that constantly improves the functionality of the system. Due to the technique of updating the life cycle of the system is increased. One more advantage of the independent expansion of the system is the possibility of not giving the information to the outsiders, i.e. preserving confidential information.

The work with large size data base. It is the requirement resulting from the aforementioned advantages. The ideal situation for this is when all the stored information is accessible with minimal expenses in a uniform data massif. It allows analyzing the information in a static diapason of the time period. The data development is of extreme use under the stabilizing of economy of Ukraine. The IS «KASKAD™» system incorporates the specialized software (OLAP), which enables the analysts and the leaders to analyze big amount of the data by means of the fast interactive their presentation and different levels of their peculiarity.

The low cost of its acquisition. It is one of the most important properties of the IS «KASKAD™», system which is connected with a program and technical architecture and the chosen platform. The total cost of the acquisition of the system is formed from buying expenses, implementation and application of software, equipment and the system of the data transfer. The decrease of the system acquisition cost is made of the buying, implementation and exploitation of the software, equipment and the data system transfer. The decreasing of the cost of the IS “KASKAD” system is made of the following:

1) The object model of the data preservation on the basis of IntelligentCore gives the possibility to the developers and the users to deal with the object data using the hierarchical structures and imitation.

2) The usage of the hierarchical model of the data/

3) The possibility of arbitrary expansion of the data structures in the process of implementation and exploitation of the system.

4) The built-in system of the development with wide possibilities of programming permits in accordance with the needs to manipulate the data structures on the spot as well as their processing (business logic). The AIS KASKAD system employs in full degree the object approach both in subprograms and in the data these programs operate with;

5) The built-in system of the accounting formation which is successfully implemented at many enterprises. The AIS KASKAD system gives a possibility to automate complex business processes quickly and qualitatively.

## 4.Conclusion

As follows, we can make the general conclusion that there is huge IT industry in Ukraine; the country is an important player within World IT-market and it is #1 in Central and Eastern Europe by IT outsourcing market size and IT talent pool.

Main trends of the Ukrainian ERP-market:

- the economy and political instability will not stimulate the interest of both foreign and domestic investors to long-term investments in ERP;
- an important factor in the demand for ERP-systems is ongoing processes of consolidation within industries: the newly created group of companies and holdings need software that allows integrating disparate information systems into a single management tool;
- in the near future ERP-systems market is expected to increase the number of domestic management software company, which sold more functionality (CRM-functions, supply chain management, etc.).

Due to the high cost of ERP systems the Ukrainian enterprises use the similar systems, the IS KASKAD being the one, for the automation of big and middle enterprises to increase the effectiveness of their work.

## References

1. Verstiak, A., Ukrainian IT-industry charakteristics, (Vol. 3 (6), 79-84). Bulletin Chernivtsi Trade Institute. Economics, 2012.
2. Matyas, M., Director SAP Ukraine. Interview, 2012, (http://itbusiness.com.ua/it-business-week/interview/31258--sap-.html)
3. Kuhar, A., Ukrainian market of business applications: it's time to SMB, 2013, (http://ko.com.ua/ukrainskij_rynok_biznes-prilozhenij_prishlo_vremya_smb_50252)
4. Yevgen Sysoyev, Strategic National Project “Innovative Ukraine”, 2014, (http://ciklumsalesoffice.com/news/ukrainian-it-community-calls-for-government-reforms-to-help-in-creating-innovative-ukraine/)
5. Anon, IDC investigations, 2012, (http://www.idcukraine.com)
6. Anon, Implementation of ERP-solutions, 2012, (http://corp.cnews.ru/news/line/index.shtml?2013/12/29/421833)

# HDFS – The File System Suitable for Large-Scale Analytics

**Vladimir Dimitrov**

FMI, University of Sofia, 1164 Sofia, 5 James Bourchier Blvd., Bulgaria
cht@fmi.uni-sofia.bg

**Abstract.** Big data emergence provokes the great interest on big data analytics tools. Pioneer in this area is Google, who with its solutions influenced the open source project Hadoop. The Hadoop Big data storage is HDFS.

**Keywords.** Big data, distributed file systems, Hadoop, HDFS.

## 1. What is Hadoop?

Apache Hadoop is an open-source Java-based framework. It stores data (Hadoop Distributed File System - HDFS) and executes jobs (MapReduce) on large clusters of commodity servers. Hadoop supports a high-level of fault tolerance. This framework is very simple but effective for a large class of big data applications. It is scalable from a single server to thousands of servers.

The main considerations in big data analytics can be defined by the following questions are:

1) How big data is stored?
2) How big data is analyzed?

Traditional database systems have limitations on the number of columns, table size, etc. Usually, they require data to be preformatted before to load them in the database. Big data, usually, are not formatted. A big data file could store many terabytes and has billions of fields. A special storage for big data is needed.

Traditional advanced analytics tools are based on relational model of data and N-dimensional cubes. The first approach is to extract data from the big data storage and format them for input to traditional analytic tools. The second approach is to develop big data analytic tools directly on big data storage. It is possible, data extraction to be combined with data processing.

Hadoop combines big data storage and big data analytics. Hadoop Distributed File System is the solution of big data storage and it is discussed here. Hadoop MapReduce is the solution of big data processing and is discussed in the next chapter.

A solution to above mentioned problems is Hadoop. Usually, when a new technology emerges, there are no standards, but only leading solutions, projects, products, practices etc. Such a leading project in the big data analytics is Hadoop. It combines together data storage and data analytics. Hadoop Distributed File System (HDFS) is the Hadoop's solution of the first problem (Big data storage) and it is discussed here. Hadoop MapReduce is Hadoop's solution of the second problem.

## 2. HDFS Design Concepts

Hadoop flagship file system is Hadoop Distributed File System (HDFS). Hadoop could be linked to any file system that support its abstract file system interface. HDFS is an

implementation of Hadoop abstract file system, but it has special place – HDFS is optimized for MapReduce jobs that are core part of Hadoop.

HDFS is suitable for big data analytics. It is optimized to write once very big files and then to read them many times. HDFS is not suitable for random reads/writes. Hadoop complements modern database systems – they are optimized for random I/O operations.

HDFS is a general purpose distributed file system for very large files. Here, under the term "very large data files" must be understood files of many terabytes and petabytes in size. HDFS could store tens of millions files on a single instance. It is designed to run on hundreds or thousands commodity servers. A general assumption about HDFS is that hardware failure is a norm, but not an exception.

HDFS is designed for applications that consecutively access (streaming) the data sets. HDFS is optimized to write once its files and read them many times. It is not suitable for small files and for direct reads and writes. Applications that run on HDFS use batch processing. HDFS is optimized for high throughput – not for fast access to data. Interactive applications on HDFS are not the standard use case, but such applications could use HDFS under special circumstances. Applications developed in MapReduce framework on HDFS data sets are the primary use case of Hadoop.

HDFS commands are POSIX like. It does not support all POSIX requirements, because its applications does not need them.

HDFS file is a sequence of blocks than are stored in the cluster on multiple servers and on multiple disks. Fault tolerance is supported at block level. HDFS blocks are big ones – 64 MB by default. If a file is smaller than a block it share the block with other small files.

Leading Hadoop idea is to move the computations to the storage nodes but not to move the data to the computing nodes. This approach is sensible when computing programs are relatively small and stored data are big enough. In this case, the network traffic decreases.

HDFS tries to distribute file blocks among the cluster servers, but at the same time, the size of the file blocks have to be big enough – bigger than the size of the computing programs to decrease network traffic. The maximal parallelism could be achieved if almost all the cluster servers are involved in the computations, but this means that the data set files have to be big enough to be distributed on almost all the cluster servers. Hence, the optimal files for Hadoop are the big one dispersed on large clusters. The file blocks have to be bigger than computing programs and the file blocks has to be evenly distributed on almost all cluster servers – that is the key for Hadoop success.

HDFS is responsible for the even distribution of the file blocks on the cluster servers, but it is not responsible for moving computing programs to the data. The last responsibility is to the MapReduce framework.

## 3. HDFS Architecture

HDFS architecture is given at Fig. 1. There are two kind of HDFS nodes: namenode and datanode. The namenode contains metadata for the file system: files, directories, file block locations etc. The datanode stores data blocks.

The distance between nodes is defined in the hierarchy: in the same host, in the same rack, in different racks. That metrics is used to calculate when a replica is nearby.

The namenode is crucial for the file system existence – if it fails then all the file system data is lost. That is why there is an option for secondary namenode to be run. The last one collect all the information needed to restart a new namenode in case of failure.

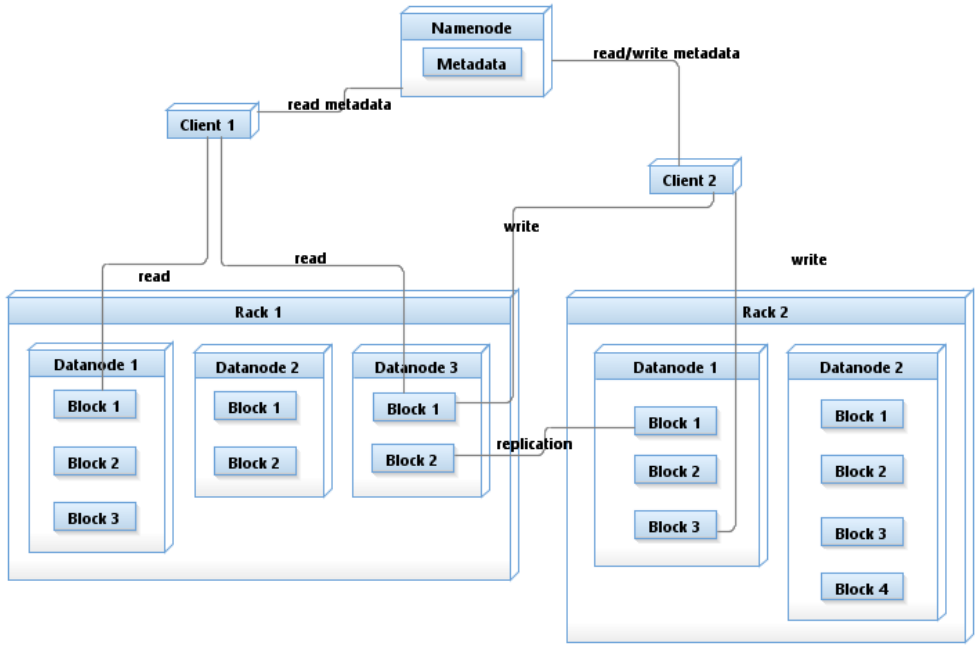



**Figure 1.** HDFS Architecture.

In HDFS version 2.xx, HDFS Federation is introduced. It permits the cluster to be scaled. New namenodes could be added. Every namenode manages different part of the file system – a namespace volume. The last one contains metadata information for the namespace and a block pool that contains all block for the files in the namespace. Namespace volumes do not communicate with each other. One datanode can store blocks from multiple namespace volumes. It contains information about that.

In case of namenode failure, a recovery can be done in 30 minutes or more. That is an HDFS problem – the high availability. The good news are that there are plans for future implementations to resolve this problem with a solution based on ZooKeeper. This solution automatically switches to secondary namenode and fences the failed node for all operations that can damage the data.

The client opens files or directories using the metadata from a namenode. After that, all operations are performed directly with the file's datanodes. In case of read operations, datanodes are accessed directly in sequential read access mode. In case of read failure a block replica is read. The failed block is marked for further recovery operations. It is reported immediately to the namenode before to continue with further reading. When all data from a given block are read, HDFS chooses the next block among all replicas – it must be closer to the client. An example of sequence of read operations in HDFS is given in Fig. 2.

When write operation is performed the block is piped through all file data blocks to the end – write operations are in sequential file access mode. All write operations are queued in the client (FSDataOutputStream). Success of write operation is piped backward to the client and removed from this queue. If a write operation fails then that information is returned to the client, then recover actions are performed to remove failed datanode and to redo all not acknowledged operations. An example of sequence of write operations is given in Fig. 3.

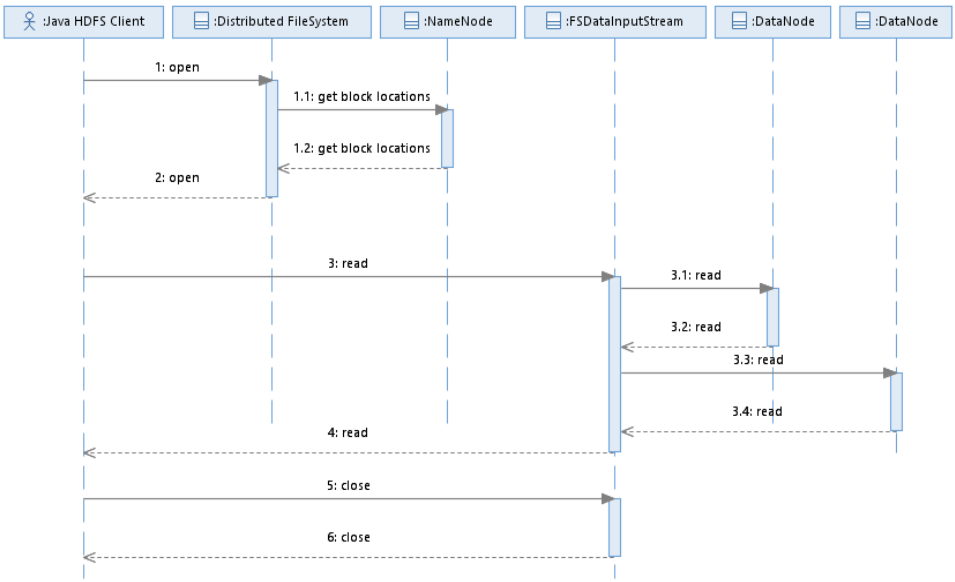



**Figure 2.** Read operations in HDFS.

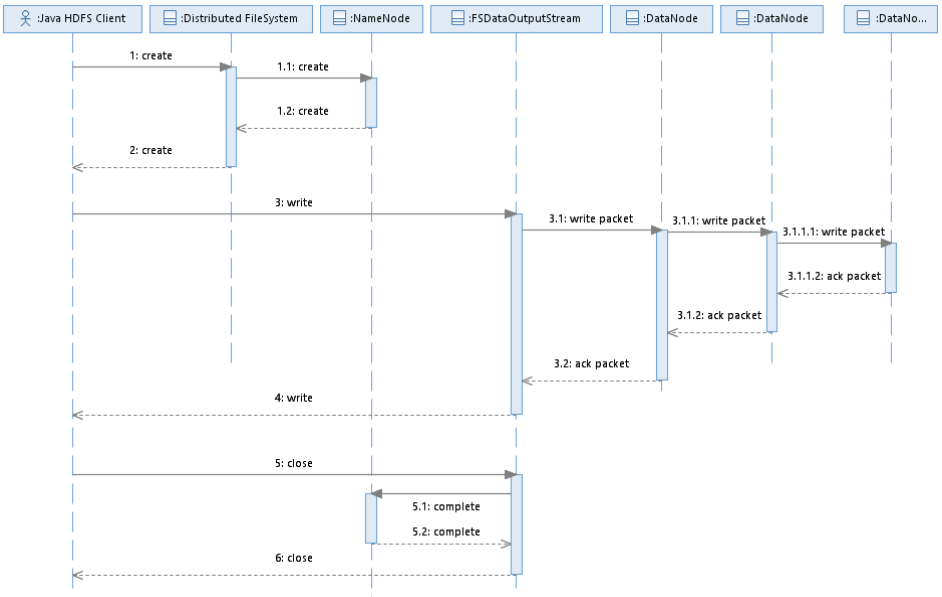



**Figure 3.** Write operations in HDFS.

It is possible to put in synchronization points among blocks to start reading from some such an intermediate point.

There are many interfaces to HDFS that can be used to access data, in Java, HTTP, C, FUSE etc.

Hadoop support two kinds of file based data structure: SequenceFile and MapFile. The first one is suitable for map inputs. The second one – for reduce inputs (see next chapter for reduce inputs). The MapFile is a kind of persistent implementation of Java maps.

There are some utilities for managing HDFS, like Data Ingest for moving large stream data into HDFS and Sqoop for importing structured data stores (relational data bases).

Hadoop archives (HAR files) can be used to package many small files. HAR files are acceptable input for MapReduce programs. They do not support archive compression. HAR archives are immutable, i.e. they cannot be changed – if a file in it is changed then the whole archive has to be recreated.

Hadoop supports file compression, but it is important compression schema to know the block boundaries for splitting. Such schemas are bzip2, and LZO schema with indexing. For Big data, file compression reduces storage needed for the file and speed up data transfers.

Hadoop is a Java implementation. It is important for RPC to support serialization. Hadoop uses its own serialization format Writables. It is compact and fast, but not extensible for other programming languages. Avro is the serialization solution of Hadoop, based on JSON. It is extensible and interoperable.

## 4. Conclusion

Apache Hadoop project is highly influenced from Google works presented in [3] and [4]. It has started as Nutch project [1], but nowadays even at Google teach on their solutions using Hadoop. More details on HDFS can be found in [5] and [6].

Big data are here and Big data analytics tools are available. Hodoop is a driving force behind the Big data industry. Its implementation of Big data storage is a leading example for large scale processing on clusters of conventional servers connected with conventional networking hardware. There are still many questions about Hadoop application areas, but many researchers are working on that.

## 5. Acknowledgements

Research in this paper are funded by Bulgarian Science Fund under contract ДФНИ-И01/12 “Modern programming languages, environments and technologies, and their application in education of software professionals”.

## References

1. Cafarella, M., and D. Cutting. 2004. Building Nutch: Open Source Search. ACM Queue, (April): http://queue.acm.org/detail.cfm?id=988408.
2. Cutting D., 2012. Intro to Hadoop and MapReduce. https://www.udacity.com/course/viewer#!/c-ud617/l-306818608/m-312934728.
3. Dean J., and S. Ghemawat. 2004. MapReduce: Simplified Data Processing on Large Clusters. Google (December): http://labs.google.com/papers/mapreduce.html
4. Ghemawat, S., H. Gobioff, and S.-T. Leung. 2003. The Google File System. Google (October): http://labs.google.com/papers/gfs.html
5. Hadoop. HDFS Architecture Guide. August 4, 2013. http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html (accessed March 11, 2014).
6. Hadoop Wiki. PoweredBy. April 23, 2014. http://wiki.apache.org/hadoop/PoweredBy (accessed April 23, 2014)

# Software for Analysis of Some Aspects of the Economic Activity of Real Estate Companies

**Marin Galabov**

UNWE, 1700 Sofia, Bulgaria
marin_galabov@dir.bg

**Abstract.** The Report focuses on some aspects of the economic activity of real estate companies such as economic efficiency and profitability. The Report presents the key correlations for measuring economic efficiency and profitability and related indicators. The Report is based on the concept that applying proper software enables quick calculation of efficiency and profitability indicators and their comparison for particular periods, which considerably shortens the time for analysing the said aspects of the economic activity of real estate companies. Thus, the management of these companies may make adequate decisions in due time.

**Keywords.** Real estate companies, economic efficiency, profitability, economic efficiency and profitability indicators, software

## 1.Brief description of real estate companies

As real estate companies perform various activities, there exist and operate construction companies, real estate agencies, real estate appraisal companies, facilities and property management companies, special purpose vehicles securitizing real estates and service companies.

Facilities and property management companies, special purpose vehicles and service companies are less known to the public than the other of the companies mentioned above. These companies are briefly described hereinafter.

Facility management performed by facilities management companies includes a wide range of activities such as door-keeping services, cleaning of common parts, maintenance of green areas, snow removal, 24-hour security guard and video surveillance, payment of bills charging electric power and water consumptions costs related to common building parts, maintenance of indoor and outdoor swimming pools and other. Facility management is performed in relation to residential and commercial buildings as well as their surroundings. Facilities management companies ensure comfortable environment to the people living and working in residential and commercial buildings.

Facility management is divided into hard and soft. Hard facility management includes all activities related to building exploitation, servicing and facilities maintenance, while soft facility management concerns all services related to cleaning, security, etc.

Property management performed by property management companies may be described as the search for and finding of lessees of private and commercial properties and management of leased properties. Property management includes the performance of a number of tasks important to owners (owners offering property lease) such as property advertising to find reliable lessees, negotiations with lessees, property stocktaking prior to property lease and after property vacating, payment of monthly bills charging electric power and water consumption, telephone, internet and other costs on behalf and at the expense of

owners or lessees, keeping lessees' contractual deposits, collecting leases from lessees and transferring them to lessors' bank accounts, regular inspections of property condition, etc.

In general, property management involves the search for and finding of property lessees and property keeping instead of owners. A property management company is a company taking care of all property-related matters, saving owners' time and efforts.

It should be noted that a number of companies perform facility management as well as property management.

A special purpose vehicle securitizing real estates (also known as real estate investment trust) is a joint-stock company, which issues securities and invests raised funds in real estates – a process known as securitization of real estates.

A real estate investment trust is registered under Article 163 of the Law on Commerce. The maximal number of founders is 50. Such registration requires the subscription of at least 30% of the capital by institutional investors (collective investments schemes, insurance companies, etc.).

The minimal capital of a real estate investment trust is BGN 500,000 and only cash contributions are allowed.

A real estate investment trust is exempted from corporate tax and allocates at least 90% of its profit for each financial year as dividends.

Service companies are commercial companies, which avails with the organization and resources required for servicing and maintenance of real estates acquired by a real estate investment trust and builds and makes improvements to real estates owned by such real estate investment trust.

The Law on Special Purpose Vehicles does not allow companies securitizing real estates to perform these activities but require their assignment to service companies.

## 2.Economic efficiency and profitability – aspects of the economic activity of real estate companies

Economic efficiency is measured through the benefits/victims correlation:

$$\frac{\text{Economic effect}}{\text{Resources (average availability)}} \qquad \frac{\text{Economic effect}}{\text{Expenses}}$$

The economic effect is a benefit in the form of revenue, profit, etc. The resources are the assets, capital and personnel of a real estate company (or a company in any other business sector). The resources figure is calculated as average availability for the particular period.

These correlations become specific through particular indicators for measuring economic efficiency. The indicators are calculated on the basis of particular economic effects (such as corporate revenue) and particular resources (such as equity), particular expenses (such as costs of services or finished products) respectively.

Some of the indicators measuring economic efficiency are:

$$\text{Ratio of capital efficiency} = \frac{\text{Revenue}}{\text{Capital}}$$

$$\text{Ratio of cost efficiency} = \frac{\text{Sales revenue}}{\text{Cost}}$$

The capital is the total of equity, liabilities and financing and deferred income. The formula includes its average figure.

The sales income generated by real estate companies, except for construction companies, represents proceeds from sales of non-material finished products (services). Construction companies realize revenue from sales of material finished products.

Profitability is the financial result whose realization employs a unit of resources (average availability) or a unit of expenses. It is described as the financial result realized through employing of a unit of expenses as well. It is measured through the following correlations:

$$\frac{\text{Financial result}}{\text{Resources (average availability)}} \qquad \frac{\text{Financial result}}{\text{Expenses}} \qquad \frac{\text{Financial result}}{\text{Revenue}}$$

These correlations become specific through a number of profitability indicators, some of which are:

$$\text{Return on equity} = \frac{\text{Financial result}}{\text{Equity}}$$

$$\text{Return on cost} = \frac{\text{Financial result (sales)}}{\text{Cost}}$$

$$\text{Return on sales} = \frac{\text{Financial result (sales)}}{\text{Sales revenue}}$$

## 3.Software application in economic efficiency and profitability analysis

The software provides real estate companies with a number of options concerning economic efficiency and profitability, including:

- ✓ Quick calculation of economic efficiency and profitability indicators for the period ended just now and their comparison with the figures planned for the same period (for example: comparison of such indicators for October with the figures planned for October);
- ✓ Quick calculation of economic efficiency and profitability indicators for the period ended just now and their comparison with the figures reported for periods during the same year (for example: comparison of such indicators for October with the figures reported for the preceding three months);
- ✓ Quick calculation of economic efficiency and profitability indicators for the period ended just now and their comparison with the figures reported for periods during previous years (for example: comparison of such indicators for October with the figures reported for the same month of the preceding three years).

Quick calculation and comparison enables analysts at real estate companies to prepare efficiency and profitability analyses in short terms. Based on such analyses, which include recommendations to the management of such companies, the latter make adequate decisions in a relatively expedient manner.

The software also offers options for working with pyramid financial analysis models (similar to the DuPont Analysis), which includes economic efficiency and profitability correlations. It ensures quick calculation and comparison of the figures expressing such correlations included in these models. This also helps the management of real estate companies to react expediently.

## References

1. Bhattacharyya, A., Introduction to Financial Statement Analysis, Elsevier, 2009
2. Ciaran, W., Key Management Ratios, InfoDAR, 2008
3. Information products of CIELA NORMA AD, Sofia

# Features of Social Customer Relationship Management Systems

**Violeta Kraeva[1], Petia Emilova[2]**

[1] D. A. Tsenov Academy of Economics, Svishtov, Bulgaria
kraeva@uni-svishtov.bg

[2] D. A. Tsenov Academy of Economics, Svishtov, Bulgaria
petiaem@uni-svishtov.bg

**Abstract.** Social networks are increasingly more intensively widespread not only in personal life but also in business. They are particularly effective in establishing continuous contact with customers, suppliers and business partners. The purpose of the paper is to present Social Customer Relationship Management - CRM - as a new tool for sharing business ideas and relevant information that supports the creation and delivery of personalized and customized products and services.

**Keywords.** CRM, Social CRM, Social Networks, Web 2.0.

## 1. Introduction

We live in a time when for any activity – work, education, entertainment, communication, etc. – people increasingly turn to the Internet and social media. For business organizations the Global Network is a major source of advice when looking for the right partners and customers, and for consumers, when looking for the most appropriate goods and services. Social media is gradually transforming business organizations now competing in the *social market* for the attention of *social customers.*

A social market is created when organizations use social media and typical technologies (Web 2.0), to collaborate and connect with customers, partners and other stakeholders to build communities around products, services and brands. The social markets and the strategies that guide them have been developed by organizations in order to build confidence in customers and partners. Through their involvement, network users can share information, seek feedback and create valuable business knowledge. Dialogues that implement a search process are performed within the social market. Business organizations use social media to aggregate loyal users, to identify key factors and targeted programmes for marketing campaigns, and to participate within the community of users, allowing them to monitor their disposition towards products and services in detail [9].

Social customers [2] expect companies to listen to their problems and worries and use the channels of social media to actively participate in transparent conversations with them. These clients, in addition to the price of the product, obtain satisfaction from the process of Web 2.0 communication with people they trust.

## 2.A Definition of Social Customer Relationship Management Systems

Customer Relationship Management (CRM) uses information and communication technology (ICT) to organize, automate and synchronize business processes such as customer service, marketing and sales. The overall goals are to find, attract and win new customers, retain those the company already has, re-attract previous customers, and reduce the costs of marketing and customer service (Galimi, J. CRM in Healthcare: The Steps to Achieve Business Value, Gartner Inc. Research, 2001).

To meet the social customer's needs, business organizations develop systems for managing relationships with customers. They use Web 2.0 technologies and include social media as a new channel for interaction with customers. In this way a new, social layer aimed at attracting Internet-based communities is added to traditional CRM.

Generally speaking, Customer Relationship Management undertaken via social media is called Social CRM [6]. According to one definition [5], Social CRM is a business strategy for engaging customers through social media with the goal of building trust and brand loyalty.

Social CRM is a *new strategy* supported by a technological platform, business rules, processes and social characteristics designed to engage the customer in a collaborative conversation in order to provide mutually beneficial value in a trusted and transparent business environment [4]. It is focused on customers and continuously attracts customer attention. It should be noted that, at the same time, social CRM represent a very good *tool* which companies actively use to find and contact a large number of customers, to maintain an interactive relationship, and to collect detailed information about them. By being based on social CRM, it's easy to construct accessible centralized databases for clients, track events and coordinate activities, and manage marketing processes.

As a tool, social CRM creates a new business model in which the processes of designing products and services include the customers themselves in the role of active participants.

The characteristics of social CRM are a projection of the opportunities offered by Web 2.0 technologies to business. The main features that distinguish social CRM from traditional CRM [2, 3, 7] are :

- *Focus on collaborative relationships.* Social CRM is focused on collaborative relationships (a more complex relationship network), unlike traditional CRM which is focused on individual relationships (C2C, C2B);
- *Content-driven.* Social CRMs are content-driven, as opposed to traditional CRMs, which are data-driven and mostly store and manage data about contacts and meetings with clients. Social CRM generates content in a variety of formats suitable for the target audience in order to for the organization to make contact with current and potential customers. The main tools used are blog posts, podcasts, YouTube videos, Webinars etc.
- *Multiple connections.* With regard to connections, social CRM provides a better understanding of customers and their community, thanks to multiple connections. As opposed to traditional CRM, it is characterized by a limited view of customers and their community preferences, habits, etc.
- *Conversation-centric*. Another significant difference is that traditional CRM aims to implement, standardize and automate the customer interaction process, while social CRM is conversation-centric. A social CRM strategy is focused on inducing

meaningful conversations with current, and especially with potential, customers seeking information and assistance.

- *Conversation generates value*. Unlike traditional CRM which uses targeted messages to generate value, social CRM generates value based on conversations.
- *Community focused*. While traditional CRM focus on operational efficiency, social CRM focuses on people and communities. It is used with popular industry bloggers and social sites which the audience frequents. This aims to induce and help conduct transparent discussions and follow discussions in the community
- *Automation.* Automation aims to increase opportunities and reduce the time for periodically generating content, primarily from social media. To drive targeted traffic to our sites tools and methods are being used such as Web site optimization, search engine marketing and others.
- *Analysis.* This means targeting the right audience in the most appropriate way to analyse various aspects of customer-generated content and its form. In this way, you can understand how potential customers use the Internet to communicate, gather information, participate in discussions, and select whom to trust etc. For example, to determine what is the most influential form for providing information to the client – e.g. e-mail, blog, Facebook page.

## 3.Issues in constructing a social CRM

Social CRM systems are in the initial stage of construction and development. We need to look for ways to use them in an effective and efficient manner. These systems can help business in different ways. The main ones are [10]:

- Directing marketing to key segments based on information about the customer's interests gathered from various social networking sites. Marketers use this information to create highly focused messages and suggestions in order to motivate customers;
- Detecting potential customers and social CRM's impact on them, and identifying influential customers. This is accomplished by following comments about the company's products and services, and also those of competitors by using appropriate keywords;
- Controlling customer opinions in order to develop an appropriate strategy and policy for the further development of the company;
- Monitoring and countering negative comments in order to be able to provide an immediate answer to them;
- Quick polling and detection of problems, together with advice for their solution;
- It is a tool for promotions, winning trust and friendship by posting photos of products, services and staff so that they can emerge from anonymity

The development of social CRM is accompanied by numerous challenges that organizations have to deal with. The main challenges relate to [1]:

- The informal and heterogeneous nature of content generated by social media.
- The identification of relevant posts, because identification by keywords and popular simple techniques cannot provide enough precision.
- The identification of illegal posts.
- Priority determination in order to achieve the optimal utilization of customer service.

## 4.Conclusion

Social CRM systems are highly competitive within the rapidly changing digital world. They help businesses to find, connect and stay in touch with a large number of customers and collect information about them. This creates a new model that alters the design of products for customers as they are designed jointly by the customers themselves. This is a new strategy that focuses on demand and attracts customer attention.

To use these systems in an effective and efficient manner it is necessary to explore the ways in which they support business and the challenges of such use.

## References

1. Ajmera, J., Hyung-il Ahn and all. A CRM system for Social Media [Challenges and Experiences], International World Wide Web Conference, May 13–17, 2013, Rio de Janeiro, Brazil. ACM 978-1-4503-2035-1/13/05
2. Brent Leary. Social CRM: Customer Relationship Management in the Age of the Socially-Empowered Customer, A CRM Essentials, LLC, White Paper, October 2008.
3. Fabio C. Social CRM; Concept, Benefits and Approach to adopt, December, 2010, http://www.slideshare.net/Subbuinblr/social-crm-by-fabio-cipriani- presentation
4. Greenberg, P. (2009). CRM at the Speed of Light, Fourth Edition: Social CRM 2.0 Strategies, Tools, and Techniques for Engaging Your Customers, 4th ed. McGraw-Hill Osborne Media.
5. Kotadia, H. Definition of Social CRM – Explained, 2010, www .customerthink.com/blog/defi nition_of_social_crm_explained
6. Mohan, S., E. Choi, and D. Min. Conceptual modelling of enterprise application system using social networking and web 2.0 \social CRM system". In Proceedings of the 2008 International Conference on Convergence and Hybrid Information Technology, ICHIT '08, pages 237{244, Washington, DC, USA, 2008. IEEE Computer Society.
7. Muhammad, A., M. Almunawar. Framework of Social Customer Relationship Management in E-Health Services. Journal of e-Health Management, Vol. 2012 (2012), Article ID 766268.
8. Woodcock, Neil and all. Social CRM as a business strategy, Journal of Database Marketing & Customer Strategy Management (2011) 18, 50 – 64.
9. Емилова, П. Информационните технологии в дейността на съвременния работник или работещият със знание. Човешкият фактор в управлението: Юбилейна научно-практическа конференция - Свищов, 12-13 октомври 2012, том 1, с. 160-165.
10. Краева, В., П. Емилова. Приложение на социалните системи за управление на взаимоотношенията с клиентите. Съвременни методи и технологии в научните изследвания: Сборник доклади от международна научна конференция - Варна, 2013 г., с. 91-96, Наука и икономика ИУ - Варна, 2013.

# Distributed Version Control Systems: Concept, Models and Perspectives

**Veselin Popov**

D. A. Tsenov Academy of Economics, Svishtov, Bulgaria
popov@uni-svishtov.bg.bg

**Abstract.** The report examines the version control system (also known as a Revision Control System), which allows different forms of distribution, collaborative development by monitoring the access of the source code of computer programs. The basic concepts of version control system are presented, the two main types of centralized and decentralized version control systems (DVCSs) are compared. DVCSs are reviewed in detail - their concept, features, types of architectural models of DVCS, suitable for different projects and teams.

**Keywords.** Version control system, distributed version control system, architecture.

## 1. Introduction

The Version Control Systems (VCSs) aim to provide simultaneous work of many people on a single project. This allows each developer to make changes, which can then be shared with the rest of the team. VCSs provide access to make a backup copy when they need to be restored or reviewed.

VCSs were originally developed to maintain the program files developed by teams of programmers. Later they began to be used for storing Web content and various other types of documents. Modern VCSs have full support for binary files such as images and others.

## 2. Conception of Version Control Systems

VCSs should be able to provide three important features: reversibility, concurrency, and annotation. Reversibility is the most important opportunity that VCSs must maintain. They must keep back up of a stored well-functioning state in order to eliminate any unsuccessful changes. Concurrency enables many people to work simultaneously on a collection of code or documents as the conflicting modifications must be able to be deleted or to resolve any arose conflicts. VCSs must maintain annotation, in order for comments and feedback for any change to be written down – what it contents and who is responsible for it.

VCSs provide simultaneous real-time work on a project of multiple users. To achieve this participants are provided with a working copy of the project files which they use. Furthermore, a repository of the files of the entire project is maintained, in which each individual user sends the modified files in order for them to be accessible to all users. This organization [1] is shown in Fig. 1.

Repository is a database that can be compared with core of VCS. The information in the data repository is stored in a hierarchical structure in the form of files and folders. The files and folders are updated by the working copies to users.

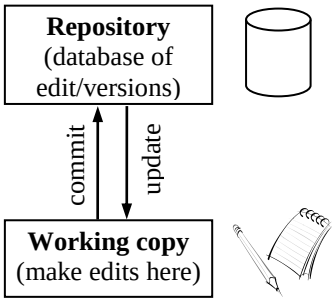


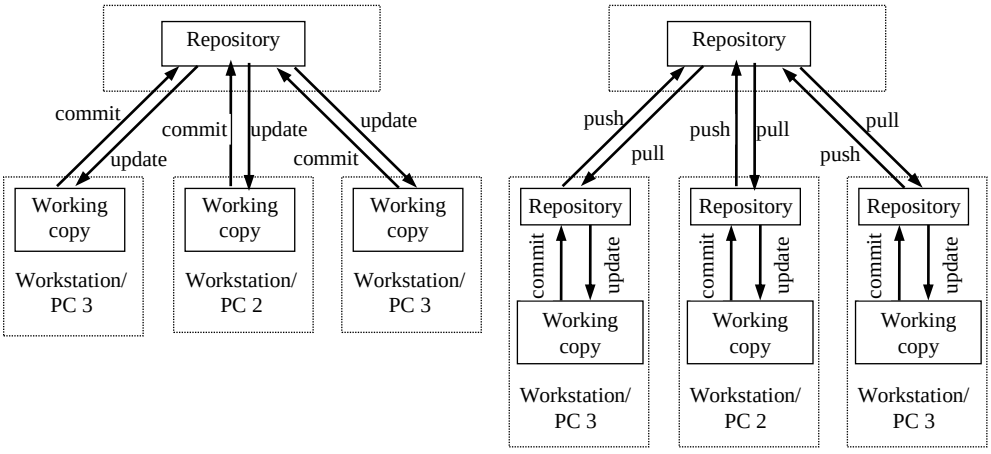

Fig 1.

Source: Ernst, M. Version control concepts and best practices. https://homes.cs.washington.edu/~mernst/advice/version-control.html

The working copy is a local copy of all data with which the current user works. It is usually located on the local disk. Changes in the working copy are sent by the user in repository.

## 3. Transition from Centralised Version Control Systems to Decentralised Version Control Systems

VCSs first appeared in the 80s of the last century. Their development goes through several phases. The first phase is characterized by the use of local systems that do not work in the network. They do not provide opportunities for competitive work on the same file, as they lock the file on which it is worked. The second phase systems are networks which work on the client/server model. These systems are centralized and they provide competitive simultaneous work on multiple files. The last phase are distributed systems which provide competitive work on changing sets that have previously checked for possible conflicts and require confirmation before performing the merge.




Centralized version control Distributed version control

Fig. 2 Difference between centralized and decentralized version control systems (Ernst, M. Version control concepts and best practices. https://homes.cs.washington.edu)

When centralized systems are used (see fig. 2) users have their own working copy and they all have one central repository. The principle of operation of centralized systems is the following:

- Everyone from the team who has completed changes in the source code have to commit them to the central repository;
- The other users of the team must perform an update, after which they receive the underlying code changes from the central repository.

In DVCSs all users have their own repository, which is independent of the central one. The principle of operation of DVCSs is implemented in 4 steps:

- The user commits the changes in the user's own repository;
- The user pushes the changes from the distributed to the centralized repository;
- The other team members pull the changes from the centralized repository in their own repository;
- Then they perform the update.

The principle of operation of CVCs is simple, similar to the client-server model, but problems arise when they are used in large projects and in teams whose members are physically separated.

# 4. Conception of Distributed Version Control Systems

DVCSs are the next generation in the development of systems for version control. They have similar concepts, such as the CVCSs but they also have a number of differences. DVCSs have local repository and can operate without being connected to the central server.

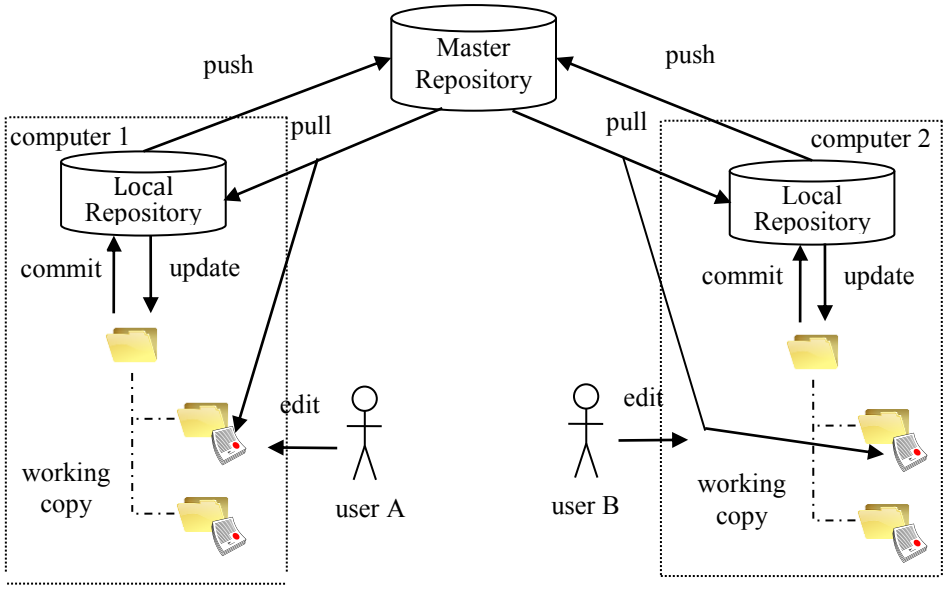



Fig. 3. Basic concept of DCVS (Adapted by: Version Control. http://pages.cs.wisc.edu/)

In the concept of the basic version of DVCSs there is a repository and a working copy. In **repository** database with previous versions of source code is stored. Information between repositories can be exchanged by pushing or pulling. If user A performs **pushing** the others working on the same project (user B, ...., User N) will get the changes made by the

user A. If user B performs pulling, he will get the changes made by user A and by other users working on the same project.

## 5. Architecture Models of Distributed Version Control Systems

The basic concept of DVCSs presented (see fig. 1) can be applied in different architectural models, depending on the size of the team, the type of the software applications, the way of organizing work.

The possible architecture solutions [2] are: the local one-person team architecture, the one-person team with off-box backup architecture, small team architecture, the open source architecture and other architectures.

**The local one-person architecture**

This is the simplest model that is suitable for use by an autonomous user. The user commits the changes to the local repository, which has the same content as the working copy. This model has many limitations. It is suitable for use by users who work alone.

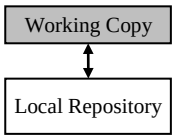



Fig. 4

**The one-person team with off-box backup architecture**

In this model, the changes that the user has made can be transferred to the local repository. Furthermore, a central repository is maintained, which is located on another computer. This ensures a secure back up of the data in case of failure of the local hard disk. The transfer of changes from the local repository to the central repository is done by pushing, and the moving of the changes in the opposite direction - by pulling.

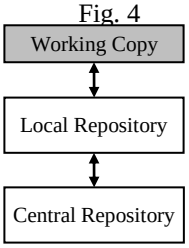



Fig. 5

**Small Team Architecture**

This model is suitable for use by a team of several users working remotely from one another. A central repository is maintained, through which users share the content of their local repositories. When, for instance, the user A is ready with the changes made, the user commits them to the local repository, which is updated from the working copy. With push the changes are transmitted to the central repository. In order to obtain the changes user B has to perform pull.

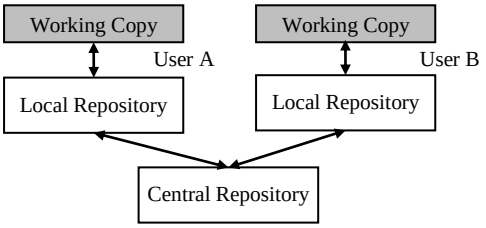



Fig. 6

**The open source architecture**

This model is suitable for use in teams developing open source project. Some of the users (a limited number) can have direct access to the central repository. The rest have access to a public clone of central repository. They commit their changes to the local repository, and then share them in the public clone. The content of the public clone is monitored by a manager who decides whether or not to withdraw the changes.

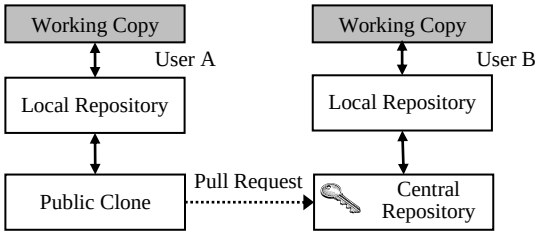



Fig. 7

**Other architectures**

For specific projects other architectural models can be built, using appropriate software tools. For example, within a very large project small teams of 5 to 10 participants can be divided which work in separate parts, and each team has its own repository, the content of which is pushed to the central repository of the project upon the completion of a particular stage of the development.

## 6. Conclusion

Version Control Systems (VCSs) provide simultaneous work for teams with one project code, but also with projects for Web sites and other types of documents.

Distributed version control systems (DVCSs) are the last stage in the development of VCSs, which provide teams, whose members are geographically distant from one another, with working opportunities. DVCSs have a local repository and can work without being connected to the central server.

DVCSs can be built in different architectural models, depending on the size of the team, the type of the software applications, and the way of organizing work. Possible architectural solutions are: the local one-person team architecture, the one-person team with off-box backup architecture, small team architecture, the open source architecture.

The latest trend in the development of VCSs is to provide them as services in the cloud, which has both advantages and disadvantages, typical for the cloud services.

## References

1. Ernst M. Version control concepts and best practices. https://homes.cs.washington.edu/~mernst/advice/version-control.html
2. Concepts of Distributed Version Control. RB Whitaker's Wiki. http://rbwhitaker.wikidot.com/mercurial-distributed-version-control
3. Knittl-Frank D. Analysis and Comparison of Distributed Version Control Systems. Hadenberg, Juni 2010. http://www.thehappy.de/~neo/dvcs.pdf
4. Yeates S. What is version control? Why is it important for due diligence? http://oss-watch.ac.uk/resources/versioncontrol
5. Otte S. Version Control Systems. https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2008-09_WS/S_19565_Proseminar_Technische_Informatik/otte09version.pdf?1346661971
6. Mukherjee P. A fully decentralized, peer-to-peer based version control systems. Darmstadt 2011. http://d-nb.info/1011040085/34
7. Santos P. Distributed VCS: Understanding the Paradigm Shift. http://www.drdobbs.com/architecture-and-design/distributed-vcs-understanding-the-paradi/240159530?pgno=1
8. Clatworthy I. Distributed Version Control System - Why and How. http://ianclatworthy.files.wordpress.com/2007/10/dvcs-why-and-how3.pdf

# Thematic Content Change in Logistics

**Julian Vasilev**[1]

[1] Varna University of Economics, 9002 Varna, Bulgaria
vasilev@ue-varna.bg

**Abstract.** The purpose of this paper is to identify the thematic content change in logistics. Logistics and supply chain management are fields of study with great interest. A grounded theory approach is used to analyse peer-review articles in online databases. First, the article describes the most commonly used sub-themes in logistics. Second, each sub-theme is checked in one online database during the last five years. Some trends are identified. The applied methodology has a wide practical application. The study is the first of its kind to analyse thematic content change in logistics.

**Keywords:** logistics, online library databases, research interests, ontology, SPSS

## 1.Introduction

Logistics is a field of study with interdisciplinary interest. The subcategories of logistics allow a great number of researchers to work in the field of logistics. A lot of articles, journals and conferences are dedicated to logistics. The change of thematic content means that some areas of research interest are not so important or researchers do not focus on them. Online libraries provide information about the latest publications in business, research papers, book reviews and other papers concerning logistics.

We suppose that the thematic content in the sphere of logistics changes. To accept or reject our initial hypothesis we have to analyse online library databases. Subcategories and key words may help us in building ontology in logistics. Ontology means the set of words used by people in a particular area.

To make the current study we start analysing the Directory of Open Access Journals (DOAJ). Future research may focus on: (1) Ebsco host, (2) Emerald Insight and (3) Inderscience.

## 2.Analysis of the thematic content in logistics in online databases

By searching the key word “logistics” 1989 articles are found. The interest in logistics increases.

**Table 1.** Number of article by keyword “logistics” by years in the Directory of Open Access Journals (DOAJ)

| Year | Number of articles |
|---|---|
| 2014 | 213 (until October) |
| 2013 | 427 |
| 2012 | 323 |
| 2011 | 245 |
| 2010 | 217 |
| 2009 | 146 |
| 2008 | 94 |

| | |
|---|---|
| 2007 | 75 |
| 2006 | 93 |
| 2005 | N/A |
| 2004 | 39 |

Without making a graph it is obvious that the number of articles in logistics increases throughout the years. The thematic content of the used key words is the following.

**Table 2.** Number of articles, grouped by key words (ordered alphabetically), concerning logistics

| Key words | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|
| Green logistics | 6 | 8 | 9 | 9 | 7 |
| Catering logistics | | | | 1 | 1 |
| City logistics | 16 | 16 | 13 | 19 | 8 |
| Continuous logistics models | 2 | 1 | 2 | 3 | 1 |
| Disaster relief logistics | | 2 | 2 | 1 | 1 |
| Discrete logistics models | | 1 | 1 | 1 | 2 |
| Door-to-door logistics service | 1 | 1 | | | 1 |
| Emergency logistics | 4 | 3 | 5 | 8 | 4 |
| Flexible multistage logistic network | | | 1 | | 1 |
| Fourth-party logistics | | | | 2 | 1 |
| Information logistics | 2 | 1 | 1 | 3 | 2 |
| Logistic functions | 1 | 3 | 2 | 3 | 3 |
| Logistical worlds | | | | | 1 |
| Logistics centre | 2 | | 4 | 8 | 2 |
| Logistics centre location selection | | | | | 1 |
| Logistics costs | 3 | 6 | 8 | 12 | 6 |
| Logistics distribution | 3 | 5 | 6 | 16 | 4 |
| Logistics distribution network | | 1 | | 1 | 1 |
| Logistics environment | | 1 | 2 | 2 | |
| Logistics industry | 5 | 8 | 9 | 11 | 7 |
| Logistics management | 6 | 8 | 16 | 16 | 10 |
| Logistics manager | | | | | 1 |
| Logistics mode selection | | | | | 1 |
| Logistics networks | 13 | 18 | 13 | 24 | 18 |
| Logistics outsourcing | 5 | 5 | 13 | 8 | 6 |
| Logistics performance index (LPI) | | | | 1 | 1 |
| Logistics process | | | 5 | 6 | 2 |
| Logistics reform | | | | | 1 |
| Logistics routing | | 1 | | 2 | 1 |
| Logistics service | 2 | 9 | 17 | 21 | 13 |
| Logistics services providers | | | | | 2 |
| Logistics support | 1 | 2 | 3 | 3 | 6 |
| Logistics support planning | | | | | 1 |
| Logistics system | 15 | 13 | 19 | 27 | 4 |
| Manufacturing logistics | | | 2 | 1 | |
| Maritime logistics | | | | | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Port logistics | | 3 | 2 | 8 | 1 |
| Port logistics industry | | | | | 1 |
| Procurement logistics | | 1 | 1 | 1 | 1 |
| Production logistics | 1 | 4 | 5 | 6 | 1 |
| Purchasing logistics | | 1 | | | 1 |
| Regional logistics network | | | | | 1 |
| Reverse logistics | 11 | 20 | 23 | 22 | 16 |
| Rural logistics | | | | | 1 |
| Third-party logistics | 1 | 5 | 6 | 9 | 3 |
| Urban logistics | 3 | 3 | 2 | 3 | |
| Warehouse logistics | | | 1 | 3 | 1 |

The most interesting spheres of logistics are: reverse logistics (92 articles during the last 5 years), logistics networks (86 articles), logistics system (78 articles), city logistics (72 articles), logistics service (62 articles), logistics management (56 articles), logistics industry (40 articles) and green logistics (39 articles).

Terms, connected with logistics, mentioned for the first time in year 2014 and not mentioned during the last 5 years are: logistics worlds, logistics centre location selection, logistics manager, logistics mode selection, logistics reform, logistics support planning, maritime logistics, port logistics industry, regional logistics networks and rural logistics.

Terms that are found only in year 2013 and year 2014 are: catering logistics, fourth-party logistics, logistics performance index and logistics routing. These terms are going to be met in new articles in the following years.

There are no terms, which are mentioned in 2010 and 2011 and are missing in 2012, 2013 and 2014. This fact shows that the thematic interest in logistics increases in most of the terms. If we compare 2010 and 2013, the number of articles in 2013 is greater than those in 2010 for each term in logistics. This fact shows an increasing interest in all spheres of logistics. Moreover, new terms are found in year 2014, which were not found during the last five years. We may conclude that logistics is developing and widening the research interest.

**Table 3.** Number of years each term is encountered

| Five years | Four years | Three years | Two years | One year |
|---|---|---|---|---|
| City logistics | Disaster relief logistics | Door-to-door logistics service | Catering logistics | Logistical worlds |
| Continuous logistics models | Discrete logistics models | Logistics distribution network | Flexible multistage logistic network | Logistics center location selection |
| Emergency logistics | Logistics centre | Logistics environment | Fourth-party logistics | Logistics manager |
| Green logistics | Port logistics | Logistics process | Logistics performance index (LPI) | Logistics mode selection |
| Information logistics | Procurement logistics | Logistics routing | Manufacturing logistics | Logistics reform |
| Logistic functions | Urban logistics | Warehouse logistics | Purchasing logistics | Logistics services providers |
| Logistics costs | | | | Logistics |

| | | | | |
|---|---|---|---|---|
| | | | | support planning |
| Logistics distribution | | | | Maritime logistics |
| Logistics industry | | | | Logistical worlds |
| Logistics management | | | | Port logistics industry |
| Logistics networks | | | | Regional logistics network |
| Logistics outsourcing | | | | Rural logistics |
| Logistics service | | | | |
| Logistics support | | | | |
| Logistics system | | | | |
| Production logistics | | | | |
| Reverse logistics | | | | |
| Third-party logistics | | | | |

Most of the terms (18 terms) in logistics exist during the last five years. This fact shows a constant interest in all subcategories of logistics. 11 terms are met just one year. It may be an episodic interest or new terms, that are discussed in logistics or a new term mentioned in 2014. Filtering the dataset in table 2 shows that these are new terms in logistics. Six terms are encountered during four of the last five years. Another group of six terms are encountered during three of the last five years. 13 terms did not exist in year 2010 but they appear in 2013.

**Table 4.** Terms, that did not exist in 2010 but they appear in 2013

| Term name | Number of articles during the last five years |
|---|---|
| Port logistics | 14 |
| Logistics process | 13 |
| Disaster relief logistics | 6 |
| Discrete logistics models | 5 |
| Logistics environment | 5 |
| Warehouse logistics | 5 |
| Logistics routing | 4 |
| Procurement logistics | 4 |
| Fourth-party logistics | 3 |
| Logistics distribution network | 3 |
| Manufacturing logistics | 3 |

| | |
|---|---|
| Catering logistics | 2 |
| Logistics performance index (LPI) | 2 |

The terms with the greatest interest in logistics are: port logistics, logistics process and disaster relief logistics. The newcomers are manufacturing logistics, catering logistics and logistics performance index.

We may make another check – which terms existed in 2010 but are not present in 2013. Just one term "Door-to-door logistics service".

A correlation matrix is made in SPSS to check the dependencies of the number articles throughout the years.

**Table 5.** Correlation matrix

| | | y2010 | y2011 | y2012 | y2013 | y2014 |
|---|---|---|---|---|---|---|
| y2010 | Pearson Correlation | 1 | ,870** | ,703** | ,776** | ,558* |
| | Sig. (2-tailed) | | ,000 | ,001 | ,000 | ,011 |
| | N | 21 | 20 | 20 | 20 | 20 |
| y2011 | Pearson Correlation | ,870** | 1 | ,867** | ,896** | ,854** |
| | Sig. (2-tailed) | ,000 | | ,000 | ,000 | ,000 |
| | N | 20 | 28 | 24 | 26 | 26 |
| y2012 | Pearson Correlation | ,703** | ,867** | 1 | ,896** | ,800** |
| | Sig. (2-tailed) | ,001 | ,000 | | ,000 | ,000 |
| | N | 20 | 24 | 29 | 28 | 26 |
| y2013 | Pearson Correlation | ,776** | ,896** | ,896** | 1 | ,785** |
| | Sig. (2-tailed) | ,000 | ,000 | ,000 | | ,000 |
| | N | 20 | 26 | 28 | 33 | 30 |
| y2014 | Pearson Correlation | ,558* | ,854** | ,800** | ,785** | 1 |
| | Sig. (2-tailed) | ,011 | ,000 | ,000 | ,000 | |
| | N | 20 | 26 | 26 | 30 | 44 |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The correlation between the articles between year 2010 and 2011 is 0.87. It shows that 87% of the terms in logistics used in 2010 are also met in the next year. 13% of the terms are new. The correlation between the articles between year 2011 and 2012 is 0.867. It shows that 87% of the terms in logistics used in 2011 are also met in the next year. 13% of the terms are new. The correlation between the articles between year 2012 and 2013 is 0.896. It shows that 90% of the terms in logistics used in 2012 are also met in the next year. 10% of the terms are new. The correlation between the articles between year 2013 and 2014 is 0.558. It shows that 56% of the terms in logistics used in 2013 are also met in the next year. 44% of the terms in logistics are new in year 2014.

Using the k-means cluster method a classification of terms in logistics is done. The classification is done in SPSS (Analyze/Classify/K-means Cluster). Three clusters are formed. Cluster 1 contains 31 terms, cluster 2 – 5 words and cluster 3 – 8 words. The clusters

have been chosen to maximize the differences among cases in different clusters. Since cluster one contains most of the words we will give the terms in the second and third cluster.

**Table 6.** Terms in cluster 2 and cluster 3

| Cluster 2 | Cluster 3 |
|---|---|
| City logistics (72) | Emergency logistics (24) |
| Logistics networks (86) | Green logistics (39) |
| Logistics service (62) | Logistics costs (35) |
| Reverse logistics (92) | Logistics distribution (34) |
| Logistics system (78) | Logistics industry (40) |
| | Logistics management (56) |
| | Logistics outsourcing (37) |
| | Third-party logistics (24) |

* The number in the brackets shows the number of articles the term is met during the last five years, according to the information in DOAJ.

The k-means classification shows that cluster 2 contains the mostly used terms in logistics. Cluster 3 contains the second group of terms that are mainly used in logistics. Cluster 1 contains all other terms that are used not so often as the terms in the second and third cluster.

The variable total number of publications is checked for normal distribution. The one-sample Kolmogorov-Smirnov test is executed. The distribution of the tested variable is normal (Asymp. Sig. 2-tailed is 0.013). Thus nonparametric tests may be done.

To check if there is a significant difference between the count of articles in each term the Friedman test is executed (Analyze/Nonparametric tests/K-related samples). The five variables (y2010, y2011, y2012, y2013, y2014) are added. The asymp. syg. value is 0.000. It is less than 0.05, which means that there is a significant difference between the groups. It means that authors are free to write articles containing any of the keywords in one of the studied year.

The parametric repeated measures ANOVA may be applied (Analyze/General Linear Model/Repeated Measures). One factor with five levels is defined. The Mauchly's test of sphericity is 0.000. It is less than 0.05. We read the p-value from the second row of the table "Tests of within-subjects effects". The p-value of Greenhouse-Geisser is 0.000. It means that we can assume sphericity. Sphericity means that there is a significant difference between the five groups.

To analyse the differences a post hoc test may be applied (Analyze/Compare Means/Paired-Samples T Test). These tests are applied when significant differences are found.

**Table 7.** Paired samples T tests

| Pairs | Sig. (2-tailed) | | Pairs | Sig. (2-tailed) |
|---|---|---|---|---|
| y2010-y2011 | 0.002 | | y2011-y2012 | 0.031 |
| y2010-y2012 | 0.000 | | y2011-y2013 | 0.000 |
| y2010-y2013 | 0.000 | | y2011-y2014 | 0.805 |
| y2010-y2014 | 0.035 | | | |
| | | | | |
| y2012-y2013 | 0.001 | | y2013-y2014 | 0.001 |
| y2012-y2014 | 0.055 | | | |

The t-test shows that y2012 and y2014 don't seem to be significantly different from each other (Sig. 0.055). The t-test shows that y2011 and y2014 don't seem to be significantly different from each other (Sig. 0.805). All other pairs seem to be significantly different (P-values are below 0.05).

The interclass correlation coefficient (ICC) may be used to estimate the reliability of the measurement method (Analyze/Scale/Reliability). The highest correlation is between y2011 and y2013 (0.896). The assumption we make is that we may try to re-measure the number of articles by terms reliably – for instance in different online library databases. When we re-measure, we will find differences. Some of them may be due to the searching in another database, not just variation in the measurements. The ICC value is 0.846. The f-test shows p-value of 0.000. It is below 0.05. The ICC value is very highly statistically significant. It is quite near the calculated correlation coefficient. The high value of ICC shows a great amount of inter-rate agreement. The Cronbach's Alpha is another measure of reliability. Its value is 0.916. A reliability coefficient of .70 or higher is considered "acceptable" in most social science research situations using Cronbach's Alpha.

## 3.Conclusion

The thematic content in logistics contains a number of terms that are used in journals. After finishing our research we may make several conclusions.

The number of articles about logistics increases during the last five years (2009-2014).

The most interesting spheres of logistics are: reverse logistics (92 articles during the last 5 years), logistics networks (86 articles), logistics system (78 articles), city logistics (72 articles), logistics service (62 articles), logistics management (56 articles), logistics industry (40 articles) and green logistics.

Terms, connected with logistics, mentioned for the first time in year 2014 and not mentioned during the last 5 years are: logistics worlds, logistics centre location selection, logistics manager, logistics mode selection, logistics reform, logistics support planning, maritime logistics, port logistics industry, regional logistics networks and rural logistics. These terms are going to be met in new articles in the following years.

There is an increasing interest in all spheres of logistics. We may conclude that logistics is developing and widening the research interest.

Future research may focus on analysing other online digital libraries, making ontologies and applying the methodology, described in this article. Future research may be focused in thematic content in other spheres of human life.

## References

1. Andrew Garth. Analysing data using SPSS. Sheffield Hallam University, 2008, p. 74-84. Available: https://students.shu.ac.uk/lits/it/documents/pdf/analysing_data_using_spss.pdf <29.10.2014>.
2. Directory of Open Access Journals (DOAJ). Available: www.doaj.org <29.10.2014>.
3. Gujarati, D. Basic econometrics. 4th edition. http://egei.vse.cz/english/wp-content/uploads/2012/08/Basic-Econometrics.pdf <29.10.2014>.
4. Hadzhiev, V. and others. Statistical and econometric software, Varna: Nauka i ikonomika, 2009 (in Bulgarian).
5. Vasilev, J. Time series analysis in loan management information systems. Theoretical and Applied Economics, Volume XXI (2014), No. 3(592), pp. 57-66. Available: http://store.ectap.ro/articole/962.pdf <29.10.2014>
6. Vasilev, J. Creating a Customer Profile in a Credit Institution. International Journal of Advanced Research in Computer Science and Software Engineering. 4(1), January - 2014, pp. 1108-1112. Available: http://www.ijarcsse.com/docs/papers/ Volume_4/1_January2014/ V4I1-0564.pdf <29.10.2014>.

# Some Opportunities of CMS OpenCart to Realization of e-Shop

**Dimiter Dimitrov[1], Svetlana Vasileva[2]**

[1] SHU, 115, Universitetska St, 9700 Shumen, Bulgaria
demen7ed.boy@gmail.com

[2] SHU, College – Dobrich, ZK Dobrotica 12, 9302 Dobrich, Bulgaria
svetlanaeli@abv.bg

**Abstract.** The Content management systems (CMS) automate and facilitate the process of adding and modifying the contents of the Web sites, organization, control and publication of a large number of documents and other content, such as images and multimedia resources. This makes the Content Management Systems attractive for specialists in various fields of human activity who want to publish on the Internet, but have little knowledge in computer programming and in particular web-programming. Therefore CMS *OpenCart* recently is preferred by companies that extend their business with the development of their e-shops. The report considers some of the capabilities of CMS *OpenCart* design and creation of e-shops. We have considered an example of e-shop for computer techniques and accessories.

**Keywords.** Content management system, web-sites, e-shop,.

## 1.Introduction

In the modern world the opportunities for electronic exchange and communication, as well as the expansion of electronic applications and services in all sectors of public life laid the foundations of the transition from an industrial to a global digital economy. The convergence of telephone, television and computers changed the traditional markets and society. [1] New business configurations and relationships are emerging. Companies create global networks linking research, production, distribution and trade. [1] The global network offers a variety of services - information, banking, transport, trade, etc.. More and more companies export trade activities and the Internet. New trends in trade are - exports and on the Internet in order to better serve the customers and greater competitiveness of traders and topicality of the market are subject to development.

The paper presents developed using the Content Management System (CMS), an OpenCart e-shop for computer techniques and accessories. Choosing the CMS OpenCart is dictated by the fact that this CMS is specifically designed for creating e-shops [2], [3], [4], [5] and others).

## 2.Features and opportunities of OpenCart

OpenCart is one of the best systems for the realization of e-shop and is completely free. With this system you can create as small e-shop, and professional online store. [2] The installation and management of the shop is very easy, there is a whole package Bulgarian version and support. Can create a myriad of categories and products in admin panel has a filter through which easily can be made available when an item in your store retailer has several thousands. In Bulgaria both for osCommerce, and for OpenCart there is a modul for

supply through the supply company Econt Express. Payment Modules are many - some of them are: Bank transfer, payment upon receipt, AlertPay, PayPal, ePay. CMS OpenCart allows the creation of vouchers or promo codes with constant and / or for a certain period of time reductions trader can cite in promotional materials for the goods / services. In the admin module there is a section Reports showing the statistics for viewed, ordered and sold products. Possibility of several currencies and languages of the e-shop, as well as additional pages in which the developer to introduce free text - information about the company, supplies, suppliers, payments, contacts, time and others. [6] OpenCart also offers a wide range of over 9000 ready-to-use free and paid applications. OpenCart features a diverse selection of functions that include, besides unlimited categories and products, unlimited number of producers, writers, all currencies, languages, product ratings multistor, product reviews, ratings, numerous integrations with popular banking services, shipping companies and payment methods. [7]

OpenCart is a system for creating an electronic shop, created by Daniel Kerr. OpenCart is based on PHP and uses MySQL database. By default, the system runs on Apache, but it is also possible to operate under nginx. It is licensed under the GNU GPL license.

**- Management of the e-shop:** Unlimited number of e-stores in one platform - OpenCart allows you to create and manage unlimited number of e-shops from an administrative panel extremely easy. Moreover, the designer has complete control over: The design of each store, so that it is in harmony with the offered product range, if needed; Which products in which store to be displayed with what prices, etc.; Supported languages, currencies and tax classes set for each specific e-shop separately; Customer groups - could easily be a separate shop with wholesale prices, for example.

- **Unlimited number of standard and electronic products:** product catalog developed with OpenCart e-shop has everything you need to allow the trader to sell both physical and electronic products without restriction on the number and types. Available to the developer are numerous additions, enabling him to describe and present the best possible products and for visitors to shop and search engines Google, Yahoo, etc.

**- Multiple currencies:** The ability of the customer eshop to view and pay the desired product in convenient for them currency advantage and Opencart trader can add and adjust exchange extremely easy. The system automatically takes care of updating the exchange rates and recalculate the prices of the products to the shop on the main currency unit, so they are always up to date without needing the intervention of the trader. Furthermore, CMS OpenCart in its admin panel supports: Multiple tax rates; Multiple delivery zones; Calculator cost of transport; 20+ international and Bulgarian payment methods; 8+ Bulgarian and international delivery methods; Electronic invoices; Edit orders - a useful feature to have can make adjustments in an order; Monitoring of cash; Reports and statistics; SSL security [8].

- **Marketing tools:** product evaluation; Rating of the products; Related products; Best-selling products; New products; Recently Viewed products; Urgent orders; Flexible system of discounts; Coupons for discounts; Reward points for customers; Product Feed; Manage own affiliate program; Built-in e-mail system; 100% structural search engine optimization (SEO).

**- Maintenance of the e-shop by CMS OpenCart:** Tools for preservation and restoration of data; Open for editing code; Rich and accessible documentation; Forum users.

## 3. Advantages of CMS OpenCart

Internet store built with OpenCart features high speed of loading pages, even with a large catalog of bulky graphical components.

With a very good for SEO optimization functionality of the online store in search engines, such as the ability to determine the url of the page manually. For each product you can set categories, keywords, meta tags and titles.

OpenCart has its own system of reports. You can configure detailed reports made orders, returns and most popular products.

Includes settings for the effective performance of products, options for discounts and promotions tied to specific products, as well as function introduction of different options for delivery.

Integrated several payment systems, including WebMoney.

OpenCart is an excellent system for small and medium business that does not require significant server resources.

## 4. Installation of CMS OpenCart

To create a web site by means of CMS OpenCart, it is necessary first of all to install the free software package XAMPP. After the installation of XAMPP we run the moduls Apache and MySQL

In our case, the administration is going through *Apache Web* browser to load the results of the index page, and for the administration of MySQL is needed again in the browser to load connector database PhpMyAdmin, which is also built into the XAMPP Control Panel and charging by Admin button to MySQL.

For a start we can launch button "Admin" for MySQL in XAMPP control panel or manually to raise PhpMyAdmin address in the address bar of your browser. Then enter a username and password to create a database.

Figuring any name with which we will install CMS and collation (encoding) of the database. In our case we have chosen the name pcshop with collation / character encoding DB utf8_unicode_ci required to be entered and displayed correctly texts that certainly dominate the characters from ASCII table data of our future web-shop for computer equipment and accessories.

The next step is to open the website of CMS OpenCart (www.opencart.com) and download the latest updated version of the product. In our case it was version 1.5.6.4. Download the archive with the system from the official website of the product (opencart-1.5.6.4.zip), then unzip it, so files upload/be extracted in the root directory *htdocs* or *html_public* (www) as hosting our event is htdocs in the local hosing of XAMPP - there will be indexed our web page in the browser from which you will configure the entire system.

Fig. 1 shows a window of the installation of CMS OpenCart. Both the status of the configuration files of the site and control panel site in checking come out with red lettering Missing (Missing), which means that no permissions assigned to write on these two important PHP-file or missing, as in the missing can be say that with another name, as in our case and should only be renamed from the index directory *htdocs*/*config-dist.php* and *htdocs/ admin/config-dist.php* without suffix -dist, so it looks config. php in both cases. Then we are ready for the actual installation.

The last step you have to do is to delete from the index directory *install* directory to have no surprises from hacker breaches site from beginner ones. After the installation, the site seems nothing has changed - Fig. 2.

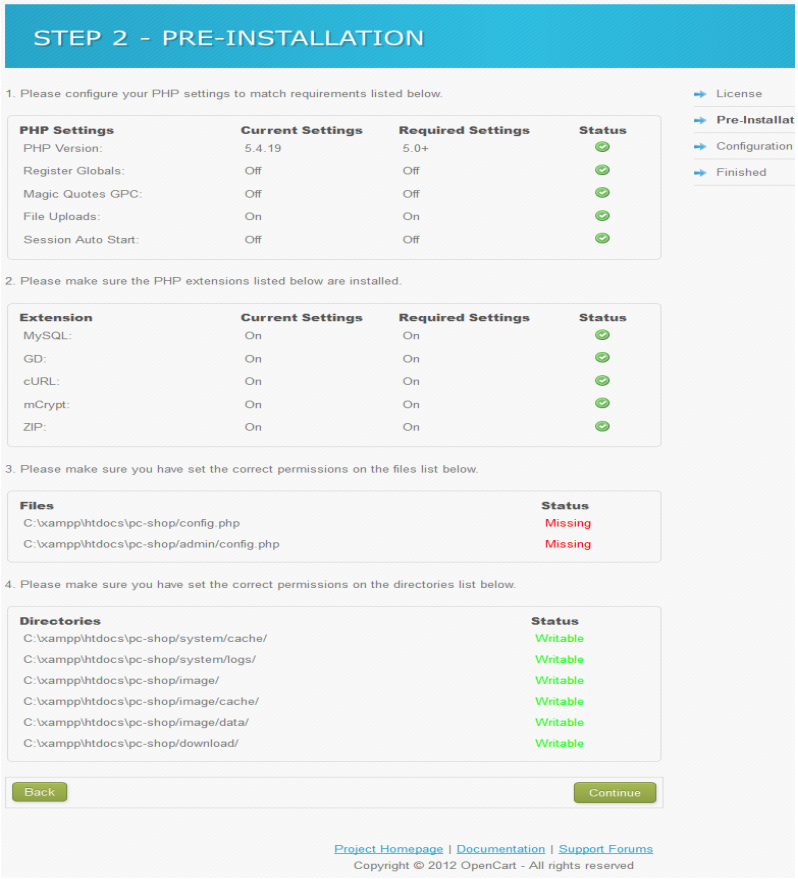



Fig. 1 Step in the installation of OpenCart. Files for the site configuration

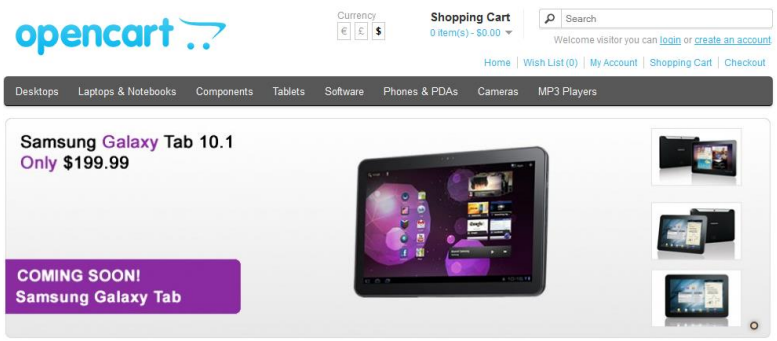



Fig. 2 “Empty” OpenCart e-shop

## 5.Creation of e-shop by means of CMS OpenCart

Fig.3 shows the admin panel of the created e-shop of the dealer. From this window is set to have or not a backup of the database.

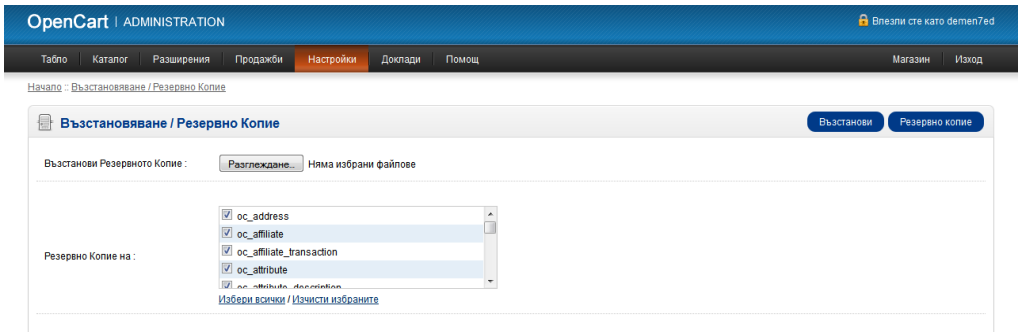



Fig. 3 AdminPanel - set to have or not a backup of the database

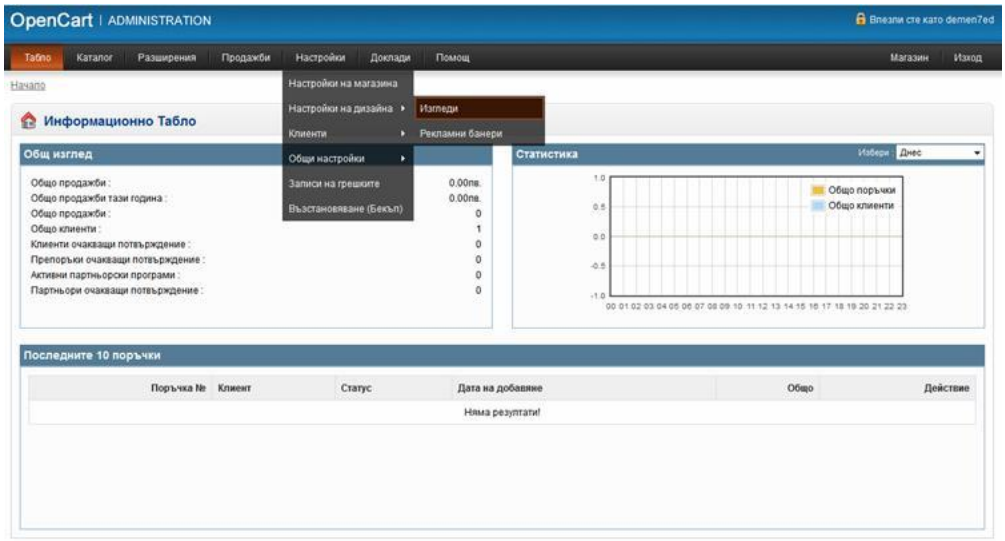



Fig. 4 Overview of the display of the e-shop

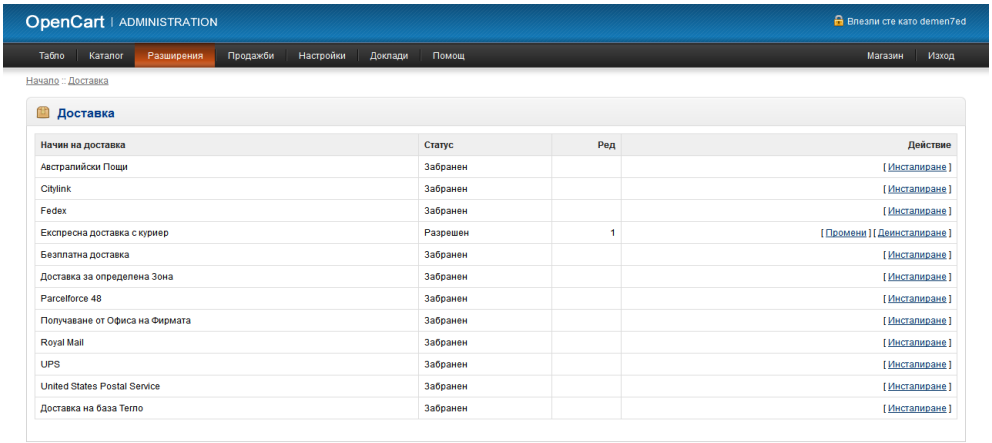



Fig. 5 Admin panel for additional delivery options

Fig. 4 shows an overview of the informational desk of the store, which the programmer is set up so that the drop-down menus can set up this module for Bulgarian language and are pre-installed. Furthermore, the trader can monitor all sales for the month on the day information about customers and orders. Can be set and awaiting orders - The exchange dealer called to confirm that the client that really is ordered, not just marked.

Fig. 5 shows the admin panel on/off additional delivery options - the trader can include / exclude companies supply which has contracts. Can be adjusted and the cost of delivery. (set the basic amount - more than 100 BGN delivery at the expense of the operator, under 100 lev is paid by the buyer.

Fig. 6 shows the Admin Panel for setting the categories of goods offered for sale online. Where there are integers greater than 0 (fig. 6), it means that the corresponding category has subcategories.

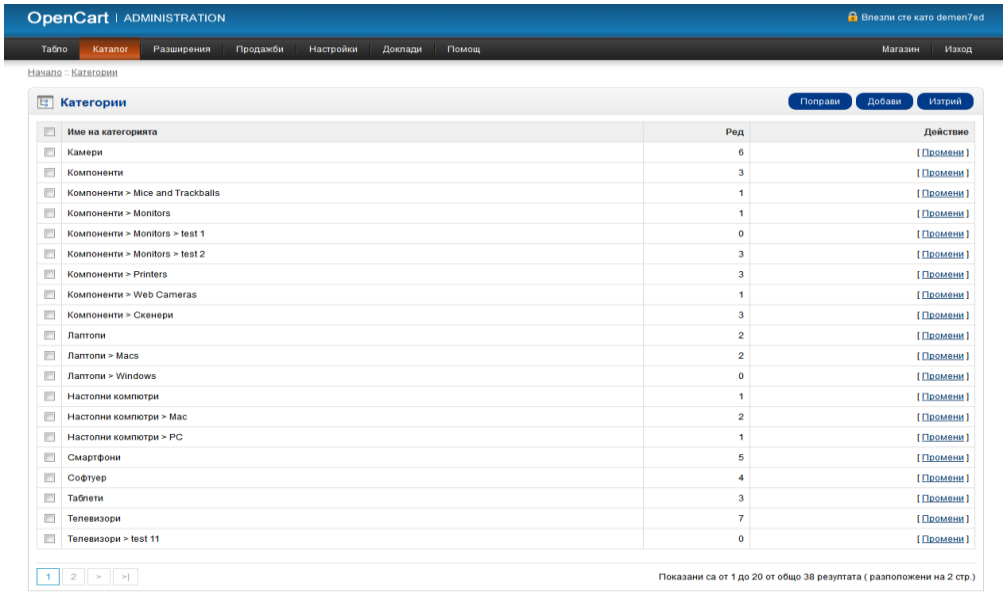



Fig. 6 Admin Panel for setting the categories

Fig. 7 shows the Admin Panel, which are derived from various reports – eg the trader to see the company's profits for the period.

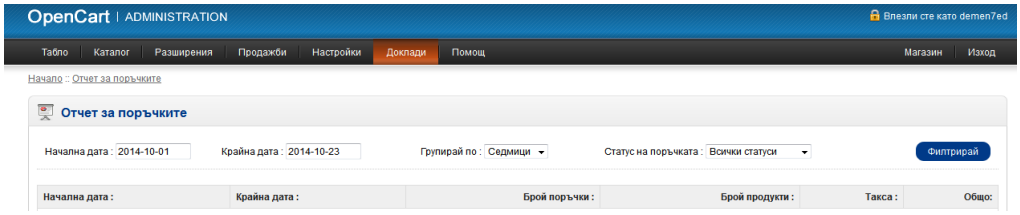



Fig. 7 Admin Panel for Reports

Fig. 8 shows the Admin Panel of the e-shop for computer techniques and accessories. From this panel trader as new product introduces a product name with description - photo, characteristics and quantity of stock and price. (If the status is prohibited - so it is no longer on sale).

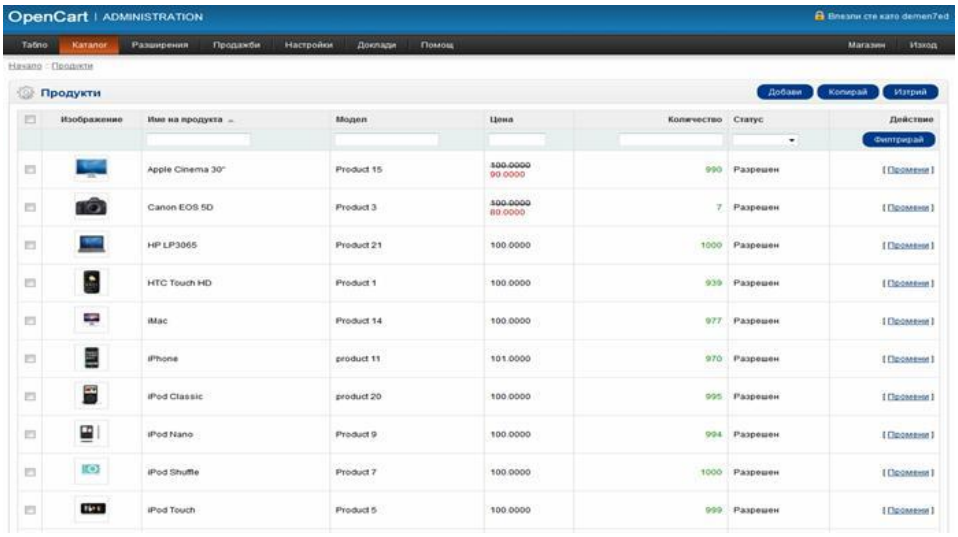



Fig. 8 AdminPanel to introduce new products

Fig. 9 shows the window of the client portion of the web site of the created e-shop. Customers can indicate products as the most liked (not even buy them) to be able to make a comparison. The client can share a product (to advertise the product that is on the screen). The customer can leave a comment about a product. The web site has a protection against bots (which spam network, advertise to another store or written obscenities) with CAPTCHA code. [8]

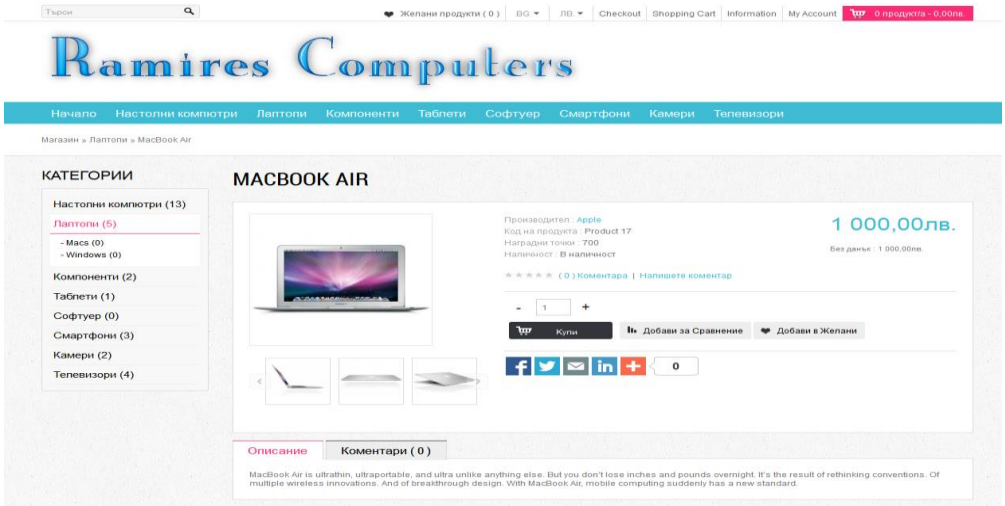



Fig. 9 Customer section of the website

Fig. 10 shows a window with the English version of the created e-shop. In this window, the client can change the currency (if the customer changes the currency shall be recalculated on the entered price formula set by the customer currency).

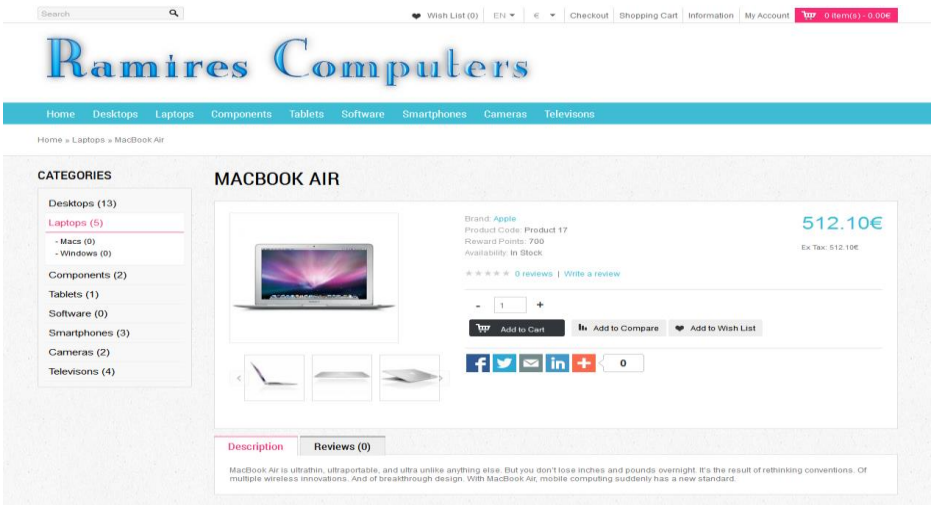



Fig. 10 English version of the created e-shop

## 6.Conclusion

CMS OpenCart is a very suitable tool to create fast dynamic web sites with a wide range of facilities and services. With the funds of OpenCart is created e-shop for computer techniques and accessories.

With the developed e-shop the Ramires company: export business on the Internet; becomes more accessible to consumers; is competitive.

The work on the project could serve as a basis to create friendly templates for e-shops for the nonprogrammers traders..

## References

1. Nelzina, O. G. *Information systems for e-commerce*, Fenix, Rostov n/D., 2008. (Нельзина, О. Г. *Информационные системы для электронной коммерции*, ООО „Феникс”, Ростов на/Д., 2008).
2. Chantharoth R., *Create e-commerce website Opencart*, (http://angkor.barcampcambodia.org/wp-content/uploads/2013/02/Create-e-commerce-website-Opencart.pdf)
3. Hasan T., *OpenCart 1.4 Template Design Cookbook*, Pack Publishing, 2011, (https://www.packtpub.com/web-development/opencart-14-template-design-cookbook)
4. Nepali R., *Getting Started with OpenCart Module Development*, Pack Publishing, 2013, (https://www.packtpub.com/sites/default/files/9781783280377_Chapter-01.pdf)
5. OpenCart 1.5 Home, HostJars, 2012, (http://opencart.hostjars.com/documentation)
6. Watson K., *OpenCart 1.5 User Manual*, OSCManuals.com, 2012.
7. Yilmaz M., *OpenCart 1.4 Beginner's Guide*, Pack Publishing, 2010, (https://www.packtpub.com/sites/default/files/3029-chapter-5-setting-payment-models.pdf)
8. Security tips for opencart, (http://angkor.barcampcambodia.org/wp-content/uploads/2013/02/Security-tips-for-opencart.pdf)

# Application of Modern Information Technologies in the Practice of Business Analysis

**Rositsa Ivanova**

UNWE, 1700 Sofia, Bulgaria
Rosi_Ivanova@abv.bg

**Abstract.** Business analysis is a sub-system of the organisation's financial management system. It should ensure information not for previous events, but rather for making forecasts for the organisation's future and its financial and market stability and sustainable development. To this end, business analysis should be performed in real time. This gives rise to the topicality of the issue for the application of modern information technologies in the practice of business analysis. The report highlights different issues related to the use of information technologies in the practice of the organisation's overall activity analysis from the point of view of system, complex and process approaches. The aim is to outline the positive aspects of the use of modern information technologies in the practice of the business analysis for making forecasts for the organisation's future development.

**Keywords:** analysis, real time, information technologies, sustainable, development

## 1.Introduction

The financial-economic analysis (business analysis) has triune expression as an independent area in the scientific space, specialised management function within the organisation's management function and practically applied work. Every company's management needs information about its financial-economic position, financial and market stability and sustainable development. Under existing conditions of globalisation, market competition, market dynamics and entropy, the analysis performed only with data from the organisation's financial statements is insufficient. The effective information about events that have already happened does not bring the usability required for the business. Management does not need information of what has already happened, but for what will happen in operating and strategic aspect in future. In practice, the role of forecast analysis of organisation's financial stability and insolvency and bankruptcy risk within the integrity model for analysis of the organisation's overall activity increases. Business analysis may provide the information required by the management if made in real time only. Thus, we should use the capacities of modern information technologies for preliminary, operating, subsequent and forecast analysis of the marketing, innovation, investment, principal, commercial and financial activity, viewed within their interrelations and dependencies, as well as the organisation's financial position and overall business efficiency. The application of information technologies in the practice of business analysis has a definite benefit, as the real time analysis gives the organisation competitive advantage.

Modern information and communication technologies are practically used in all fields of economic practice. Since their birth, until nowadays, information technologies are considered specific matter intended for a close circle of specialists. The aim of this study is to break this view down by highlighting the positive aspects of the information technologies

use, the possibilities for business analysis they ensure, and the problems that should be solved.

The platform for entirely managed cloud is based on four pillars: work with large data amounts; cloud technologies; mobile technologies and social technologies.

Each organisation may use the capacities of this platform in two ways. The first one is the direct use of the platform, and the second one – through a strategic partner.

The importance of the issue for information technologies use, including cloud services for business analysis, is determined by the following objective circumstances:

1) The amount of information used for the purposes of the business analysis is significantly increasing.

2) Business analysis use varied sources of information, i.e. there is a process of information source diversification, required for the complex analysis of the organisation's overall activity.

3) Business analysis' methodology and technique are further complicated by developing and use of complex models that describe different relations and dependencies between the different components of the organisation's activity.

4) Associates' mobility increases and thus the management faces the objective need to process information in real time for the purposes of the organisation's operating management.

Integrated business analysis application can be developed and used within the platform for cloud services. It helps for the development and analysis of huge amount of operating data in real time. On this basis, it is possible to study and critically analyze the parameters of different scenarios for the behaviour of a number of indicators that are important for the business, defining income, expenses, financial result, financial balance, financial position and efficiency both from the use of resources and from the overall business activity. Thus, it is objectively capable to manage key business indicators in operation aspect (see Figure 1).

In parallel, the organisation's management may manage the key internal business process in real time (see Figure 2).

The use of integrated application for business analysis within the cloud services platform should be reviewed in its complexity from three aspects: benefits; risks and expenses.

In our view, the benefits from using the capacities of the platform for business analysis in a cloud environment could be classified as follows:

1) Economic benefits, the more important of them being as follows:

- the platform offers a complex approach along the chain: software licenses, engineering; maintenance. It uses pre-set and feasible business analysis platforms, which are certified;

- the input data handling efficiency, increases thus providing effective analytical information required for the organisation's operating management;

- the organisation gets competitive advantage, as it achieves the best presentation to the outside world at the best price and by means of easily scaleable architecture;

- objective possibilities for long-term business and future growth of the company are ensured;

- the organisation's management is allowed to make operating management decisions;

- the return of investments in cloud technologies increases due to the improvement of processes that maintain, analyse and manage within cloud environment.

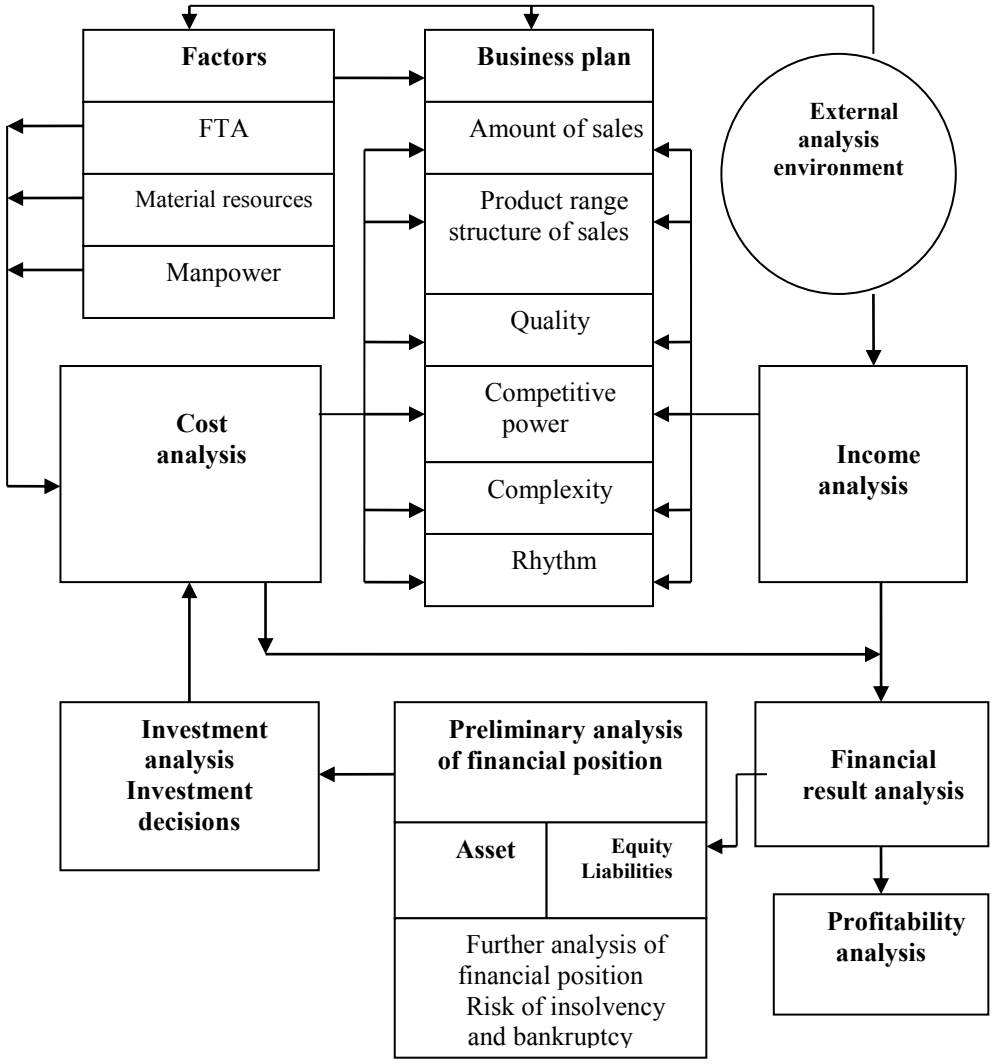



Figure 1. Complex business analysis by objects

2) Technological benefits, such as:

- information is automatically collected from different sources. This means that the business analysis' information basis may be automated to a great extent;
- quick access to information is ensured;
- information processing rate is improved;
- the integral business analysis application is an open system. Some objective possibilities for future upgrade of the application exist on the basis of the expansion of the business analysis' subject matter and the inclusion of new objects being subject to analysis and assessment. For example, organisation's sustainable development and the methodology for its analysis;
- flexibility of the business analysis cloud application is achieved, i.e. it may be easily adapted to the needs of the management for effective analytical information;

- a package of different business analysis applications is integrated;
- transferability and reliability of information processing are improved;
- high information processing productivity is ensured;
- the business analysis application is easily restored after the occurrence of unforeseen events.

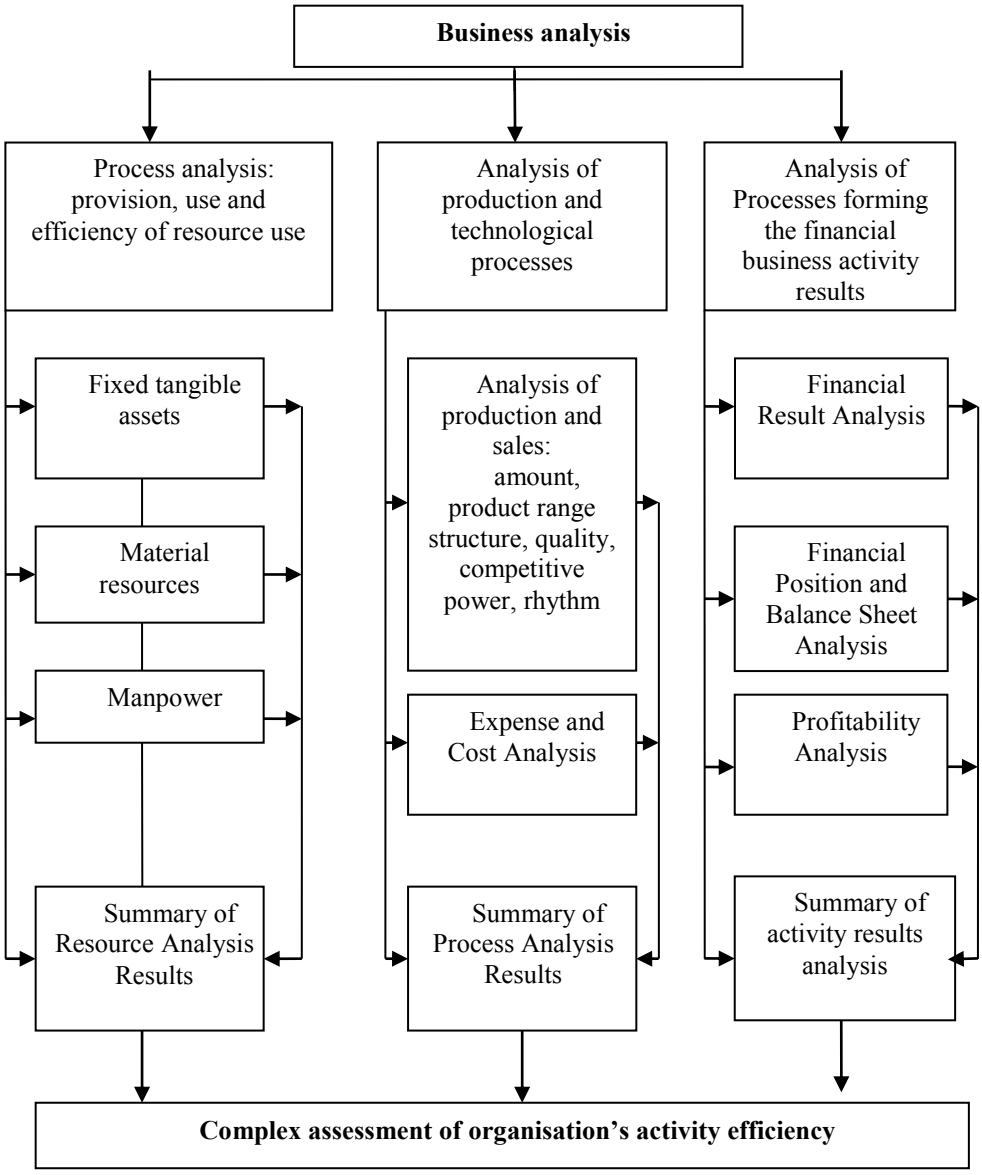



Figure 2. Complex business analysis by processes

3) Methodology and technique benefits, the more important of them being as:

- cloud technologies ensure better choice of IT solutions for business analysis;
- they offer possibilities for improvement of the business analysis method and increasing the use of mathematical methods for formation and behaviour of key business indicators;
- they offer possibilities for using improved or high quality new methodologies and models for business analysis that describe important interrelations and dependencies between different aspects of organisation's activity, including its sustainable development.

To our view, the risks of using the capacities of business analysis cloud platform may be classified according to the following aspects:

- Firstly, there is a risk of occurrence of system risk for the information security and protection of secret of data relevant to the business of an organisation. Here, with some justification, we can consider the issue of having technological capacity needed by the organisation for the purposes of protecting the information and its protection in cloud environment;
- Secondly, there is a risk relevant to the efficient use of business analysis cloud technologies. The organisation should ensure possibilities to use the business analysis cloud platform by a greater number of internal users (associates), without the need to provide special training. Provided business analysis cloud services are used by a limited number of associates, the efficiency of their use will be lower;
- Thirdly, there is a risk related to the return of investments for business analysis cloud technologies. The major problem here relates to the measurability of benefits from the use of business analysis cloud technologies in value expression. Thus, we should define a key indicator measuring the return of capital invested for integrated business analysis application in cloud environment. For the purposes of solving this issue, the organisation's management should correctly consider and answer the following questions: how will the business analysis application improve clients' satisfaction and will it achieve better customer relationship management; how will the organisation's current expenses change; will the supplier relationship management improve and will the time from order until the time of delivery decrease; will business efficiency improve; how and with what kind of indicators can information processing productivity be measured; how and with what kind of indicators can competitiveness of offered products and services, and of the organisation in general, be measured?

## 2.Conclusion

Perspectives: It is foreseen to develop a common European cloud (CoCo Cloud - Confidential and compliant) with a common, reliable, secure and Europe-wide cloud environment. Companies like HP, SAP, Atos, and other, as well as academic partners: such as the Italian National Research Council, University of Oslo, Norway, London Royal College, etc. take part in the working party. The project is funded under the European Union's Seventh Framework Program for Information and Communication Technologies and its duration is three years, from 2013 to 2016. Specific aims comprise achievement of confidentiality of information for each client of the cloud, as well as data compatibility. Thus, users will be able to work with their data in a secure information environment. The successful completion of the project is expected to overcome both the lack of synchronisation in the regulatory cloud service environment within the European Union, and the legal discrepancies between the EU Member States.

# Analysis of Migration - Interaction of Human Populations

**Elena V. Nikolova[1], Ivan P. Jordanov[1,2], Nikolay K. Vitanov[1]**

[1] Institute of Mechanics - BAS, Acad. G. Bonchev Str., Bl. 4, 1113 Sofia, Bulgaria
[2] UNWE, 1700 Sofia, Bulgaria
i_jordanov@email.bg

**Abstract.** Dynamics of many social systems is nonlinear. In this aspect we model the spatial and temporal dynamics of human populations by a system of nonlinear partial differential equations. We assume that the migration is a process influenced by the changing values of the growth rates and coefficients of interaction among the migration groups. We reduce the general model to analytically tractable partial differential equations (PDE) with polynomial nonlinearity up to second order for the particular case one population and one spatial dimension. We obtain an analytical solution which explains nonlinear kinks and solitary waves in the human migration dynamics.

**Keywords.** Nonlinear dynamics, population dynamics, migration of human populations, PDEs, balance equation, immigration waves.

## 1. Introduction

Migration of people from one region to another (within the country or abroad) is a global problem. It generally occurs as a result of any economic, political, ideological and social constraints in the target country. For each country, demographic statistics are available for this migration, but it is not enough to provide for the additional spatial and temporal characteristics of the dynamics of the human population [1,2]. Mathematical modeling provides this opportunity and helps to generate testable predictions [2,3,19]. It also ensures to design and evaluate protocols and thus to manage and control the dynamics of immigration. The systems studied the dynamics of the population are some of the best examples of nonlinear systems in quantitative social sciences [2,15,18]. Nonlinearities occur in the model equations of such systems is due to the interaction between individuals or populations, and limitations in the environment [8,9]. Such nonlinear model systems require application of the methods of nonlinear dynamics [6,11], chaos theory [7], and the theory of stochastic processes.

In recent years there has been rapid development of models of migration, because with their help we can adequately describe the processes in a number of social systems, such as the migration of different individual groups of the human population [1,2,24,25].

If the number of different groups and chaos in their behavior is small, it is possible to describe the system for interacting with a more or less - less complex systems of ordinary or partial differential equations. Such systems should describe the characteristics of the collective behavior of individual groups (their movement in space). We have just such modeling is necessary to describe the interaction between different social groups. Therefore, the deterministic model using the different types of systems, differential equations is further widespread. For the case of interest and spatial detection should be available to individuals in the system, which should be used more partial differential equations and systems.

There are various methods for obtaining a solution to these systems of equations. Some of them are derived from the values corresponding to certain temporal and spatial

coordinates. In some cases, however, a good idea will give us the decision itself, although collected under appropriate assumptions for its type. Recently especially interesting from a practical point of view, are the solutions obtained in the form of waves that describe the immigration of persons in space, ie so-called population waves [10,11,24,25].

Of course the decisions of linear equation we get a linear wave. In this case, a dispersion phenomenon can be observed, which leads to a reduction of the wave. However, when the equations are nonlinear, we can get waves in which the non-linearity compensate dispersion. These are so-called solitary waves, a special case of which are solitons [13,14]. We will examine the nonlinear waves, which are received by a parabolic partial differential equation with polynomial nonlinearity of second order.

## 2. Mathematical Model

Differential form of the equation describing the migration is [16]:

$$\frac{\partial \rho_i}{\partial t} + \nabla . \vec{I}_i = R_i, \quad i = 1,2,\ldots,n \tag{1}$$

The first term in (1) describes the net rate of increase of the density of *i*-th group, the second term describes the net rate of immigration into the area and the third term describes the net rate of increase without migration. Similar to the generalized Lotka–Volterra equations we assume that the changes of the density because of reproduction and interaction between the populations. Then the function $R_i$ can be written in the form [4,17]:

$$R_i = r_i^0 \rho_i (1 - \sum_{j=1}^{n} \alpha_{ij}^0 \rho_i + \sum_{j=1}^{n} r_{ij} \rho_j) \tag{2}$$

where $r_i$ is the growth ratio of the *i*-th population and $\alpha_{ij}$ is the interaction coefficient measuring to what extent the growth of the *i*-th population is influenced by the *j*-th population. We assume also that both coefficients depend on the density of the individuals in the population in the following manner [4,5]:

$$r_i = r_i^0 (1 + \sum_{k=1}^{n} r_{ik} \rho_k) \qquad \alpha_{ij} = \alpha_{ij}^0 (1 + \sum_{l=1}^{n} \alpha_{ijl} \rho_l) . \tag{3}$$

In addition, we assume that $I_i$ has a form similar to the general form of the linear multi-component diffusion.

$$\vec{I}_i = -\sum_{j=1}^{n} D_{ij} \nabla \rho_i \tag{4}$$

where $D_{ij}$ are diffusion coefficients, assumed as constants. Here we will consider the particular case when the number of the members of the population does not influence the interaction between the members. For the case of one population and one spatial dimension, (1) reduces to the following model equation:

$$\frac{\partial \rho}{\partial t} - D \Delta \rho = r^0 \rho [1 - (\alpha^0 - r_{11}) \rho] \tag{5}$$

Next we re-write (4) in the following manner:

$$\frac{\partial \rho}{\partial t} - D\frac{\partial^2 \rho}{\partial x^2} = G\rho^2 + H\rho \tag{6}$$

where

$$G = r^0(r_{11} - \alpha^0 - 3\rho_0 r_{11}\alpha^0), \quad H = r^0[1 - \rho_0(3\alpha^0 r_{11}\rho_0 + 2r_{11} - 2\alpha^0)].$$

Below, we will treat (6) as a general form of a model equation, which for the particular case when $G$ and $H$ satisfy the above relationships describes population density waves.

## 3. Application of the Modified Method of Simplest Equation

We consider the equation (11) and search for solutions of the kind of travelling waves: $\rho(x,t) = \rho(\xi) = \rho(x - vt)$, where $v$ is the velocity of the wave [22,23]. We have to solve the equation:

$$D\frac{d^2\rho}{d\xi^2} + v\frac{d\rho}{d\xi} + G\rho^2 + H\rho = 0. \tag{7}$$

We assume that $\rho(\xi)$ has the form [20,21]:

$$\rho(\xi) = \sum_{i=0}^{n} a_i \phi^i, \quad \frac{d\phi}{d\xi} = \sqrt{\sum_{j=1}^{r} c_j \phi^j}, \tag{8}$$

where $a_i$ and $c_j$ are parameters that we will determine below. Following the steps of the modified method of the simplest equation, we substitute (14) in (13) and obtain an equation that contains powers of $\phi$. Next, we balance the highest power arising from the second derivative in (13) with the highest power arising in the term containing $\rho^3$ in the same equation. The resulting balance equation is: $r = n + 2, \quad n = 2,3,\ldots$.

If $n = 2$ then. $r = 4$. Assuming that $c_0 = c_1 = 0, c_2 = p^2 \neq 0, c_3 = 2pq, c_4 = q^2 \neq 0$, we search for a solution of the form:

$$\rho(\xi) = a_0 + a_1\phi + a_2\phi^2, \quad a_2 \neq 0, \quad \frac{d\phi}{d\xi} = p\phi + q\phi^2 \tag{9}$$

Substituting this in (7), we obtain the following system of 5 algebraic equations:

$$6Dq^2 + Ga_2 = 0$$

$$5Dpa_2 q + Dq^2 a_1 + va_2 q + Ga_1 a_2 = 0$$

$$3Da_1 q + 4Dp^2 a_2 + 2va_2 p + va_1 q + Ha_2 + 2Ga_0 a_2 + Ga_1^2 = 0$$

$$Dp^2 + 2Ga_0 + vp + Ha_a = 0$$

$$Ga_0 + Ha_0^2 = 0$$

The system implies that $a_1 = 0$. The solution of (17) is:

$$a_2 = -6\frac{Dq^2}{G}; \quad a_0 = -\frac{H}{G}; \quad v = \frac{5\sqrt{6HD}}{6}; \quad p = -\frac{v}{5D}. \tag{10}$$

The expression for the solitary wave depends on the solution of the differential equation in (6) and it is given by:

$$\rho(\xi) = a_0 + a_2\left(\frac{pe^{[p(\xi+c)]}}{1-qe^{[p(\xi+c)]}}\right)^2 \qquad (11)$$

where $a_0, a_2, p$ are given by (10), and $q$ is a free parameter.

## 4. Conclusion

In this paper, we discuss a model partial differential equation for description dynamics of the human population (the populations can be any social, ethnic, religious group of people, etc.). From a practical point of view Eq. (6) can be used for eventual prognosis and control of the human immigration, for example in the borders of a specific country. Each country can be separated into a number of regions for which some data on the number of immigrants can be found. Birth and death rates are often available from statistical data. From these sources estimations of the coefficients can be obtained. The most difficult task is the determination of the coefficients and, but they can be obtained by standard or appropriate statistical procedures as well. After determining the coefficients we can simulate Eq. (6) with known or appropriate initial conditions. Thereby we can obtain an approximate picture of the human immigration in the observed region for some period of time.

For the purposes here, we discussed in detail to a simple case of migration of the population in case of a human populations. In this case, the model is a nonlinear PDE. This equation describes the evolution of the spatial density of the individuals in the population. Solitary waves can travel through the system, if the amplitude of such density is high. In addition, using an appropriate method, we received exactly special analytical solution of the equation of the model. This decision describes the nonlinear solitary wave whims and expression of the distribution of the change in density in space.

## Acknowledgment

We would like thank to the "ISCH COST Action IS1104 the EU in the new complex geography of economic systems: models, tools and policy evaluation" for the support of our research.

## References

1. Витанов Николай К., З. Димитрова, С. Панчев. Популационна динамика и национална сигурност. 2005: Академично издателство "Марин Дринов" на БАН, 2005.
2. Витанов Николай К., З. Димитрова, С. Панчев. Социална динамика без формули. София: Академично издателство "Марин Дринов" на БАН, 2008.
3. Anderson R. M., R. May, Infectious diseases of humans. Dynamics and control. Oxford University Press, Oxford, 1992.
4. Dimitrova, Z. I., N. K. Vitanov, Influence of adaptation on the nonlinear dynamics of a system of competing populations. Phys. Lett. A 272, pp. 368-380, 2000.
5. Dimitrova, Z. I., N. K. Vitanov, Dynamical Consequences of Adaptation of Growth Rates in a System of Three Competing Populations. J. Phys. A: Math Gen., 34, pp. 7459-7473, 2001.
6. Dimitrova Z. I., N. K. Vitanov, Adaptation and its impact on the dynamics of a system of thire competing population. Physica A, 300, 91-115, 2001.
7. Dimitrova Z. I., Vitanov N. K., Shilnikov chaos in a generalized system for modelling dynamics of competing populations. Compt. rend. Acad. bulg. Sci., 58, N 3, pp. 257-264, 2005.

8. Dimitrova Z. I., N. K. Vitanov, I.P.Jordanov, On nonlinear dynamiks of agent systems: A case of competing population. Proceedings of the 10th Jubilee National Congress of Theoretical and Applied Mechanics, pp. 41-48, 2005.
9. Jordanov I. P. On the nonlinear waves in (2+1)-dimensional population systems. Comp. rend. Acad. Sci. Bulg. 61 (3), pp. 307-314, 2008.
10. Jordanov I. P. Nonlinear waves caused by diffusion of population members. Comp. rend. Acad. Sci. Bulg. 62 (1), pp. 33-40, 2009.
11. Kaplan, D., L. Glass, Understanding Nonlinear Dynamics, Springer, New York, 1995.
12. Malfliet, W., Solitary Wave Solutions of Nonlinear Wave Equations. American Journal of Physics, 60, pp. 650 – 654, 1992.
13. Martinov, N., N. Vitanov, On the Self-consistent Thermal Equilibrium Structures in Two-dimensional Negative-temperature Systems. Canadian Journal of Physics, 72, pp. 618–624, 1994.
14. Martinov, N., N. Vitanov, New Class of Running-wave Solutions of the 2+1-Dimensional Sine - Gordon Equation. J. Phys. A: Math. Gen., 27, pp. 4611 – 4618, 1994.
15. Montroll, E. W., W. W. Badger, Introduction to Quantitative Aspects of Social Phenomena, Gordon and Breach, New York, 1974.
16. Nikolova E. V., Jordanov I. P., Vitanov, N. K., Mathematical Modeling of the Migration of Human Populations, Proceedings of the ICAICTEE -2013, pp. 474-483, 2014.
17. Richmond P., L. Sabatelli, Peer pressure and generalized Lotka-Volterra models. Physica A, 344, 2004
18. Richmond P., L. Sabatelli, Langevin processes, agent models and socio-economic systems. Physica A, 336, 27, 2004.
19. Sheyretski K., Predictability of Dynamical Systems, Proceedings of the International Conference on Application of Information and Communication Technology and Statistics in Economy and Education-2013, pp. 484-488, 2014.
20. Vitanov, N. K., Application of simplest equations of Bernoulli and Riccati kind for obtaining exact traveling-wave solutions for a class of PDEs with polynomial nonlinearity, Communications in Nonlinear Science and Numerical Simulation 15 (8), pp. 2050-2060, 2010.
21. Vitanov, N. K., Modified method of simplest equation: powerful tool for obtaining exact and approximate traveling-wave solutions of nonlinear PDEs, Communications in Nonlinear Science and Numerical Simulation 16 (3), pp. 1176-1185, 2001.
22. Vitanov, N. K., Z. I. Dimitrova, Application of the method of simplest equation for obtaining exact traveling-wave solutions for two classes of model PDEs from ecology and population dynamics, Communications in Nonlinear Science and Numerical Simulation 15 (10), pp. 2836-2845, 2010.
23. Vitanov, N. K., Z. I. Dimitrova, Kantz, Modified method of simplest equation and its application to nonlinear PDEs, Communications in Nonlinear Science and Numerical Simulation 16 (3), pp. 1176-1185, 2011.
24. Vitanov, N. K., I. P. Jordanov, Z. I. Dimitrova, On nonlinear dynamics of interacting populations: Coupled kink waves in a system of two populations, Commun.Nonlinear Sci. Numer. Simul. 14 (5), pp. 2379–2388, 2009.
25. Vitanov, N. K., I. P. Jordanov, Z. I. Dimitrova, On nonlinear population waves, Applied Mathematics and Computation 215 (8), pp. 2950–2964, 2009.

# Optimization of the Shape of the Car Body with FEM Analyis

**Petko Nedyalkov[1], Ivan Ivanov[2] Miroslava J. Ivanova[3]**

[1]Technical University of Sofia,
Mechanical Engineering Faculty,
1000 Sofia, Bulgaria
nedpetko@tu-sofia.bg

[2,3]UNWE, 1700 Sofia, Bulgaria
[2]ivanov_i_s@abv.bg [3]mi_jordanova@abv.bg

**Abstract.** The last few decades have seen a deepening of economic problems in both developed and developing countries. Many economical subjects fall into "energy dependence" because they must import basic types of fuels or energy sources. It is therefore obvious that the economy should aim to optimize its products to reduce energy consumption. It is essential to optimize fuel consumption in road transport. In this article it is explored the shape and configurations of the car compartment by fluid-flow analysis tools in FEM. The study is based on comparative analysis of full-scale geometric model of car bodies. It is introduced numerical criteria calculated through FEM model for evaluation of different configurations - aerodynamic drag coefficient, aerodynamic resistance force and power to overcome aerodynamic resistances. Numerical modeling and stability of the results for this simulation testing allowed a design body of a car in reduction of about twice the quantitative data for several iterations.

**Keywords.** Energy efficiency, aerodynamic resistance, fluid simulation, CFD, FEM.

## 1.Introduction

Energy efficiency is both technically challenging examination by technical problems and their solution and economic challenge in terms regarding consumption of resources and energy. In this aspect, examining the fluid dynamics - aerodynamic and hydrodynamic, has their peculiarities and advantages and disadvantages. Treatment in this case exclusively relies on the comparison between the simulation of several geometrically similar panels because theoretically - calculation methods are not sufficiently precise or add idealizations to models which do not allow refine results [5].

In globally development and accessibility of fluid analyzes in recent years has led to major development of the turbo, optimizing the aerodynamic performance of cars and other vehicles, optimization of the cooling of electronic systems and thereby reduce the dimensions of the cooling radiators and boxes of electronic devices.

In the case discussed optimization of the external shape of the car bodies designed for racing car to participate in the competition for energy efficiency. The dimensions of the body are bound by the rules of competition, the structure of the suspension, frame, steering, wheels and openings for them specified in advance. The purpose of the development and optimization of the body is to lower the utmost resistance force of aerodynamic resistances of cars on the road.

## 2.Theoretical Background

Comparison between different types of compartments and/or surfaces is appropriate to use a coefficient of drag, total drag force and power spent to overcome the aerodynamic resistance. These parameters are integrated criteria, as they contain the surface area as a parameter, where the dynamic pressure of the flow acts. Also, these parameters are dependent on the flow and velocity rate, which generally can not be uniform or fixed to a specific value. Due to the above facts, those parameters are difficult calculable in theoretical or analytical modeling [2, 4]. Using FEM fluid analysis these parameters are introduced for calculation of the computing module (solver) and are reported as a numerical result after conducting a simulation analysis.

It is introduced the following criteria:

• drag force -

$$F_d = c_d \cdot p_d \cdot A_d = c_d \cdot \frac{1}{2} \cdot \rho \cdot \mathrm{v}^2 \cdot A_d, \quad N \ , \tag{5.3.1}$$

where:

• dynamic pressure -

$$p_d = \frac{1}{2} \cdot \rho \cdot \mathrm{v}^2, Pa \tag{5.3.2}$$

• drag area - $A_d$ ;

• drag coefficient -

$$c_d = \frac{2 \cdot F_d}{\rho \cdot \mathrm{v}^2 \cdot A_d}, \tag{5.3.3}$$

• drag power -

$$P_d = F_d \cdot \mathrm{v}, \quad W \ ; \tag{5.3.4}$$

Results such as - distribution of the flow rate, the dynamic pressure of the flow on the surface of the passenger compartment provides information on the obstacles and flow splitting and separation of the laminar and turbulent layers which aids in the removal of obstructions and local modification of the total body shape.

## 3.Software and model setup

In the creation of numerical model in FEM first step is the cross-linking of the geometric model and in this case was used the volume of a parallelepiped with a spacing of the walls about 3 times the linear dimensions of the passenger compartment in the direction except the floor which is level with the lower tangent line of the tire. Dimensions of the geometric model in scale 1: 1 in order to avoid the transfer of results through different sizes similarity theory.

Removed volumes are formed and united as one body to avoid having to create links between mesh models which extremely aggravated and complicated cross-linking meshing processes. Fig. 1 shows in the mesh model in the axial section along the long axis of the car body [1, 3]. A subjective criterion to track the accuracy of a meshing in this case is the difference between the dimensions of the individual elements along tetrahedral mesh elements. In the case clearly see the large size of the tetrahedral in the free volumes and different comminution in case of sections and volumes of the cabin.

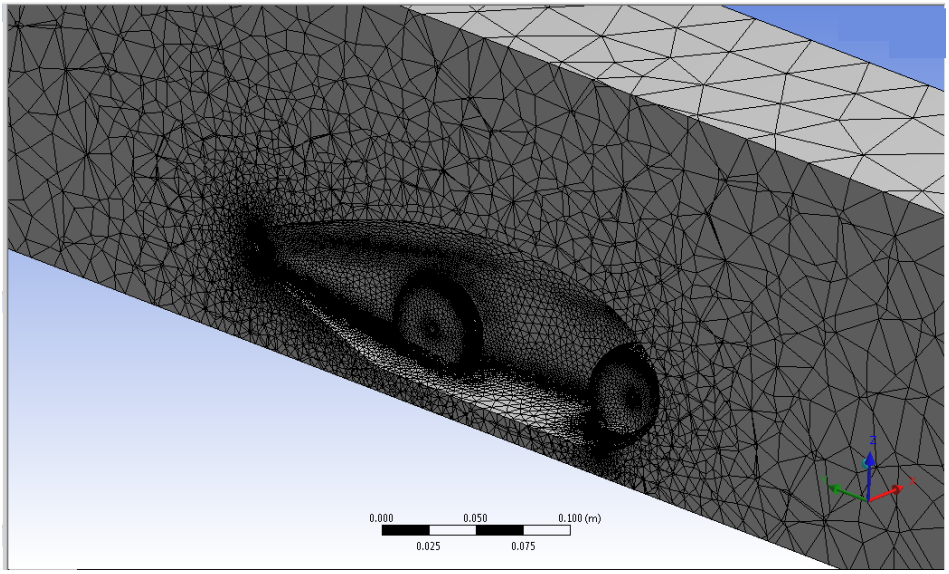



Fig. 1. Mesh setup and preparation of the simulation model (intersection).

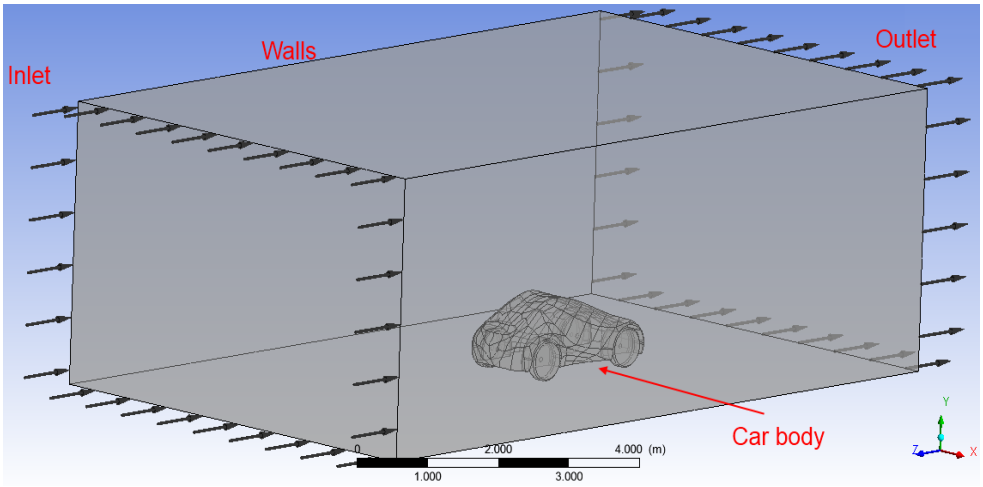



Fig. 2. Model setup and preparation of the simulation model (intersection).

The configuration of the model as it is shown on fig. 2, uses:

- fixed walls, floor and ceiling;
- an inlet flow from the nose of the vehicle to control of the rate of flow, uses flow velocity rates between 25 and 50 km / h, such as speed ranges in city traffic and race;
- outlet flow from the tail of the car with the control of the output type – or a normal flow rate, or static pressure on the output surface, the two kinds of output do not substantially alter the results;

• type of fluid – air at 20 ºC, with the control of turbulence, for these simulation tests shall not be tested turbulent flows and seek their prevention;

• the absence of other sources of air or water flow, and lack of heat sources

## 4.Research and Results

Consider the case comparative analysis between evolutionarily developed car bodies, as base used body of Lyna H2 of 2010 shown in fig. 3.a. and developed several versions of the coupe ECO from 2012 - fig. 4 as a final version is shown in fig. 3.b.

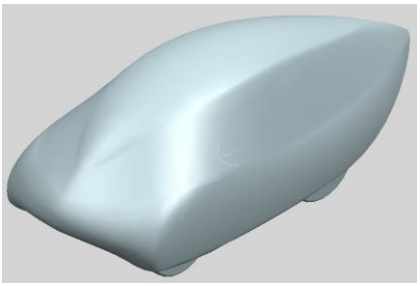
a)  b) 

Fig. 3. Evolution and development of automotive body
a) version Lyna H2 – 2010 , b) version ECO 1 – 05 (final) – 2012.

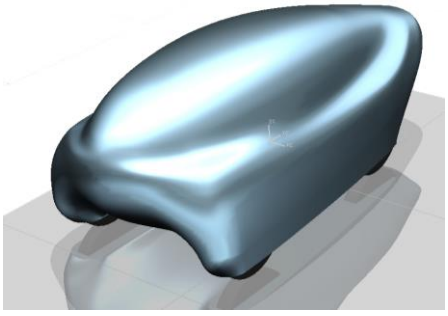
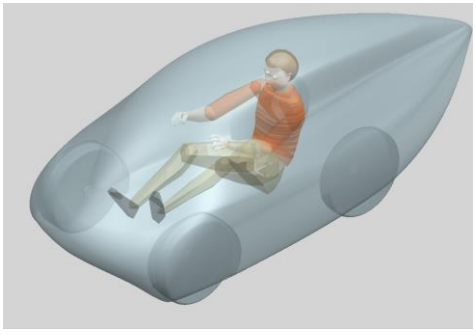
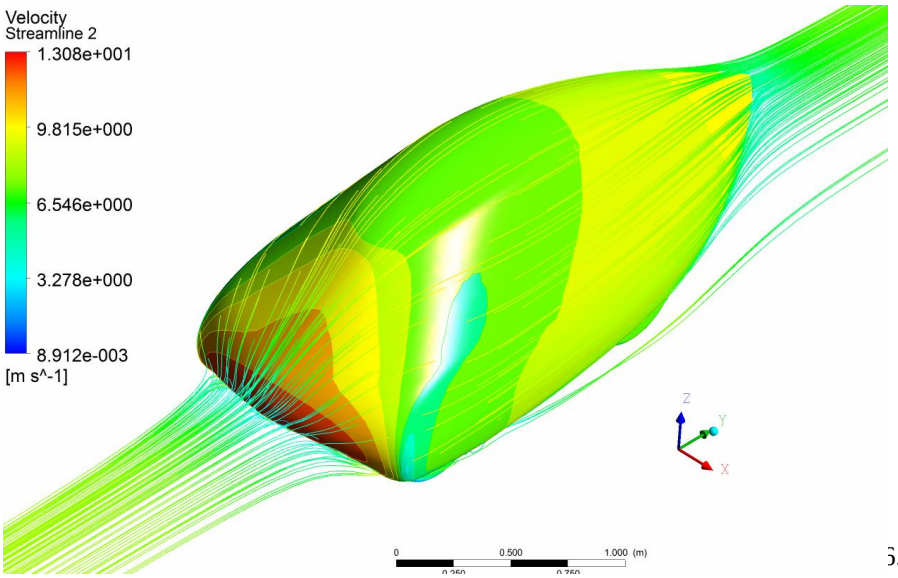
a)  b) 

Fig. 4. Version development of automotive body ECO 1,
a) ECO1-04, b) ECO1-05.

Subjective indicators in aerodynamic research with FEM over the presented bodies are shown in a series of figures.

Fig.5 shows the streamline distribution of velocity flow over the surface of the passenger compartment. A diagram such like this is used in iteration design process to remove and improve behavior of obstacles and flow splitting and analysis of dynamically loaded areas of the body.




On fig. 6 to 9 shows similar streamline flow velocity distributions combined with area distribution on the surface of the car from 2010. Importance mention is the influence of car tail, fenders and openness of tires.

Comparison on such subjective criteria would serve on the form optimization, but criteria comparison could be performed on numeric criteria in table 1 and 2.

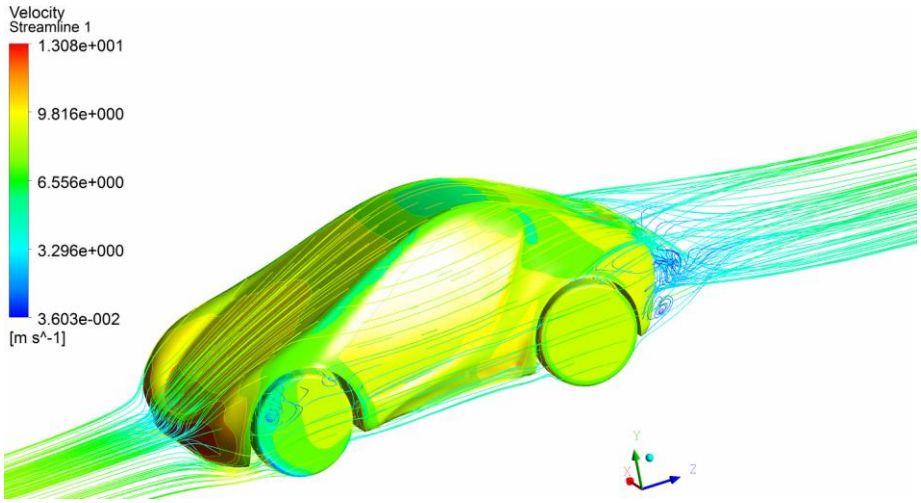



Fig. 6. Speed flow and of dynamic pressure map over the body variant Lyna H2.

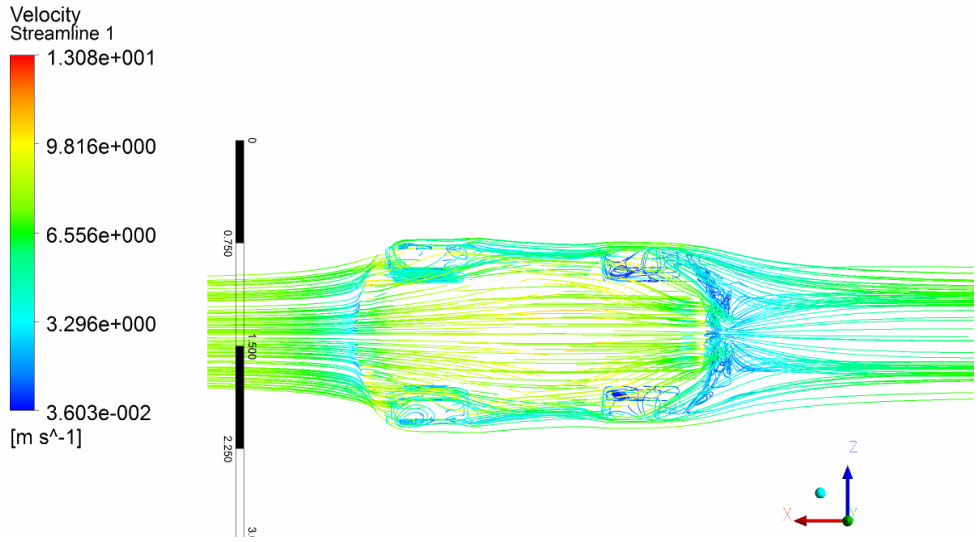



Fig. 7. Speed flow in the plane XZ – streamline

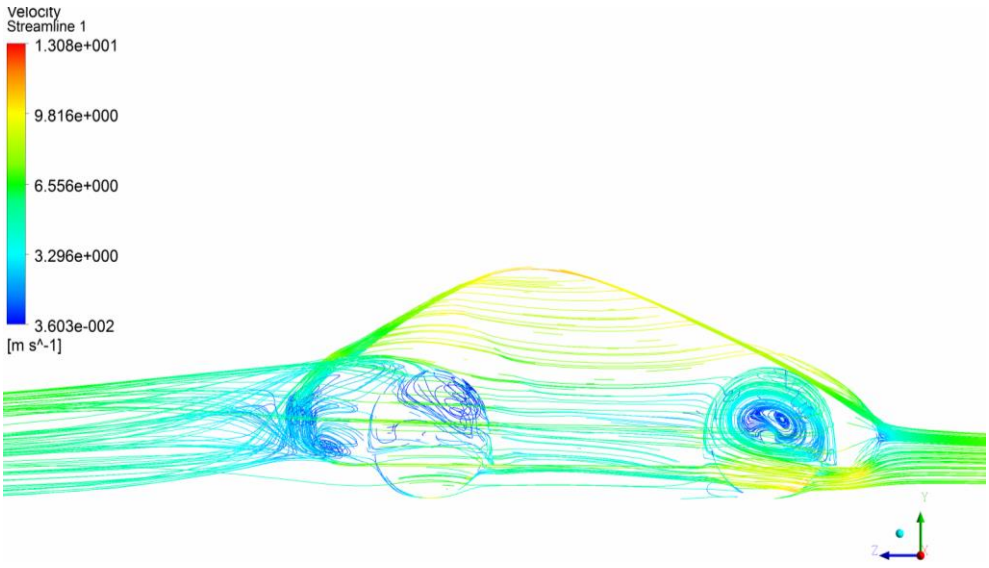



Fig. 8. Speed flow in the plane YZ – streamline presentation version Lyna H2.

**Table 1.** Comparative results for version Lyna H2

| output control | outlet static pressure | | |
|---|---|---|---|
| mesh | coarse | fine | fine |
| model | *Lyna H2* | *Lyna H2* | *Lyna H2* |
| speed, km/h | 30 | 30 | 50 |
| cD | 0.147 | 0.143 | 0.138 |
| Drag Force, N | 24.566 | 23.847 | 65.412 |
| Drag Power, W | 204.72 | 198.73 | 908.50 |

**Table 2.** Comparative results for versions ECO

| output control | outlet - normal speed | | | outlet static pressure | | |
|---|---|---|---|---|---|---|
| mesh | coarse | medium | fine | fine | fine | fine |
| model | *05n119* | *05n119* | *05n119* | *05n119* | *05n119* | *05n119* |
| speed, km/h | 25 | 25 | 25 | 25 | 35 | 45 |
| cD | 0.081 | 0.079 | 0.078 | 0.079 | 0.077 | 0.075 |
| Drag Force, N | 13 | 12.6 | 12.374 | 12.537 | 24.018 | 39.042 |
| Drag Power, W | 90.28 | 87.50 | 85.93 | 87.06 | 233.51 | 488.03 |

## 5.Conclusion

Under this comparative analysis is used full-scale geometric model of car bodies in order to directly transfer the results to the real car or a full-scale wind tunnel test. Large-scale tests at reduced scale body sense when using an experimental trial in miniature wind tunnel and in the case is not preferred. Control over mesh refinement in FEM shows satisfactory convergence of the results at medium and high reference values suggested from software automation, as in this case is preferred to use high refinement (fine mesh).

Numerical modeling and stability of the results in this simulation testing is balanced with time for configuring the model and its solution, in which case (2012) at the computing and computational software to calculate a model configured in this manner is spent about 1 hour computing time on a personal computer with a dual-core processor and 4 GB of RAM. The results of numerical modeling of flood dynamic FEM analysis allowed to design a coupe race car participating in the competition for Energy efficiency in 2012 and 2013. The performance of the new body expressed by numerical indicators coefficient of drag resistance and total force of aerodynamic resistances were lowered twice for about 50 - ish iterations. The body is produced and show expected performance and racing cars won two second places in their respective classes in 2012 and 2013 in participation of Shell ECO Marathon Europe. The average power consumption of the car when driving at a speed in the range 30-35 km/h is in the range of 300-600 W, which shows good experimental convergence between the numerical results and the behavior of the track

Over discussed methodology are developed training exercises in "Computer Simulation Analysis" and "Integrated Design and Production Systems" of students from the Mechanical Engineering Faculty of the Technical University - Sofia.

## References

1. ANSYS CFD Help
2. Damyanov Tsvetan, Lachezar Lazov, Petko Nedyalkov, "Mathematical modeling and modal analysis of the two drum driver system for belt conveyors". ISSN: 1312 – 1820Annual of the University of Mining and Geology "St. Ivan Rilski", Vol. 51, Part III, Mechanization, electrification and automation in mines, pp. 39 - 42, publ. UMG "St. Ivan Rilski, Sofia, 2008.
3. http://www.cfd-online.com/Wiki/Main_Page
4. Nedyalkov Petko, FEM modelling and characteristics research of cylindrical journal bearing for cone inertial crusher /KID – 300. ISSN 1582-0246, RECENT Vol 15 (2014), No 2(42), pp. 112-116, Romania, 2014
5. Vitanov N. K., I.P. Jordanov, I. S. Ivanov, "Time Series Analysis of Successful Economics Regulation Politics". ISBN 978-954-644-586-5, 3-rd ICAICTSEE 2013, pp. 489-494, Sofia, 2013.

# Isomorphism Between Economic Models and Models of Natural Environment

**Kostadin Sheyretski, Meglena Lazarova**

UNSS, Studentski grad, "8-th December" blvd, 1700 Sofia, Bulgaria
sheyretski@unwe.bg
meglena.laz@unwe.bg

**Abstract:** In this paper some models of natural environment suitable for economic applications are considered. The development of methods for studing of some dynamic systems in fundamental science and its practical applications is favorable for modeling of economic processes based on similar mathematical models. Some prognoses and planning of economic phenomena, crises etc. can be done by means of analytic and numerical mathematical methods.

**Keywords:** dynamical systems, economic models

## 1.Introduction

The modern science imposes an integration of different fields for solving of concrete economical, technological or scientific problems. In the last few days some interdisciplinary fileds like synergetics, econophysics and etc. received a great development.

The purpose of this report is to review a brief survey of some modern ways in the development of economics on the base of symbiosis between physical and economical models. The development of the modern digital technics gives way to the analysis of complex multi-dimensional systems which we can see in the process of studying of many phenomena in the field of natural sciences.

Finding of some isomorphous models which make a relation between the economics and hydrodynamics, the plasm physics and the atmospheric physics gives us an opportunity to use the obtained results in the description of complex economical processes and also to forecast crises and disastrous phenomena.

## 2.Tsunami and Shock Waves

One of the manifestation of the nonlinear phenomena are the *solitons* or the so called localized waves. They are found by Jhon Scot Russel who observed the stopping of the barge pulled up by two horses in a narrow channel. The wave broke off from its bow part and with a great speed began to go forward like a small hill, which continued its way in the channel with a constant speed without changing its form.

For the solitons, defined by the equation of Kordeveg - de Vriz is specific that the soliton is as higher as it's narrower and if the soliton is wider then its movement is slower and also its amplitude is smaller. The soliton's balance is keeping up from the balance between the non-linearity and the dispersion. Also the non-linear properties are known for the tsunami waves. The tsunami seems to be harmless in the open sea and it's height in the see is under 30 centimeters. By the way the distance between the crests may reach hundreds

of kilometers. So a huge mass of water is concentrated in the tsunami, which sometimes is moving with 180 kilometers per hour.

The tsunami slows down its movement in the seacoasts, especially in the shallows and its height is at about tens of meters. After that this powerful wave is spilling over the shore as it creates huge destructions.

Interesting phenomena are the *shock waves*. One typical example for shock wave is the eruption of volcanoes. In 1883 a volcano, situated in the inhabitable Ireland with volcano origin – Krakatau between the Irelands Java and Sumatra erupts. The shock wave goes around the earth circle three times. The explosion was heard in 4 800 kilometers from the source and at about 4,6 cubic kilometers of volcanic ashes were thrown up in the atmosphere. The energy of the explosion was at about 250 thousand bombs thrown up over Hiroshima (5000 Mt trotyl).

The law of the shock wave's movement can be given by the method of dimensions. The wave's distance from the center at the moment of the explosion $r$ depends on the variable – time $t$, the energy of the explosion $E$ and the density of the atmosphere $\rho$.

| $\dim r$ | $E$ | $\rho$ | $t$ |
|---|---|---|---|
| $L$ 1 | 2 | $-3$ | 0 |
| $M$ 0 | 1 | 1 | 0 |
| $T$ 0 | $-2$ | 0 | 1 |

When we solve the given linear equations we obtain the following:

$$(1) \qquad r = C\left(\frac{E}{\rho}\right)^{\frac{1}{5}} t^{\frac{2}{5}}, \quad v = \frac{2}{5} C\left(\frac{E}{\rho}\right)^{\frac{1}{5}} t^{-\frac{3}{5}},$$

where $C$ is a constant and $v$ is the speed of the shock wave.

In the modern context the economic's development still more is entering the methodology of economic dynamics, which investigates the unstable systems by contrast with the orthodox economic theories. In the work [2] is made an intermediate approach, used for studing of some wave's properties of the economic subjects. In author's opinion there are two brightest manifestations of the economic soliton. The first one is the shady economics and the second one - the urban marginal societies.

The waves which describe the collective behavior of the economic subjects are quite difficult case of spreading the multi-dimensional non-linear oscillations. More over these non-linear waves are not spreading independently. They are also interacting. Similar phenomena are observed in studying of the plasm or in the hydrodynamic systems.

The solitons in the economics are defined as a stable localization in specified field of the economic density's space of the probability some microeconomic subjects to fall into this concrete field. The economic soliton's introduction enables to work out three alternative forms of collective behavior of the economic trace elements: chaotic, organized with localized instability and organized with stable localization in the economic space – solitone's type. From the economic development's point of view, the soliton's form is the most preferable for structuring the economic development. The main reasons for this can be found in the opportunity for prognosis and analysis of such systems. The economic policy is the main instrument which can control the soliton. This can be expressed by a model which contains all the effects' combinations with different character over the economic trace

elements' behavior.

The shock waves' effect is usually observed when there is an economic market collapse and also in the suddenly arised economic crises. Still in the literature there is a low presence of such models that use nonlinear waves, but the closer integration of the different sciences will bring the opportunity of making such models that can assist the quick and effective decisions about crises situations and even for disastrous phenomena.

## 3.Intensification's restrictions in the nature and in the economic

Sergei Pavlovich Kurdjumov [1,10] analyzes over quick processes like explosions, instabilities in plasm and etc. The attendant description of such processes is called *a restriction with intensification* when one or some variables in the studied system is growing indefinitely for a short time.

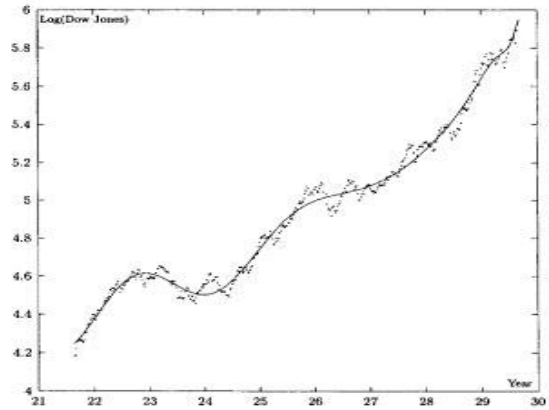



Figure 1. Dow Jones index's logarithm before the economic collapse on 23 October 1929. The dense line corresponds to the approximation of the formula (2).

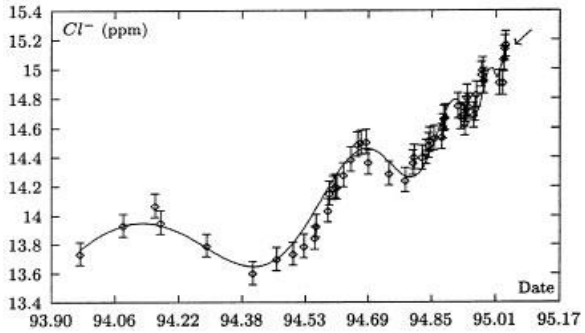



Figure 2. The chlorine ions' concentration in the underground sources before the earthquake in Kobe on 17.01.1995.

This theory has an immediately application to the crises processes. On Fig.1 and Fig.2 are represented the two complex systems' indicators due to we may judge their dynamics. On Fig.1 is seen the Dow Jones index's logarithm before the Great Depression, where it is observed a quick development with growing oscillations, which can be a presage of a catastrophe. On Fig. 2 is shown the chlorine ions' concentration in the mineral sources

before the huge earthquake in Kobe. In the two cases is seen one and the same behavior with different scale and character of the systems. A good approximation for the two processes is given by the formula:

(2) $$I(t) = A + B(t_f - t)^{\alpha}\left[1 + \cos(\omega \log(t_f - t))\right],$$

where $\alpha < 0, A, B, \omega$ are costants. As $\alpha < 0$ then $I(t) \to \infty$ when $t \to t_f$ where $t_f$ is intensification's restriction. A special feature of synergetic is that it makes a relation among totally different systems which have a similar behavior.

## 4.Similarity between the climate and the economic models

The planet's climate is defined by three main factors [4].

**a/ Astronomical factors** – the Sun like main source of energy, the Earth's movement parameters – eccentricity, big semi-axis and the axis's slope. The Moon's influence is also important.

**b/ External geophysical factors –** Earth's shape, dimensions and mass, rotation's speed around its axis, the gravitational and magnetic field, volcanic properties and etc.

**c/ Internal geophysical factors –** an atmosphere's mass and content, also the surface's character and properties.

For the internal factors there is a negative back relation: the worming of the weather leads to evaporation and increasing of the humidity, which enhances its radiation cooling.

Also there exists a positive back relation – the temperature's decreasing leads to the increasing of the areas, covered by snow and ice, which have a great albedo i.e. a reflection coefficient of the solar radiation. This decreases the absorbed heat's quantity and leads to temperature's decreasing. The cloudiness and the temperature have as positive and negative back relation. The temperature's increasing leads to the cloudiness's increasing. This leads to bigger solar rays' reflection and finally the temperature is decreasing. But the clouds strongly absorb the earth thermal radiating and with the back solar radiating they decrease the effective radiating and this leads to temperature's increasing. If the system evolves to one attractor and has one and the same characteristics in a long period of time then the system is called *a transitive*. If the system's behavior depends on the initial conditions in time, then it's called *an intransitive* system. Lorentz thinks that the climate system is almost intransitive. For a short period of time the meteorological system is stable – transitive. An example for intransitivity is the fact that from constant external factors, from one and the same initial states the system may evolve in different ways. An example for system's atmosphere-ocean intransitivity is the stream El Ninio. Once in two to ten years in the east equatorial part of the Pacific Ocean (Peru's and Ecuador's coasts) usually on the oceanic's surface are shown warm waters for few months before and after Christmas.

The hypothesis for the intransitivity gives the opportunity some significate climate changes to be explained only by the external factors. This means the unpredictability and the possibility that the anthropogenic factors have strongly influence on the climate. The unpredictability is strengthened if we also take in mind the outer factors.

Lorentz had examined a thermo convection of warmed up from below fluid's layer in the field of the weight's power. In this model the warmer fluid's parts are becoming lighter at the expense of the thermal expansion. Archimedes' power pushes these parts out and after that on their place come other colder layers. So Lorentz worked out three ordinary differential equations [3] which in spite of their simplicity have complex properties.

*The Lorentz's system* is:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y-x), \\ \frac{dy}{dt} &= rx - y - xz, \\ \frac{dz}{dt} &= xy - bz. \end{aligned} \tag{3}$$

In the upper equations the parameters $\sigma, r, b$ are nondimentional parameters. The variables $x, y, z$ are variables that characterize the system, and $t$ is a parameter for time.

On certain parameters' values the system has a chaotic behavior and in the state space is settled the so called *strange attractor*. In common the dynamic system is making a complex chaotic movement between two unstable focuses – the stationary points beat off the points which describe the state trajectory and are moving on a spiral. Starting its movement from an arbitrary point the system gets into unstable focus where it starts to move on an increasing spiral. After some rotations the system unexpectedly goes to the second focus where it also becomes to move on an increasing spiral. After some time the system jumps again on the first focus and so on. No periodicity is observed on this movement. The rotation's numbers around the spiral and the time being spent around the focuses are completely random.

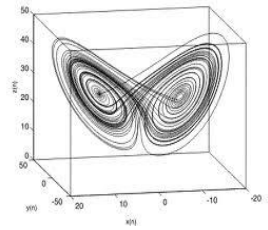



Fig. 3 Lorentz's attractor

We will give an example which will connect the Lorentz's system with an economic model. The model is considered in a big city where a subsystem without influence on the economic conditions exists. We accept that the economic conditions are stationary in a given period of time. We take the local system's characteristics as follows:

$X$ – the output produced by the city subsystem,

$Y$ – the local population's number,

$Z$ – the ground rent.

The output may be consumed by the local population or it may be exported. The model consists of a system of three differential equations of first order. There is one differential equation for every variable. Using some mathematical transformations we obtain the Lorentz's system. In short this city subsystem's economic behavior is equivalent to the climate system's behavior given by the meteorologists. Some analogous problems for forecasting the system's future condition and management arise. It is clear that even so simple the economic model have the property to be unpredictable and dependent on the determined chaos. This fact shows that the chaotic behavior is an irreversible part of the complex economic processes. The solution of some basic questions in the science ***"Chaotic dynamics"*** will bring the clearer understanding of the economic phenomena which are

considered to be a product of the stochastic. The break in the real problems of the nonlinear dynamic systems will bring a new reading of the economic theories and models. As the Lorentz's system have no claims on describing the whole complexity of the climate systems, so every economic model can't completely be adequate to the observed behavior. However the value of such models is that in simple form they give tendencies and important characteristics of the system which in a complicated model hardly can be observed in a pure form.

The Lorentz's system is a *dissipative system*. For a conservative system where the energy is constant in time, the Liouville's theorem holds true – the state volume is retained. Also the Poincare's theorem holds true – almost every point from an arbitrary area of the state space, moving on the trajectory will return to the same area again. In dissipative systems the state volume is compressing. That's why the dissipative character and the instability of the state trajectories give an opportunity in certain system's parameters to have a chaotic regime.

## 5.Conclusion

From the given examples is seen how perspective is the connection among the different fields of science. Such method of approach can bring to an economic phenomena's prognosis which at this stage of development are considered to be too complex and with slight known manage mechanism. On the other hand such method of approach helps the management and the optimization of the economic systems. The interdisciplinary fields are main instrument of taking some difficult decisions in crisis situations. They can be used for stability and limitation of the negative consequences.

## References

1. *Future Applied mathematics. Lecture notes for young explorers*, Moscow, URSS 2004 (in Russian)
2. Ogorodnikova T. *The economic soliton like a stable form of the macroeconomic subjects' collective wave behavior*, Notifications IGEA. 2004 Num 2. (39) (in Russian)
3. Lorentz, E. *The nature of Chaos*, Sofia University Publishing House St. Kl. Ohridski, 2000. (in Bulgarian)
4. Pantchev, St. *Chaos Theory*, Sofia, Academic Publishing House" Prof. Marin Drinov", 1996. (in Bulgarian)
5. Puu, T. *Nonlinear economic dynamics, Lecture notes in Economic and Math. Systems*, Vol. 336 (Springer Verlag), 1989.
6. Trubetzkov, D. I. *An introduction to Synergetics, Fluctuations and waves*, Moscow URSS, 2011 (in Russian)
7. Trubetzkov, D. I. *An introduction in Synergetics, Chaos and structure*, Moscow URSS, 2012 (in Russian)
8. Zang, B.V. *Synergetic economy, Time and change in the nonlinear economic theory*, M Peace 1999 (in Russian)
9. http://www.ufpi.br/menezes/materias/index/mostrar/id/5486
10. http://spkurdyumov.narod.ru

# Analysis of Virtual Waiting Time of PH|G|1 Queue Under Heavy Traffic

**Mitko Dimitrov**

UNWE, 1700 Sofia, Bulgaria
mcdimitrov@unwe.bg

**Abstract. .** We consider a PH|G|1 queue and prove a heavy traffic limit theorem for the distribution of the stationary virtual waiting time W. If the traffic load $\rho \to 1$, then W multiplied by an appropriate "coefficient of contraction", that is a function of $\rho$, converges in distribution to exponential distribution.

**Keywords:** PH|G|1 queue, Virtual waiting time, Heavy traffic limit theorem.

## 1.Introduction

The $PH|G|1$ queue was introduced and investigated by Marcel Neuts [1,3]. This paper studies $PH|G|1$ queue under heavy traffic assumptions. We first derive the mean value of the virtual waiting time using the Laplace-Stiltjes Transform of virtual waiting time, and then present heavy traffic limit theorem for the virtual waiting time in steady state. Obviously, rigorous proof of the convergence of the stationary distribution and its mean value requires more complex types of convergence of the stochastic processes and is not a trivial mathematical problem. In this paper we use another approach to the problem. It is based on certain "semi-explicit formulas" for the stationary distribution of the virtual waiting time and its mean values and gives a rigorous proof of the heavy traffic limit theorem directly for the desired characteristic.

This paper is organized in four sections. In Section two we introduce some preliminary results and derive some basic results. In Section three we introduce the $PH|G|1$ queue. In Section four we derive the main results. In Section five we draw the conclusions.

## 2. Some preliminary results

Let us consider an irreducible continuous Markov chain with finite state space $\{0,1,\cdots,K\}$ with infinitesimal generator $Q=\|q_{ij}\|$, $q_{ii}=-\sum_{j\neq i} q_{ij,}$ $q_i=-q_{ii}$ and stationary probability distribution denoted $\pi=(\pi_0,\pi_1,\cdots,\pi_K)$. We will use further the solution $a=(a_0,a_1,\cdots,a_K)$ to the set of linear equations

$$Q(x_0,x_{1,\cdots,}x_K)^T=(b_0,b_1,\cdots,b_K)^T \tag{1}$$

Obviously, the set of linear equations has a solution iff the vector $b=(b_0,b_1,\cdots,b_K)$ is orthogonal to the vector $\pi$. From $(1)$ and $rangQ=K$ follows that the matrix $A=\|q_{ij}\|_{1\le i,j\le K}$ is nonsingular and $(q_{01},q_{02},\cdots,q_{0K})A^{-1}=-\frac{1}{\pi_0}(\pi_1,\pi_2,\cdots,\pi_K)$. Let us introduce the matrix $R$. The first row and first column of matrix $R$ are zero vectors, and next $K$ rows and $K$ columns form matrix $A^{-1}$. Then matrix $QR$ looks as follows: the first column is equal to zero vector ,the first row is equal to $\left(0,-\frac{\pi_1}{\pi_0},\cdots,-\frac{\pi_K}{\pi_0}\right)$, and the other elements form the identity matrix and for any vector $x=(x_0,x_1,\cdots,x_K)$ we have

$$xQR=x-x_0\frac{\pi}{\pi_0}. \tag{2}$$

It should be noted that the variables $a_0,a_1,\cdots,a_K$ introduced with the equations (1) (for $a_0=0$) could be obtained with the help of the matrix $R$ as follows

$$(0,a_1,\cdots,a_K)=Rb^T. \tag{3}$$

## 3. Model description

A (continuous) probability distribution of phase type, introduced by M. Neuts [1] is any continuous probability distribution on $[0,\infty)$ which is obtainable as the distribution of the time till absorption in a continuous time finite state space Markov chain with a single absorbing state into which the absorption is the certain. The class of such distributions includes a number of well - known particular cases such as generalized Erlang and hyper exponential (i.e., a mixture of a finite number of exponentials) distributions and due to its interesting closure properties constitutes a versatile class with properties especially useful in the algorithmic solution of several queueing models. Let us consider a Markov chain $Z(t)$ with state space $\{0,1,\cdots,K+1\}$, initial probability vector $(\alpha,\alpha_{K+1})$ and infinitesimal generator $Q=\begin{pmatrix}T & T^0\\ 0 & 0\end{pmatrix}$, where $\alpha=(\alpha_0,\alpha_1,\cdots,\alpha_K)$. Matrix $T$ is nonsingular $(K+1)\times(K+1)$ matrix with $T_{ii}<0$ and $T_{ij}\ge 0$ for $i\ne j$, and $T^0\ge 0$ is an $K+1-$ vector satisfying $Te+T^0=0$ with $e^T=(1,\cdots,1)$. For such a Markov chain the probability distribution of the time till absorption in K+1 state, is given by the cdf, $A(x)=1-\alpha\exp(Tx)e, x\ge 0$. Each time the Markov chain $Z(t)$ becomes absorbed in the state $(K+1)$, restart it by performing a multinomial trial with possible outcomes $0,1,\cdots,K$, and probabilities $\alpha_0,\alpha_1,\cdots,\alpha_K$ to pick a new "initial state". Considering each absorption into the state $K+1$ as renewal we obtain a renewal process for which the time

between any two successive renewals has cdf $A(x)$. Such a renewal process is called a renewal process of phase type ($PH-$renewal process). The above procedure also constructively defines a new Markov chain with state space $\{0,1,\cdots,K\}$ initial probability vector $\alpha$ and infinitesimal generator $Q^* = T + T^0\alpha$. In [1,2] it is shown that one may without loss of generality, assume the representation $(\alpha, T)$ of $A(t)$ is so chosen as to make $Q^*$ irreducible and we shal henceforth assume that this is indeed the case.

We let $\pi = (\pi_0, \pi_1, \cdots, \pi_K)$ denote the stationary probability vector of the Markov chain $Q^*$ i.e., the unique (strictly positive) vector satisfying $\pi Q^* = 0,\ \pi e = 1$.It may be easily verified that $\pi = -\beta\alpha T^{-1}$, where $\beta^{-1} = -\alpha T^{-1} e$ is the mean of $A(x)$.

By $PH\,|\,G\,|\,1$ we denote a queue in which the arrival process is a $PH-$renewal process. Let us assume there is a single server and the queue discipline is first come, first served.

The arrival process is the $PH-$renewal with representation $(\alpha, T)$ and the service time cdf is given by $H(x)$, $h(s)$ is Laplace-Stiltjes Transform, $h_1$ and $h_2$ the first and the second moments of the service time. The $PH\,|\,G\,|\,1$ queue is stable iff $\rho = \pi T_0 h_1 < 1$ and $\rho$ is the traffic intensity of the $PH\,|\,G\,|\,1$ queue. Also let $J(t)$ and $W(t)$ denote the current phase of the customer arriving at time $t$, and the virtual waiting time at the same time.

Our standing point is the Markov process $(J(t), W(t), t \geq 0)$. Let $F(X) = P(J = i, W < x)$ is the stationary distribution of the phase of the arrival process and the virtual waiting time with the LST $\varphi_i(s)$, where

$$\varphi(s) = sF(0)(sI + T + T^0\alpha h(s))^{-1} \tag{4}$$

$$F(0) = (F_0(0), F_1(0), \cdots, F_K(0)),\ \varphi(s) = (\varphi_0(s), \varphi_1(s), \cdots, \varphi_K(s)),$$

$$F_i(0) = P(J = i, W = 0).$$

## 4. The Main Results

Our goal is to find asymptotic behavior of virtual waiting time in steady state under heavy traffic, when $\rho \equiv \alpha T^0 h_1 \to 1^-0$.To avoid unnecessary complication with minor details we consider the following specific problem setting: we assume $T = \lambda S$ and $S^0 = -Se$, and all parameters except for $\lambda$ are fixed, and $\lambda$ increases in such a way that $\rho = \lambda \pi S^0 h_1 \to 1^-0$. Obviously, the vector $\pi$ is stationary distribution of Markov chain with infinitesimal matrix $S + S^0\alpha$. Now, we will consider the set of equations (1) with $Q = S + S^0\alpha$ and $b^T = (\pi S^0)e - S^0$, or

$$(S + S^0\alpha)a = (\pi S^0)e - S^0. \tag{5}$$

Theorem 1. The mean virtual waiting time is given by the formula

$$EW = \frac{1}{1-\lambda\pi S^0 h_1}[\lambda\pi\frac{S^0 h_2}{2} + \pi\left(\lambda S^0\alpha h_1 - I\right)ah_1 + F\left(0\right)ah_1]. \quad (6)$$

Proof: From (4) follow equations

$\varphi\left(s\right)(T+T^0\alpha) = \varphi(s)(T^0\alpha(1-h(s)) - sI) + sF(0)$ and

$$\varphi\left(s\right)\lambda(S+S^0\alpha) = \varphi(s)(\lambda S^0\alpha(1-h(s)) - sI) + sF(0) \quad (7)$$

After multiplying equation (7) from the right by vector of ones $e = \left(1,\cdots,1\right)^T$ we get

$$F(0)e = \varphi(s)(I - \lambda\frac{1-h(s)}{s}S^0\alpha)e. \quad (8)$$

After we let $s \to 0$ we obtain relation $F(0)e = \pi(I - \lambda S^0\alpha h_1)e = 1-\lambda\pi S^0 h_1.$

Differentiating the equation (8) with respect to $s$ and setting $s = 0$ in the result gives the following

$$EW(I - \lambda S^0\alpha h_1)e = \frac{1}{2}\lambda\pi S^0 h_2 \quad (9)$$

Now, multiplying both sides of (7) from the right with vector $a$ which was introduced in Subsection 2 and is a solution to the set of equations (5). Taking into account (5) we get

$$\varphi(s)\lambda((\pi S^0)e - S^0) = \varphi(s)(\lambda S^0\alpha(1-h\left(s\right) - sI)a + sF(0)a. \quad (10)$$

After differentiating the last equation with respect to $s$ and setting $s = 0$ in the result gives the following

$$-EW\lambda\left(\left(\pi S^0\right)e - S^0\right) = \pi\left(\lambda S^0\alpha h_1 - I\right)a + F(0)a. \quad (11)$$

Now, multiplying (11) by $h_1$ and summing with (9) we get the equation

$EW((I - \lambda S^0\alpha h_1)e - \lambda((\pi S^{0)})e - S^0)h_1) =$

$\pi\frac{\lambda S^0 h_2}{2} + \pi(\lambda S^0\alpha h_1 - I)ah_1 + F(0)ah_1,$ and

$$EW = \frac{1}{1-\lambda\pi S^0 h_1}\left[\frac{\lambda\pi S^0 h_2}{2} + \pi\left(\lambda S^0\alpha h_1 - I\right)ah_1 + F(0)ah_1\right].$$

Thus, the theorem is proved.

Theorem2.Under heavy traffic

$$(1-\rho)EW = \frac{h_2}{2h_1} + \frac{\pi\left(S^0\alpha - (\pi S^0)I\right)ah_1}{\pi S^0} + 0(1).$$

Proof: After we let $\rho \to 1^{-}0$ we get $\lambda \to \frac{1}{\pi S^0 h_1}$ and all the probabilities $F_i(0)$ $i = 0,1,\cdots,K$ tend to zero. Now, our theorem follows directly from (6).

Now, consider asymptotic behavior of the stationary distribution of the process ( W(t), J(t)).

Theorem 3. Under heavy traffic random variables $W(t)$ and $J(t)$ are asymptotically independent and the random variable $(1-\rho)W(t)$ is asymptotically exponential with the mean

$$M = \frac{h_2}{2h_1} + \frac{\pi(S^0\alpha - (\pi S^0)I)ah_1}{\pi S^0}.$$

Proof: After multiplying both sides of (7) from the right by the vector of ones $e$ and since $(S + S^0\alpha)e = 0$, $F(0)e = 1-\rho \equiv \varepsilon$ we get

$$\varphi(s)\left[\lambda S^0\alpha(1-h(s)) - sI\right]e + s\varepsilon = 0. \tag{12}$$

Postmultiplying (7) by the matrix $R$ which was introduced in Section 2 for the case when $Q = S + S^0\alpha$ and noting that for any vector $x = (x_0, x_1, \cdots, x_K)$ we have

$xQR = x - \frac{x_0}{\pi_0}\pi$, yields

$$\varphi(s) = \frac{\varphi_0(s)\pi}{\pi_0} + \frac{1}{\lambda}\varphi(s)\left[\lambda S^0\alpha(1-h(s)) - sI\right]R + \frac{1}{\lambda}sF(0)R$$

(13) Replacing the vector $\varphi(s)$ in the right hand side of (13) with the help of (13)

$$\varphi(s) = \frac{\pi\varphi_0(s)}{\pi_0}\{I + \frac{1}{\lambda}\left[\lambda S^0\alpha(1-h(s)) - sI\right]R\} + sY_0(s), \text{ where} \tag{14}$$

$$Y(s) = \frac{1}{\lambda^2}s\varphi(s)\{[\lambda S^0\alpha\frac{1-h(s)}{s} - I]R\}^2$$
$$+\frac{F(0)R}{\lambda^2}\{\lambda + \frac{1}{\lambda}[\lambda S^0\alpha(1-h(s)) - sI]R\} \tag{15}$$

Substituting (14) into (12) we have

$$\varphi_0(s) = -\pi_0\frac{B(s)}{C(s)}, \tag{16}$$

where $B(s) = \varepsilon s + SY(s)[\lambda S^0 \alpha - sI]e$, $C(s) = C_1(s) + C_2(s)$, and

$C_1(s) = \pi[\lambda S^0 \alpha(1-h(s)) - sI]e$

$C_2(s) = \frac{1}{\lambda}[\lambda S^0 \alpha(1-h(s)) - sI]R[\lambda S^0 \alpha(1-h(s)) - sI]e$.

Now let us replace s by $\varepsilon s$ and let $\lambda$ increase in such a way that $\varepsilon \to 0$. Since $|\varphi_i(\varepsilon s)| \le 1$ then from $1-\rho \equiv \varepsilon$ follows that $Y(s) \to 0$ and $\frac{1-h(\varepsilon s)}{\varepsilon s} \to h_1$,

, $\frac{h(\varepsilon s)-1+h_1 \varepsilon s}{(\varepsilon s)^2} \to \frac{h_2}{2}$. Also $\frac{B(\varepsilon s)}{\varepsilon^2 s} = 1 + sY(\varepsilon s)[\lambda S^0 \alpha \frac{1-h(\varepsilon s)}{\varepsilon s} - I]e \to 1$,

$\frac{C_1(\varepsilon s)}{\varepsilon^2 s} = -1 - s\frac{\lambda \pi S^0 (h(\varepsilon s)-1+h_1 \varepsilon s)}{\varepsilon^2 s} \to -(1+\frac{h_2}{2h_1}s)$ ,

$\frac{C_2(\varepsilon s)}{\varepsilon^2 s} \to s\frac{h_1}{\pi S^0}\pi[S^0\alpha - (\pi S^0)I]R[S^0\alpha - (\pi S^0)I]e$ .

At last, from (3) and (5) follows $R(S^0 - (\pi S^0)e) = -a$, where $a$ is a solution to (5).

Thus, (16) yields that there exists $\lim_{\varepsilon \to 0} \varphi_0(\varepsilon s) = \frac{\pi_0}{1+Ms}$, where

$M = \frac{h_2}{2h_1} + \frac{1}{\pi S^0}\pi(S^0\alpha - (\pi S^0)I)ah_1$ .Finally, from (13) we get

$\lim_{\varepsilon \to 0} \varphi(\varepsilon s) = \lim_{\varepsilon \to 0} \varphi_0(\varepsilon s)\frac{\pi}{\pi_0} = \pi \frac{1}{1+Ms}$, which implies the desired result.

## 5.Conclusion

The rigorous proof of the convergence of the distribution of virtual waiting time is based on the certain "semi-explicit" formulas for the stationary distribution of the virtual waiting time These formulas contain the unique unknown term $Y(s)$, which vanishes under the heavy traffic. This approach could be used for solving other problems.

## References

1. Neuts M.F. Probability distributions of phase type. In Liber Amicorum Professor Emeritus H.Florin . 173-206.Dept.Math., Univ. Louvain, Belgium., 1975.
2. Neuts, M.F. Renewal processes of phase type. Naval Res. Logist. Quart. 25, 445-454.
3. Neuts, M. F. Matrix-geometric solutions is stochastic models: an algorithmic approach. John Hopkins University Press, 1981.

# One New Solution of a Functional Equation

**Diko Souroujon, Teodora Zapryanova**

University of Economics - Varna, 9002 Varna, 77 Knyaz Boris, Bulgaria
teodorazap@abv.bg

**Abstract.** This paper presents one new solution of a functional equation - one problem from group A of the National Student Olympiad of Mathematics, held in Sozopol, 30May – 1 June, 2014. New ideas to the solution are considered. It contains some methodological details.

**Keywords.** NSOM 2014, Functional equations

## 1.Introduction

Mathematicians have been working with functional equations for much longer period of time than the formal discipline has existed. Examples of early functional equations can be traced back as far as the work of fourteenth century mathematician Nicole Oresme who provided an indirect definition of linear functions by means of functional equation. The function $f$ is defined to be linear, if

$\frac{y-x}{z-y} = \frac{f(y)-f(x)}{f(z)-f(y)}$, for all distinct values of $x, y, z$. (1.1)

What makes Oresme's definition a functional equation is that $f$ is treated abstractly and one may pluq any function into tis equation to see whether the equation is satisfied for all possible $x, y$ and $z$. We can compare this with the standart definition of linear function of the form

$f(x) = ax + b$, for some $a, b$. (1.2)

Oresme's equation (1.1) is a functional equation. The definition in (1.2) with $a \neq 0$ is its solution. Note that Oresme's definition does not allow the constant liner function where $a = 0$.

Over the next few hundred years, functional equations were used but no general theory of such equations arose. Notable among mathematicians was Gregory of Saint-Vincent (1584-1667), whose work on the hyperbola made implicit use of the functional equation

$f(xy) = f(x) + f(y)$.

We now recognize this equation as the distinctive functional equation for the family of logarithms. However, the theoretical work which links this functional equation to the family of logarithms had wait for the work of Augustin-Louis Cauchy.

Although Nicole Oresme's definition of linearity can be interpreted as an early example of a functional equation, it does not represent a start in point for the theory of functional equations. The subject of functional equations is more properly dated from the

work of A. L. Cauchy born in Paris, France. In 1821 it was proved by Cauchy that the only continuous solution of the functional equation

$$f(x+y) = f(x) + f(y) \tag{1.3}$$

from $\mathbb{R}$ to $\mathbb{R}$ are those of the form $f(x) = kx$ for some real number *k*. If the function *f* with $D_f = \mathbb{Q}$ satisfies only the additive condition (1.3) then the solution is again of the form $f(x) = kx$ for some real number *k* . In 1905 G. Hamel proved that there are non-continuous solutions of the Cauchy functional equation (1.3) using Hamel bases. Every non-continuous solution is necessarily non-measurable with respect to the Lebesgue measure. The fifth of Hilbert's problems is generalization of this equation.

**Definition** We say that the function $f$ is additive if satisfies (1.3).

It is known that if $f: \mathbb{R} \to \mathbb{R}$ is additive function ( $D_f = \mathbb{R}$) and in addition satisfies one of following condition;

1) $f(x)$ is continuous at fixed point
2) $f(x)$ is monotonous in the given fixed interval
3) $f(x)$ is bounded in a neighborhood of given point

then $f(x) = kx$ for some real constant *k*.

Cauchy's functional equation (1.3) is a tool in the solution of other functional equations.

**Example 1** Let $f: \mathbb{R} \to \mathbb{R}$ be strongly increasing (decreasing) function defined on $\mathbb{R}$ and satisfies the condition

$$f(x+y) = f(x)f(y) \quad \forall x, y \in \mathbb{R}.$$

Then $f(x) = a^x$, for $a \in (0,1)$ $(a \in (1,\infty))$ when $f$ is decreasing (increasing).

Putting *y=x* we show that $f(x) \geq 0 \ \ \forall x \in \mathbb{R}$, then we show that it is impossible $f(x) = 0$, i.e. $f(x) > 0 \ \forall x \in \mathbb{R}$, then putting $g(x) = \ln f(x)$ it is obvious that $g(x)$ is additive function.

**Example 2** Let $f$ be monotonous function, defined on $(0,\infty)$ and satisfies condition

$$f(xy) = f(x) + f(y) \ \ \forall x, y \in \mathbb{R}^+.$$

Then

$$f(x) = a \ln x \text{ for arbitrary const. } a \in \mathbb{R}.$$

The proof is obtained putting $x = e^t$, $y = e^s$ and $g(t) = f(e^t)$. Then obviously $g(t)$ is additive function.

**Example 3** Let $f: \mathbb{R}^+ \to \mathbb{R}^+$ be continuous function, defined on $(0,\infty)$ and satisfies the condition

$$f((xy)^a) = \big(f(x)f(y)\big)^a \ \ \forall x, y \in \mathbb{R}^+ \text{ for some real number } a > 0.$$

Then $f(x) = x^c$ if $a \neq 1/2$; $f(x) = Ax^c$ if $a = 1/2$, $c$ and $A>0$ are real constants.

**Proof.** If we put $x = e^t$, $y = e^s$ we obtain $f\left(e^{a(t+s)}\right) = \left(f(e^t)f(e^s)\right)^a$,

$\ln f\left(e^{a(t+s)}\right) = a(\ln f(e^t) + \ln f(e^s))$. We put $g(t) = \ln f(e^t)$. Then $g: \mathbb{R} \to \mathbb{R}$ and

$$g\big(a(t+s)\big) = a\big(g(t) + g(s)\big). \tag{1.4}$$

If $t = s = 0$, then $g(0) = 2ag(0)$.

Case 1. $a \neq 1/2$.

In this case $g(0) = 0$. From (1.4) with $s$=0 we have $g(at) = ag(t)$. From (1.4) we get

$$g\big(a(t+s)\big) = ag(t) + ag(s) = g(at) + g(as).$$

If we replace $at$ and $as$ by $t$ and $s$, then $g(t+s) = g(t) + g(s)$, i.e. $g$ is an additive function. Since $g$ is continuous $g(t) = \ln f(e^t) = ct$ for some const. c. Thus $f(e^t) = e^{ct} = (e^t)^c$ and putting $e^t = x$ we get $f(x) = x^c$, where c is an arbitrary constant. We verify that this function satisfies the desired equation.

Case 2. $a = 1/2$,

i.e. $f$ satisfies the equation $f\left(\sqrt{xy}\right) = \sqrt{f(x)f(y)}$ $\forall x, y \in \mathbb{R}^+$. As in Case 1, $x = e^t$, $y = e^s$, $f\left(e^{\frac{t+s}{2}}\right) = \sqrt{f(e^t)f(e^s)}$, $\ln f\left(e^{\frac{t+s}{2}}\right) = \frac{1}{2}(\ln f(e^t)) + \ln f(e^s)$ and hence

$$g\left(\frac{t+s}{2}\right) = \frac{g(t)+g(s)}{2} \quad \forall t, s \in \mathbb{R}. \tag{1.5}$$

The function $\varphi(t) = g(t) - g(0)$ also satisfies (1.5), i.e.

$$\varphi\left(\frac{t+s}{2}\right) = \frac{\varphi(t)+\varphi(s)}{2} \quad \forall t, s \in \mathbb{R} \text{ and } \varphi(0) = 0.$$

For $s$=0 we have $\varphi\left(\frac{t}{2}\right) = \frac{\varphi(t)}{2}$, i.e. $\varphi(t) = 2\varphi\left(\frac{t}{2}\right)$ and hence

$$\varphi(t+s) = 2\varphi\left(\frac{t+s}{2}\right) = \varphi(t) + \varphi(s) \quad \forall t, s \in \mathbb{R},$$

i.e. $\varphi(t)$ is additive function. Since $\varphi(t) = \ln f(e^t) - \ln f(e^0)$ is a continuous function

$$\varphi(t) = \ln f(e^t) - \ln f(1) = ct$$

Hence $\ln f(e^t) = \ln f(1) + ct$, $f(e^t) = e^{\ln f(1)+ct} = f(1)e^{ct} = f(1)(e^t)^c$.

Thus $f(x) = Ax^c$ $\forall x \in \mathbb{R}$, where $A = f(1) > 0$ and $c$ are arbitrary constants. This function satisfies the desired equation.

**Example 4** Let $f: \mathbb{R} \to \mathbb{R}$ be continuous function, defined on $(-\infty, \infty)$ and satisfies the condition

a) $xyf(x+y) = (x+y)\big(x\,f(y) + y\,f(x)\big)$ $\forall x, y \in \mathbb{R}$;

b)$(x+y)f(x)f(y) = f(x+y)\big(x\,f(y) + y\,f(x)\big)\ \forall x, y \in \mathbb{R}$;

c) $(x+y)f(x)f(y) = xy\,f(x+y)\ \forall x, y \in \mathbb{R}$.

Then a) $f(x) = c\,x^2$; b)$f(x) = c$; c)$f(x) = x\,e^{cx} = x\,a^x$ for some const. c and $a = e^c$.

The proof follows from the fact, that the function

a)$\frac{f(x)}{x}$; b) $\frac{x}{f(x)}$; c)$\ln\frac{f(x)}{x}$ is additive.

There is no theory attached to functional equations. Each such equation requires different techniques to solve it. Functional equations are often the most difficult problems to be found on mathematics competitions because they require a minimal amount of background theory and a maximal amount of ingenuity. The great advantage of a problem involving functional equation is that student at all levels can understand and play with.

In the oldest mathematical competition for university students in European Union – Vojtěch Jarnik International Mathematical Competition, Ostrava, 13th -14th April 1993, we found the following closely related to Cauchy equation

**Problem** *Prove that if* $f:\mathbb{R} \to \mathbb{R}$ *fulfills the inequalities*

$$f(x) \le x, \quad f(x+y) \le f(x) + f(y)$$

*for all* $x, y \in \mathbb{R}$ *then* $f(x) = x$ *for all* $x \in \mathbb{R}$.

For more details see http://vjimc.osu.cz/

The paper is dedicated to one new solution of a problem from the National Student Olympiad of Mathematics, held in Sozopol, 30 May – 1 June, 2014. Detailed information is published at http://nsom2015.heliohost.org/nsom15.html, where we can find the author's solution of the problem. Here we present second solution, different from author's of the following

**Problem** *For* $p \ge 0$, *let* $f_p(x) = 2014x - p\sin x$. *Find all numbers p, such that* $f_p\left(f_p(x)\right) = x$ *implies* $f_p(x) = x$.

This problem can be generalized using an arbitrary positive real number A instead the number 2014. The generalization is

**Problem** *For* $p \ge 0$ *let* $f_p(x) = Ax - p\sin x$, *where A>0 is an arbitrary real constant. Find all numbers p, such that* $f_p\left(f_p(x)\right) = x$ *implies* $f_p(x) = x$.

## 2.Solution

Case 1. Let $0 \le p < A+1$ and $f_p\left(f_p(x)\right) = x$. We will employ the Lagrange's theorem of finite growth

$$0 = \left(f_p\left(f_p(x)\right) - f_p(x)\right) + \left(f_p(x) - x\right)$$

$$= \left(f_p(x) - x\right)f_p'(\theta) + \left(f_p(x) - x\right) = \left(f_p'(\theta) + 1\right)\left(f_p(x) - x\right) \qquad (2.1)$$

for some $\theta \in (-\infty, \infty)$.

But

$$f_p'(\theta) = A - p\cos\theta \geq A - p > A - (A+1) = -1$$

Hence

$$f_p'(\theta) + 1 > 0.$$

Thus from (2.1) we get

$$f_p(x) - x = 0, \ i.e. \ f_p(x) = x.$$

Case 2. Let $p = A + 1$ *and* $f_p\left(f_p(x)\right) = x$ ,

where $f_p\left(f_p(x)\right) = f(x) = Ax - (A+1)\sin x$. Putting $f(x) = y$ we obtain

$$0 = f_p\left(f_p(x)\right) - f_p(x) + f_p(x) - x = f(y) - f(x) + y - x =$$

$$(Ay - (A+1)\sin y) - (Ax - (A+1)\sin x) + y - x =$$

$$(A+1)y - (A+1)\sin y - (A+1)x + (A+1)\sin x = 0$$

(due to $A+1>0$)

$$\Leftrightarrow \sin y - \sin x = y - x \quad \Leftrightarrow \quad 2\sin\frac{y-x}{2}\cos\frac{x+y}{2} = y - x \quad \Leftrightarrow$$

$$\sin\frac{y-x}{2}\cos\frac{x+y}{2} = \frac{y-x}{2} \qquad (2.2)$$

But $\left|\sin\frac{y-x}{2}\cos\frac{x+y}{2}\right| \leq \left|\sin\frac{y-x}{2}\right| \leq \left|\frac{y-x}{2}\right|$ and since $\left|\cos\frac{x+y}{2}\right| \leq 1$, the equality (2.2) is possible iff

$$\frac{y-x}{2} = 0 \ \ i.e. \ \ y = x \ i.e. \ f(x) = x.$$

Case 3. $p > A + 1$.

In this case we show that there exists $x_0 \neq 0$ such that $f_p(x_0) = -x_0$.

Indeed $f(x) = -x \ \Leftrightarrow \ (A+1)x - p\sin x = 0 \ \Leftrightarrow \ \frac{\sin x}{x} = \frac{A+1}{p} \in (0,1)$.

We have $\left(\frac{\sin x}{x}\right)' = \frac{x\cos x - \sin x}{x^2} = \frac{\cos x}{x^2}(x - \tan x) < 0 \quad \forall x \in (0, \pi)$.

As $\tan x > x$, $\cos x > 0$ for $x \in (0, \frac{\pi}{2})$ we have $\left(\frac{\sin x}{x}\right)' < 0$ for $x \in (0, \frac{\pi}{2})$; if $x \in [\frac{\pi}{2}, \pi)$, then $x\cos x \leq 0$, $\sin x > 0$ and hence $\left(\frac{\sin x}{x}\right)' < 0$ for $x \in [\frac{\pi}{2}, \pi)$, Thus

$$\left(\frac{\sin x}{x}\right)' < 0 \quad \forall x \in (0, \pi).$$

Therefore the function $\frac{\sin x}{x}$ is decreasing for $x \in (0, \pi)$.

Since

$$\lim_{x \to 0+} \frac{\sin x}{x} = 1, \quad \frac{\sin \pi}{\pi} = 0 \text{ and } \frac{\sin x}{x} \text{ is continuous in } [0, \pi]$$

(at $x$=0 we define by continuity $\frac{\sin x}{x} = 1$ ), there exists $x_0 \in (0, \pi)$:

$$\frac{\sin x_0}{x_0} = \frac{A+1}{p} \quad \Leftrightarrow \quad f_p(x_0) = -x_0.$$

Thus

$$f_p\left(f_p(x_0)\right) = f_p(-x_0) = -f_p(x_0) = x_0$$

(since $f_p(.)$ is odd function), i.e. there exists $x_0 \in (0, \pi)$:

$$f_p\left(f_p(x_0)\right) = x_0,$$

but

$$f_p(x_0) = -x_0 \neq x_0, \quad \text{as } x_0 \neq 0.$$

Thus we get that only when

$$0 \leq p \leq A+1 \quad f_p\left(f_p(x)\right) = x \quad \text{implies} \quad f_p(x) = x.$$

## References

1. Aczel, J. Lectures on Functional Equations and Their Applications. New York: Academic Press, 1966.
2. Cauchy, A.L. Cours d'Analyse de'lEcole Royale Polytechnique. Chez Debure freres, 1821.
3. Hamel, G. "Eine Basis aller zahlen und die unstetigen Lösung der Funktional-gleichung $f(x+y) = f(x) + f(y)$". Math. Ann. 60, 459-462, 1905.

# Innovative E-System for Strategic Analysis of New Technology Ventures

**Sia Tsolova[1], Petko Ruskov[2]**

[1]Sofia University "St. Kliment Ohridski", 1164 Sofia, Bulgaria
s.valentinova@gmail.com
[2]Sofia University "St. Kliment Ohridski", 1164 Sofia, Bulgaria
petkor@fmi.uni-sofia.bg

**Abstract.** The aim of this report is to present the research and the current results from a research on strategy management for technology new ventures, by proposing an innovative e-system for strategic analysis for technology new ventures. The innovative e-system is designed based on investigation amongst 121 entrepreneurs in the technology sphere and development of innovative tool for strategic management, specifically designed for technology startup companies strategic analysis and implementation. The system includes selection on the most useful for technology new ventures strategic analysis tools and the developed innovative strategic analysis tool for technology new ventures, including guidelines and developed process of use of the innovative tool. The presented e-system can be used for further incorporation into an overall strategic management system supporting all stages of the strategic management process for technology new ventures.

**Keywords:**strategy; analysis; innovation; e-system; technology new ventures

## 1.Introduction

The fast development of technology and the high amount of startup companies in the technology sphere (technology new ventures) is bringing up the question of creation of strategic modeling tools, designed for technology new ventures, supporting each of the strategic management stages for those type of companies – strategic analysis, strategic formulation, strategic implementation, strategic execution and strategic control and evaluation.

According a research of Bloomberg 80% of the startup companies are closing their activities in 18 months. According another research of National Statistics Institute of Bulgaria from 2012, only 8.1% of the startup companies are still functioning after 5 years of their development. All these statistics are showing an emerging need for development tools and systems, corresponding to the specific needs of the startup companies. Development of such tool is also the focus of this report.

The presented in this article and developed by the author innovative canvas for strategic analysis and identification of the basic competitive advantages of a technology new ventures (technology startup company) is designed to support the process of strategic analysis and modeling, by presenting an unified instrument, consisting in its structure the most important elements from the researched classical instruments for strategic analysis, which elements contribute in highest degree to the gathering of key information for the strategy modeling process through identifying and analysis of the characteristics and competitive advantage of the company.

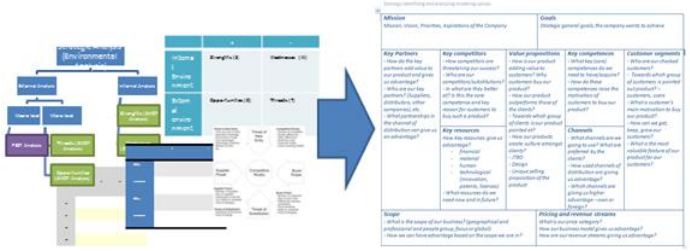


Figure 1: Development of the unified Strategy analysing and identifying modeling canvas

The place of application of Strategy analysing and identifying modeling canvas is exactly before the start of the strategic modeling process for technology new ventures. The developed instrument can be applied as a stand-alone instrument, or together with the classical instruments for strategic management, as its focus is the strategic analysis for the following strategy modeling process.

The process of development of the Strategy analysing and identifying modeling canvas includes: (1) extraction and grouping of the key for the technology new ventures elements from the classical strategic analysis and modeling tools; (2) ordering of the elements of the Strategy analysing and identifying modeling canvas; (3) defining of the key questions for each of the categories of the canvas usage and (4) defining of the logical sequence of work with the different categories of the canvas and work with its elements.

## 2. Elements of the Strategy analysing and identifying modeling canvas

The development of the Strategy analysing and identifying modeling canvas begins with defining of the key elements from the classical strategic analysis tools and their grouping in thematical groups.

As a basis for the development of the Strategy analysing and identifying modeling canvas, the type of instrument which was chosen is "canvas" (Strategy analysing and identifying modeling canvas is similar to the instrument Business Model Canvas, which lies closest to the idea of the developed instrument). The Business Model Canvas, also, is the only instrument from all strategic analysis tools, developed specifically for startup companies, and technology new ventures as a part, and due to this reason all its elements have been considered with the needs of the startup companies in the development of their business models. At the strategic modeling process, the elements, which are included in the Strategy analysing and identifying modeling canvas are different, due to the different focus (i.e. strategy analysis and modeling), but also there are common elements, between the two instruments. Both, the process of analysis and strategies modeling, as well as the most important elements from the different strategic management tools are researched for the defining of all elements of the presented instrument.

The development of each strategy starts with definition of: (1) Mission; (2) Vision and the extracted from them (3) Strategic goals, which the companies defined for reaching in its existence.

An important note is to be made for close link between the product modeling in this early stage of the company's existence and the strategy modeling. According the theoretical basics of the methodology, developed by Prof. Steve Blank and his team, while in the past the developed companies were using the Product Development Model as standard methodology for product development, nowadays, a better for the startup companies methodology has emerged – The Customer Development Model.

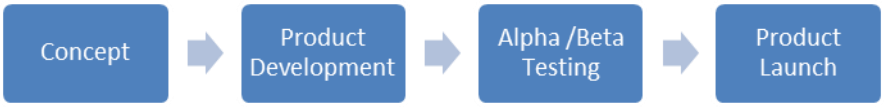



Figure 2: Product Development Model

Using Product Development Model doesn't use feedback from customers, while at Customer Development Model, the product development is implemented in close connection with the clients, together with clients' identification, which leads to developing more successful products of higher quality and with less risks for the startup companies

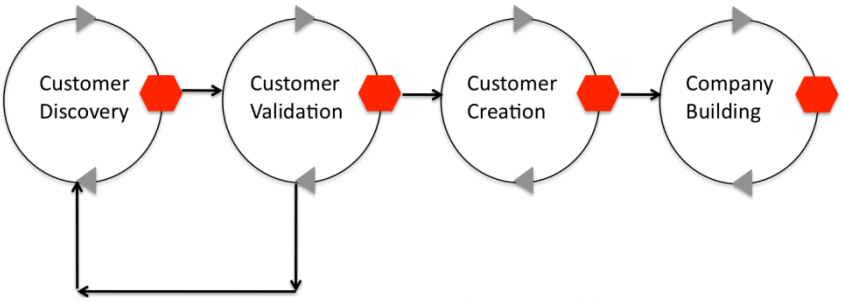



Figure 3: Customer Development Model

This specifics in the methodologies of early product modeling, together with the "search mode" of work of the startup companies, are also connected with the existence of several categories of elements, included in the developed Strategy analysing and identifying modeling canvas, which are also available at the business model development, connected with the client segment and the value proposition of the product.

The following key elements of the developed Strategy analysing and identifying modeling canvas were identified, based on the researched elements of strategic management and modeling:

- Category „Mission" – extracted from the overall process of strategic management
- Category „Goals" – extracted from the overall process of strategic management and included in the instrument GAP analysis
- Category „Key competence/skills" – extracted from the instruments SWOT analysis and Core Competences analysis
- Category „Key resources" - extracted from the instruments GAP analysis, Business Model Canvas, SWOT analysis, TOWS matrix
- Category „Key competitors" - extracted from the instruments Niche/GAP analysis, SWOT analysis and Porter's Five Forces analysis
- Category „Key partners" - extracted from the instruments Porter's Five Forces analysis, SWOT analysis, TOWS matrix and included in Business Model Canvas
- Category „Clients" – extracted from the instruments Business Model Canvas, Porter's five forces analysis, included in Balanced Scorecard and Customer Development Model
- Category „Product value proposition" – extracted from the instruments: Unique Selling Proposition analysis, Business Model Canvas, included in Customer Development Model
- Category „Pricing and revenue streams" - extracted from the instruments Business Model Canvas and defining in large amount the business model of the startup company
- Category „Market scope" – extracted from the instruments Niche/GAP analysis, Business Model Canvas and included in Customer Development Model

- Category „Channels of distribution“ – extracted from the instruments Business Model Canvas, included in Customer Development Model

The defined categories are aiming support of the process of strategic analysis and identification of the key characteristics and competitive advantages of the startup company in the process of its strategy modeling, which on the other hand, brings up the need for development of questions for each of the defined categories, explaining and supporting the usage of the developed instrument, as well as a sequence of working with the categories, which will be the process of usage of the developed Strategy analysing and identifying modeling canvas.

The division of the chosen categories on the canvas, together with the developed for each one of the categories key supporting questions, are presented in the following table.

Strategy identifying and analyzing modeling canvas

<table>
<tr><td colspan="3">Mission<br>Describe the Mission, Vision, Values, Priorities and Aspirations of the Company</td><td colspan="3">Goals<br>What are the Strategic general goals, the company wants to achieve</td></tr>
<tr><td rowspan="2">Key Resources<br>How Key resources give us advantage?<br>- financial<br>- material<br>- human<br>- technological (innovation, patents, licenses)<br>- What are the techno-logical capabilities of our company?<br>- What resources do we need now and in future?</td><td>Key Partners<br>- How do the key partners add value to our product and gives us advantage?<br>- Who are our key partners? (Suppliers, distributors, other companies), etc.<br>- What partnerships in the channel of distribution can give us an advantage?</td><td rowspan="2">Value Propositions<br>- How is our product adding value to customers? Why customers buy our product?<br>- How our product outperforms those of the clients?<br>- Towards which group of clients is our product pointed at?<br>- How our products create culture amongst clients?<br>- JTBD<br>- Design<br>- Unique selling proposition of the product<br>- What level of quality is enough to gain competitive advantage?</td><td>Scope<br>- What is the scope of our business? (geographical and professional and people group, focus or global)<br>- How we can have advantage based on the scope we are in?</td><td rowspan="2">Customer segments<br>- Who are our checked customers?<br>- What is the scope of the market?<br>- What is the maturity of the market?<br>- Towards which group of customers is pointed out product? – customers, users<br>- What is customer's main motivation to buy our product?<br>- How can we get, keep, grow our customers?<br>- What is the most valuable feature of our product for our customers?</td></tr>
<tr><td>Key Competitors<br>- How competitors are threatening our success?<br>- Who are our competitors/substitutions?<br>- In what are they better at? Is this the core competence and key reason for customers to buy such a product?</td><td>Channels<br>- What channels are we going to use? What are preferred by the clients?<br>- How used channels of distribution are giving us advantage?<br>- Which channels are giving us higher advantage – own or foreign?</td></tr>
<tr><td colspan="3">Key competences<br>- What key (core) competences do we need to have/acquire?<br>- How do these competences raise the motivation of customers to buy our product?</td><td colspan="3">Pricing and revenue streams<br>What is our price category?<br>How does our business model give us advantage?<br>How are our revenue streams giving us advantage?</td></tr>
</table>

Table 1: Strategy analysing and identifying modeling canvas

Based on the developed canvas a research amongst 121 entrepreneurs in the technology sphere was implemented, which confirmed the levels of importance of the included in the developed instrument elements.

## 3. Results from the conducted research on Strategy analysing and identifying modeling canvas

The research implemented amongst 121 participants, includes research on the level of usage of classical strategic management instruments and the level of importance of the elements, included in the developed Strategy analysing and identifying modeling canvas.

The results from the research are showing the high levels of importance of the included in the instrument elements, confirming the choice of the elements and categories in the Strategy analysing and identifying modeling canvas.

The research was implemented on the basis of questionnaire amongst 121 entrepreneurs in the technological sphere (63% men and 37% women).

The first part of the research aimed identification of the level of usage of classical strategic management tools. The results from the implemented research are:

| What methods and instruments in the market research process are you using for the strategy modeling of your business? | Yes, I use it | No, I don't use it | I am not acquainted with it |
|---|---|---|---|
| SWOT analysis | 70% | 7% | 22% |
| PEST analysis | 30% | 10% | 60% |
| GAP analysis | 32% | 12% | 55% |
| Unique Selling Proposition analysis | 26% | 9% | 64% |
| Business Model Canvas | 54% | 3% | 43% |
| Balanced Scorecard Methodology | 23% | 12% | 64% |

Table 2: Levels of usage of Classical Strategic Management Tools in the Strategic Management process

In direct connection with the upper results is also the asked question on level of education in the sphere of entrepreneurship, business and management amongst the participants. The results are as follows:

| Have you been studying entrepreneurship and management? | Results |
|---|---|
| Yes, I have. | 65% |
| No, I haven't. | 20% |
| Self-learning | 15% |

Table 3: Levels of usage of Classical Strategic Management Tools in the Strategic Management process

An interesting fact in the research is the close result between the participants of participants in the research, who are not using any tools for strategic analysis and the percentage of participants, who haven't been studying entrepreneurship and management before.

The participants in the research were asked about the level of development of their startup ideas. The results according the stage of business development amongst the participants are:

| Stage of startup development | Results |
|---|---|
| Atstage "idea development" | 69% |
| Atstage "business starting" | 12% |
| At stage "new working business idea" | 19% |

Table 4: Stage of startup development amongst participants in the research

Totally over 81% of the participants from the startup technology sphere are at stage "idea development" and "business starting", where the stage of "idea development" includes the period before having first sales amongst the clients. Exactly this is the period of strategy, business and product modeling and the period in which future plans of realisation aiming at the successful product launch are made. For these reasons, this is also the key period, in which the foundations of a future successful development of the technology startups is made and this is the period for initial application of the presented in this report instrument and the process of its application.

The implemented research on the importance of the elements included in the developed Strategy analysing and identifying modeling canvas are presented in the following table:

| Categories in the developed instrument | High importance | Average importance | Total: High and Average Importance | Low importance |
|---|---|---|---|---|
| Clients | 93% | 7% | 99% | 1% |
| Product | 88% | 11% | 99% | 1% |
| Strategic goals | 80% | 19% | 99% | 1% |
| Mission | 63% | 30% | 93% | 7% |
| Competitors | 61% | 36% | 98% | 2% |
| Key resources | 60% | 33% | 93% | 7% |
| Market scope | 60% | 37% | 98% | 2% |
| Pricing and revenue streams | 60% | 37% | 97% | 3% |
| Key competence/skills | 50% | 45% | 94% | 6% |
| Key partners | 48% | 43% | 91% | 9% |
| Channels | 46% | 47% | 93% | 7% |

Table 5: Results from research on the level of importance of the included in the developed instrument categories

The implemented research on the level of importance of the chosen categories for the Strategy analysing and identifying modeling canvas proves the suitable choice of the categories n the developed instrument and supports also the following the next stage in the instruments development, which is the process of defining the sequence of work and usage of the Strategy analysing and identifying modeling canvas.

## 4. Process of workflow with the developed Strategy analysing and identifying modeling canvas

The next stage in the development of the Strategy analysing and identifying modeling canvas is the defining of the workflow and sequence of work with the developed tool.

The development of the process of work with this instrument includes defining of the sequence of filling in information in the different categories and development of overall process of workflow with the instrument.

The process of work with the developed Strategy analysing and identifying modeling canvas includes the following three stages: (A) Defining of the scope of competition; (B) Defining of competitive advantage; (C) Defining of key elements from the strategic plan of actions

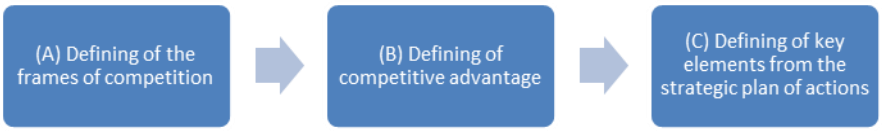



Figure 4: Process of work with the developed Strategy analysing and identifying modeling canvas

Each of these stages includes work on a particular set of categories of the developed Strategy analysing and identifying modeling canvas.

Stage (A) Defining of the scope of competition includes work on the categories in sequence as follows: 1) Mission; 2) Goals; 3) Product value proposition; 4) Clients; 5) Market scope. The sequence of work with all categories is recommended, but not obligatory.

Stage (B) Defining of competitive advantage includes work on the following categories: 6) Key competence/skills; 7) Key competitors; 8) Key partners; 9) Types of pricing and revenue streams.

Stage (C) Defining of key elements from the strategic plan of actions includes the following categories: 10) Key resources; 11) Channels of distribution (incl. advertising activity).

The following scheme presents visual representation of the sequence of work with the developed instrument.

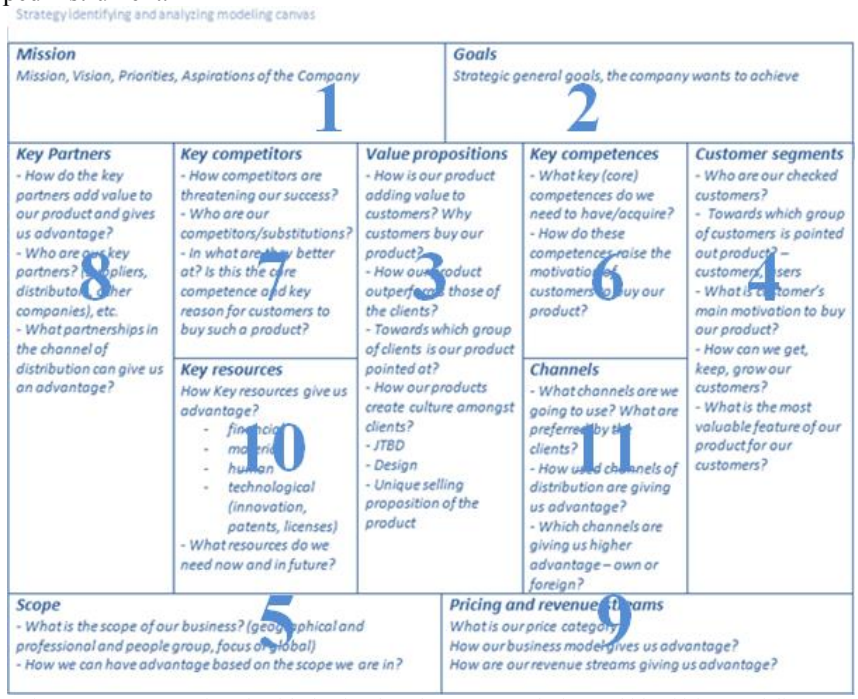



Figure 5: Sequence of work with the developed Strategy analysing and identifying modeling canvas

There are similar processes of work with almost all instruments for strategic management, where some parts of the sequence of work is obligatory and other is recommended only. The proposed process has recommendatory character, as the sequence of filling in the information in the canvas does not define the content of the canvas, but the defined sequence would enormously support and enhance the logical work with the instrument and would increase the level of its efficiency.

Additional testing of the instrument and the process of its work was implemented amongst 16 technology startup companies, which results showed high level of acceptance of the instrument – 81% of the participants of the approbation have evaluated the results from using of the instrument and its algorithm as highly accurate and 94% of the participants have evaluated the instrument and the process of its work as highly applicable in technology entrepreneurship strategic analysis. Additionally 72% of the participants in the approbation have evaluated the instrument and te developed process as easy to use in high degree and 18% have evaluated it as easy to use in highest degree.

## 5. Conclusions

The presented Strategy analysing and identifying modeling canvas for technoe instlogy new ventures includes all basic elements in the formulation of strategies, where the answers of these questions in the different categories are aiming support of the successful strategies modeling process for technology new ventures. The categories, in which the factors of success and the questions for their identification are based on extraction of the most important elements of the number of instruments of strategic analysis, the types of strategies and their characterising elements, as well as the specifics of activities' focus at the technology startup companies. The Strategy analysing and identifying modeling canvas corresponds to all upper mentioned specifics of strategic analysis and strategy modeling for technology new ventures.

The developed Strategy analysing and identifying modeling canvas can be used also over previously developed typological strategies, supporting in this way the development of strategies for a technology startup company and can be used as a basis for further development as part of E-Systems supporting the process of strategic analysis and strategic modeling for technology new ventures.

## References

1. Osterwalder, I. Peigner (2010), Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers, Wiley, John & Sons Inc.
2. A.Haberberg, A.Rieple (2008), Strategic Management: Theory and Application, Oxford University Press.
3. B.Yankov (2013), Research of Success Factors for Start-up Companies, Doctoral Conference in Mathematics, Informatics and Education 2013, ISBN 978-954-07-3596-2, pp. 119-124.
4. Harvard Business Essentials (2005), Harvard Business Essentials: Strategy: Create and Implement the Best Strategy for Your Business, Harvard Business Press.
5. Harvard Business Press (2005), Strategy: Create and Implement the Best Strategy for Your Business, Harvard Business Press.
6. Harvard Business Press (2006), Harvard Business Essentials: Strategic management, Harvard Business Press.
7. J. Thompson (2005), Strategic management: Awareness and change, Cengage Learning EMEA.
8. M. Hitt, R. Ireland (2008), Competing For Advantage, Cengage Learning.
9. M. Panova, T. Panov (2008), Strategic management.
10. S.Tsolova (2014), Results from Research on Algorithm for Strategy Modeling Implemented by e-Governance System, VI International Scientific Conference "E-Governance" 2014, ISSN 1313-8774, pp. 224-232.
11. Steve Blank, Bob Dorf (2012), Startup Owners Manual: The Step by Step Guide to Building a Great Company.
12. Steven Gary Blank (2006), The Four Steps to The Epiphany – Successful Strategies to Products that Win.

# Teaching Self-Service BI in a Master Program of Business Administration

**Diana Tisheva**

Sofia University "St. Kliment Ohridski"
Faculty of Economics and Business Administration
dianaht@feb.uni-sofia.bg

**Abstract.** During the last decade almost all of the digital companies have included Business Intelligence (BI) tools in the stack of their corporate data management solutions. The ability to apply BI as a core technology to support their operational and strategic management practices has allowed them to react promptly and adequately to the dynamic changes in the business environment and to leverage it as an additional competitive advantage. The most recent move is towards utilization of the Self-Service BI as one of the most important tools to support the decisions of the chief executives and the knowledge workers. Thus it was supposed that the inclusion of this technology in the curriculum for a Master Program in Business Administration will make it more attractive and oriented towards the needs of the business. This paper presents the approach used to deliver the subject matter and a study of the student's feedback.

**Keywords:** Self-Service BI, Business Intelligence, Power BI.

## 1. Introduction

Most of the MBA programs include a course in Information Systems and Technologies, under this or some other name, which delivers to the students the basic knowledge for the broad area of management information systems. The topics in such a course are more or less standard among the curriculum in the different universities. To name a few, those are topics discussing the digital company, its specific organization and processes, the different information systems that support the business activities, and also their core functional elements – the databases. Although a central theme in the subject matter are the information technologies, the content is actually cross-disciplinary and appropriate for students with different background from their bachelor studies. The fact that some of the topics in the theoretical part of the course have a strong business administration outline is well appreciated by all the students, even the ones with an IT background. However the task to make the laboratory classes interesting and compelling for everyone is a challenging one.

Half of the credits on this course at the Faculty of Economics and Business Administration at Sofia University "St. Kliment Ohridski" (for short FEBA) come from the laboratory classes. The classes are led in computer labs and the students are separated into groups of around 20 people. Most of the courses with subject matter in the information technologies, which are delivered in a non-technical specialty have an audience with a very different background. Thus in some of them the lab classes are scheduled to deliver just a basic knowledge for working with word processing, spreadsheet applications, databases and Internet. However these topics are usually part of some high school or bachelor subjects and thus seem too basic for a master course. Moreover the universities have to get ready to meet in their master programs the so called Generation Z – those are the people born after the mid of 1990s. This generation has grown up with the variety of technologies all around and even

the students with a non-technical bachelor degree possess impressive technical skills and experience. Thus the decision to teach self-service BI by including it as a topic in the lectures and as a main subject matter for the laboratory classes seemed to be a way to simultaneously achieve several goals – deliver more advanced knowledge in the areas of spreadsheet processing and databases, show the potential business impact of the technology and motivate the students to participate actively in the classes by blending their IT and business skills.

## 2. Literature overview

In (Mundy J., 2011) we see an extensive usage of the term DW/BI. These are actually two terms used together – data warehouse and business intelligence. Thus the DW/BI term refers a technology stack that covers everything from esoteric source system knowledge to user interface design and best-practice BI applications (Mundy J., 2011, p.24). More specifically the data warehouse should be the foundation for effective business intelligence (Kimball R., 2010, p.84). Of course in the early days of the data warehousing technology in the late 1980s there were only some ancestors of the BI tools like we know them today. These are the Executive Information Systems (EIS) which have provided key information on the desktops of the executives (Rasmussen, 2002, p.99) and the Decision Support Systems (DSS) which were a broad class of applications classified by (Alter S. L., 1980) in the range between the two extremes – data-oriented DSS that provide data access and simple possibilities for analysis, and model-oriented DSS that apply comprehensive models on the data. Although some of these systems were using really advanced theoretical models, highly tuned for the specific practical needs, at that time, they had a crucial drawback - they were completely independent of the data warehouse. Thus Ralph Kimball has actively promoted the need for restructuring the DSS architecture by using a data warehouse as the single source of information and respectively using the DSS as a source for the EIS. Today many practitioners consider data warehousing to refer to populating the warehouse with data, and business intelligence to refer to using the warehouse data (Ponniah P., 2010, p.xxvii). The terms DSS and EIS are almost not used nowadays and instead BI is the term accepted to denote the array of analytical applications. For the year 2009 the area of Business Intelligence was among the top ten strategic technologies according to Gartner (Gartner, Inc., 2008). During the recent years the functionalities of the different BI toolsets have been constantly extended and the technology has been massively applied in the practice to boost companies' business performance. Respectively there is an increased search for employees that possess MBA or doctoral degrees and relevant analytical skills.

We have to stress out here what is the difference between business intelligence and self-service business intelligence. According to Gartner "Business intelligence is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance". Nowadays business intelligence is not considered anymore as an expensive and exotic approach for data analysis. Many companies have well established platforms and processes to secure the delivery of all related services for data integration, data quality, master data management, etc. The big organizations that have such an extensive IT platform based on a data warehouse are practicing the so called corporate BI. Of course many of the small and medium enterprises do not generate large volumes of data, maintain more simple IT platforms and have smaller budgets for data management and analysis. Due to the new concept of self-service BI these companies could also benefit from the BI if it is applied over a proper data set. Self-service BI is a form of business intelligence in which the end-users can independently generate their own reports, run their own queries, and conduct their own analyses, without the need to engage the IT department (Alexander M., 2014, p.1). This new

type of BI should be considered also by the big organizations as a very important part of the company's endeavors to analyze successfully the variety of internal and external data. The highly dynamic business environment and the related constant changes in the data often cause the corporate business intelligence to fail to deliver its analyses on time or on the basis of the newest available data. This happens because there are usually many steps in the respective data governance methodology and it takes time to complete the process. In such cases an analysis with the self-service BI approach might help. However, whatever the organizational context – a small, medium or big company, in order for self-service BI to occur and bring value, the data architecture needs to provide the 5C's of information – consistent, comprehensive, conformed, clean and current (Sherman R., 2014, p.135).

## 3. Teaching the Self-Service BI topics at FEBA

The common architecture of a corporate BI platform is presented on figure 1.

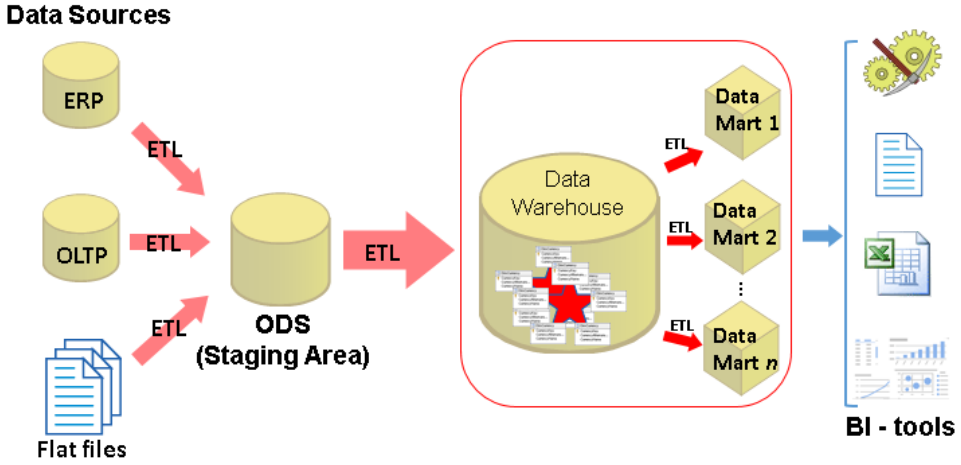



Figure 1. The common architecture of a corporate BI platform.

The implementation and the support of such a complex platform requires extensive resources (technical, labor, financial, time) and is putting a lot of challenges even in front of the big organizations. (Betts M., 2010) and (Mrdalj S., 2011) outline the mounting pressure on the Masters of Business Administration (MBA) programs to provide courses to prepare their students to become the next generation of BI workers but also point out the insurmountable obstacles like high technology costs, complex maintenance requirements, and steep learning curves. A big part of these issues become non-relevant in the case when the goal is to teach self-service BI. For the purposes of delivering the theoretical background to the MBA students at FEBA we have presented in two lectures the whole BI architectural platform (figure 1), its important role and the approaches to include any external information in a self-service BI analysis.

For the laboratory classes we have considered the data warehouse-data mart infrastructure as a black box that provides clean and integrated data sets for the purposes of our analyses. Thus we have concentrated on the BI tools and we have used an MS SQL Server just as a data source (the main one but not the only one).

As a BI tool we have used Excel from MS Office Professional Plus 2013, extended with the Power Query, Power Pivot, Power View and Power Map add-ins, installed on stand

alone machines in a computer laboratory. We have used Power Query and Power Pivot to do rapid prototyping of the data integration – to include the information from multiple sources in our own Power Pivot data model and next we have extended it with custom calculations. We have created different data presentations and explorations through Power View and the students have witnessed the idea to use the BI for intuitive data navigation – to give opportunity to the user to explore and analyze the data without the necessity of any training or support from IT professionals. The students have seen also the value of Power map when the analysis involves some geo-located data.

## 4. Student's feedback

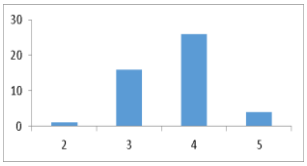
The students have been active listeners during the lectures and keen participants in the classes of this course - they seemed to be motivated to gain more BI-related skills. Trying to quantify the presence of a higher motivation as well as to get the students' opinion for the specific tool that was used the tutor has decided to conduct a short survey.

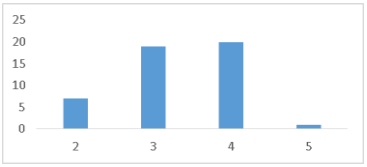
**Table 1. The survey questions.**

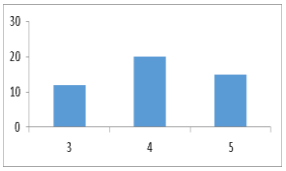
| | **Question** |
|---|---|
| 1. | How much intuitive and well organized are the BI add-ins for Excel, according to you? |
| 2. | To what extent you find the Excel add-ins for BI easy for use? |
| 3. | How much the offered features would facilitate teamwork on an issue? |
| 4. | To what extend the features that have been presented to you meet your expectations for a functional tool for business intelligence? |
| 5. | According to you, how much the professional decisions of a management specialist are being delivered not only on the basis of some theoretical knowledge and practical experience, but also on the basis of an adequate interpretation of the data from the internal and the external company environment? |
| 6. | How much you estimate your initial Excel skills (at the beginning of the course) as enough to allow you to use the core functionalities of the tool? |

**Table 2. Descriptive statistics for the data set.**

| Statistic | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 |
|---|---|---|---|---|---|---|
| Mean | 3,7021277 | 3,3191489 | 4,0638298 | 3,787234 | 4,4680851 | 3,0212766 |
| Std. Error | 0,0957704 | 0,1101105 | 0,1113547 | 0,1135426 | 0,1131954 | 0,1157742 |
| Median | 4 | 3 | 4 | 4 | 5 | 3 |
| Mode | 4 | 4 | 4 | 4 | 5 | 3 |
| Standard Deviation | 0,6565686 | 0,7548793 | 0,7634093 | 0,7784089 | 0,7760285 | 0,7937079 |
| Sample Variance | 0,4310823 | 0,5698427 | 0,5827937 | 0,6059204 | 0,6022202 | 0,6299722 |
| Kurtosis | -0,0288422 | -0,637444 | -1,2449364 | -0,2958394 | 1,1123482 | 0,0376765 |
| Skewness | -0,083235 | -0,296914 | -0,109506 | -0,1808347 | -1,3468079 | 0,5063072 |



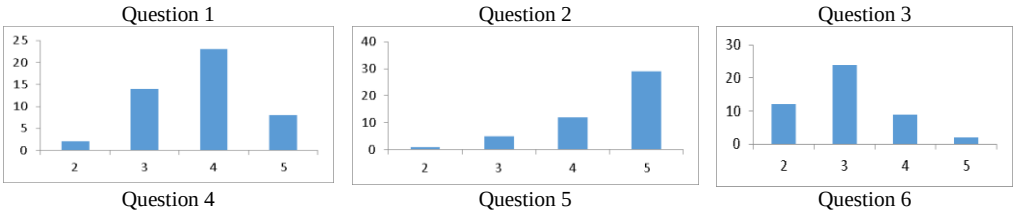



Figure 2. Frequency diagrams for the answers to each survey question.

There were six questions in the survey and the responses were according to a 5-point ordered Likert scale - ranging from 1 as the least assessment to 5 as the highest assessment (Likert, 1932). Table 1 lists the 6 questions in the survey. For example the least assessment for question 2 was "It is very hard to use the Excel add-ins" and the highest assessment was "It is very easy to use the Excel add-ins". The students had one week to reply to an on-line survey and the response rate was 50% which is 47 students.

In Table 2 we have provided the descriptive statistics for the data set composed of the answers. The standard error that is observed allows us to consider the results as confident. Looking at the mean and median, which are very close as values we see that the highest assessments were provided for Q5 and the lowest ones – for Q6 (these are respectively the fifth and the sixth questions in Table 1). From the frequency diagrams presented on Figure 2 we could see that the distributions for these two questions are skewed – respectively negatively for Q5 and positively for Q6.

We have used the Table Analysis Tools and Data Mining client for Excel to analyze the data. From the report presented on figure 3 we see that the question that has received from most of the responders the highest score of 5 is the one which has explored the view that the decisions are data-based (this is Q5). 64% of the students have answered to this question absolutely confirmative and another 21% have agreed with this view by assigning a grade of 4. We could also look at the bottom of the report and see that only 3% of the students think that sometimes the decisions are taken without consideration of the pertinent data.

| Or | Variables | Values | Probability |
|---|---|---|---|
| 1 | The decisions are data_based | 5 | 64 % |
| 2 | Meet my expectations | 4 | 58 % |
| 3 | Intuitive and well organized | 4 | 58 % |
| 4 | Easy to be used | 4 | 52 % |
| 5 | I can use the core Excel functionalities | 3 | 45 % |
| 6 | Support the team work | 4 | 39 % |
| 7 | Easy to be used | 3 | 33 % |
| 8 | Support the team work | 3 | 30 % |
| 9 | Support the team work | 5 | 30 % |
| 10 | Intuitive and well organized | 3 | 30 % |
| 11 | I can use the core Excel functionalities | 2 | 27 % |
| 12 | I can use the core Excel functionalities | 4 | 21 % |
| 13 | The decisions are data_based | 4 | 21 % |
| 14 | Meet my expectations | 5 | 18 % |
| 15 | Meet my expectations | 3 | 18 % |
| 16 | The decisions are data_based | 3 | 12 % |
| 17 | Easy to be used | 2 | 12 % |
| 18 | Intuitive and well organized | 5 | 9 % |
| 19 | Meet my expectations | 2 | 6 % |
| 20 | I can use the core Excel functionalities | 5 | 6 % |
| 21 | Intuitive and well organized | 2 | 3 % |
| 22 | The decisions are data_based | 2 | 3 % |
| 23 | Easy to be used | 5 | 3 % |

Figure 3. Characteristics of the answer set.

Looking at rows 2, 3 and 4 of the report we see that most of the students think that it is easy and intuitive to work with the Excel Power BI add-ins and that the BI-features meet their expectations for a functional tool for business intelligence analysis (they have given again a high grade of 4).

An alternative view of the report on figure 3 is presented on figure 4. From this figure we could easily see that there are no answers in the lower extreme – thus the assessment 1 is

not listed in the legend at all. We see also that most of the answers were neutral or positive, with the prevalence of the positive ones (in green and purple colors).

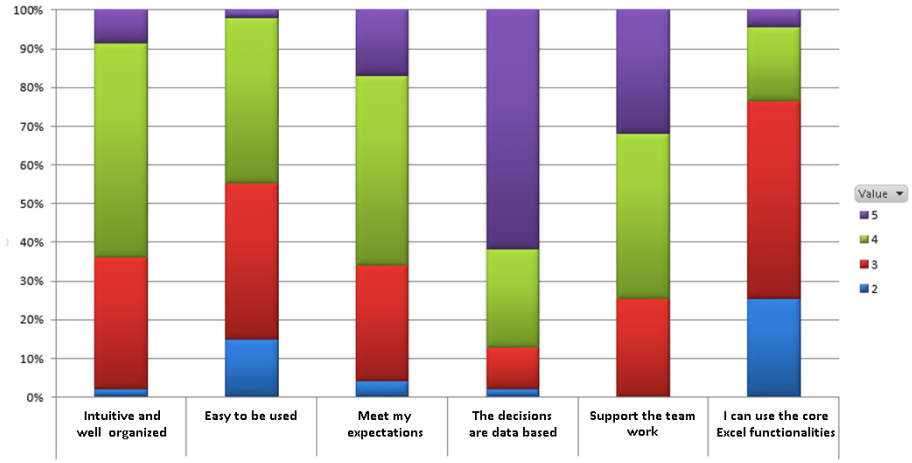



Figure 4. Characteristics of the answer set by question.

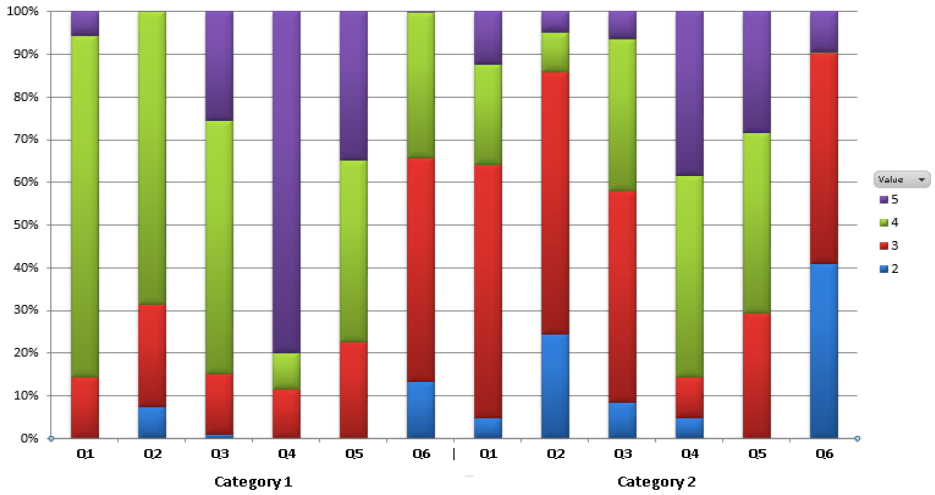



Figure 5. Separation of the answer set to two categories by the data mining algorithm.

If we let the data mining clustering algorithm split the students on 2 groups and perform the same type of analysis we get the chart that is presented on figure 5. We have labeled the questions on this figure by Q1 to Q6 to save space and make the presentation more clear – the order of the questions is the same like on figure 4 (and in Table 1).

When we compare Category 1 with Category 2 we see that the second one contains more negative answers (represented in blue). This is the category where most of the students do not have extensive Excel skills - about 40% of them have assigned grade 2 and about 50% - grade 3 to question 6. We see also that this group of students has answered again more neutrally or negatively to the questions related to the ease of use and intuitivity (Q1 and Q2)

and respectively the expectations for the BI characteristics of the product are being met to a lower extent (Q3).

## 5. Conclusion

The paper has presented an approach to teach Self-Service BI as a part of the topics in an Information Systems and Technologies course from an MBA program. We have delivered the material through a combination of lectures and classes conducted in a computer lab with stand-alone installations of Microsoft Excel Professional Plus 2013, extended with the Power Query, Power Pivot, Power View and Power Map add-ins, as well as with stand-alone installations of MS SQL Server 2014 Business Intelligence Edition.

The students have enhanced their core competencies in terms of improved perception of the opportunities that a good BI solution could deliver to a company and of additional Excel knowledge and skills that could allow them to do self-service BI analyses.

The impression of the tutor from the direct feedback during the course as well as the results from the formal survey support the assumption that the students have appraised positively the inclusion of this topic in the curriculum and have been highly motivated to learn.

We should note also that the survey described in this paper was conducted right after the students have been introduced to the topic of BI through two lectures and one class per group where they had the opportunity to play with a ready-to use sample BI analysis in Excel. The author provides for a second survey, which will include as a basis the same questions and is scheduled to happen at the end of the training – less then a month ahead from the time of writing of this paper.

## References

1. Alexander M., D. J. (2014). *Microsoft Business Intelligence Tools for Excel Analysts.* John Wiley & Sons.
2. Alter S. L. (1980). *Decision Support Systems: Current Practice and Continuing Challenge.* Reading, Massachusetts: Addison-Wesley Publishing Co.
3. Betts M., K. C. (2010, April 5). Professors cite challenges in teaching BI in college. *Computerworld*. Retrieved December 9, 2014, from http://www.computerworld.com/s/article/347880/Professors_Struggling_to_Teach_BI
4. Gartner, Inc. (2008). *Gartner Identifies the Top 10 Strategic Technologies for 2009*. Stamford, Conn.: Gartner, Inc. Retrieved Dec.r 9, 2014, from http://www.gartner.com/newsroom/id/777212
5. Kimball R., R. M. (2010). *The Kimball Group Reader - Relentlessly Practical Tools for Data Warehousing and Business Intelligence.* Wiley Publishing, Inc.
6. Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 1–55.
7. Mrdalj S. (2011). Would Cloud Computing Revolutionize Teaching Business Intelligence Courses? *Issues in Informing Science and Information Technology, 8*, 209-217.
8. Mundy J., T. W. (2011). *The Microsoft Data Warehouse Toolkit - With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset.* Wiley Publishing, Inc.
9. Ponniah P. (2010) *Data warehousing fundamentals for IT professionals*, (2 ed.), John Wiley&Sons
10. Rasmussen, N. G. (2002). *Financial Business Intelligence.* John Wiley & Sons Inc.
11. Sherman R. (2014) *Business Intelligence Guidebook: From Data Integration to Analytics.* Newnes

# System for Interbank Payments - the Effectiveness of Public Versus Private Investment

**Maya Tsoklinova**

University of Forestry, 1756 Sofia, Bulgaria
malenna85@yahoo.com

**Abstract.** The main objective of the research report is to analyze the role of the Central Bank to maintain a stable and smooth operation of payment systems. Special emphasis is placed on the option in which the system is maintained by the private sector with minimal involvement of the Central Bank and the ability to form a market failure. The attention is focused on the participation of the various factors for construction an optimal policy of a Central Bank. It turns out that the optimal Central Bank policy depends largely on whether the payment services are free services, as well as the market structure of the banking industry.

**Keywords.** System, interbank payments, public investment, private investment, factors.

## 1. Introduction

In many countries, the Central Bank is responsible as for price stability and the smooth working of payment systems. Under the Maastricht Treaty, the European Central Bank is obliged to contribute to the smooth operation of payment systems. But if the system is supported by the private sector with minimal involvement of the Central Bank, will there be a market failure? As theorists largely ignore the issue, it is difficult to justify with solid economic arguments the existence of laws that oblige Central Banks to contribute to the payment systems. This report aims to clarify these issues.

**Object of study** in this research report is to define the essential characteristics of a system for interbank payments and the real benefits to society from its implementation.

**Subject of study** in this research report is to determine the optimal participation of the Central Bank and its role in contributing smooth operation of payment systems at impact of specific factors, including the analysis of the effectiveness of public to private investment.

**The main objective** of this research report is to explore and analyze the system for interbank payments, which is supported by the private sector with minimal involvement of the Central Bank and the possibilities for the formation of market failure in its functioning under real conditions.

The nature of optimal participation of the Central Bank may depend on various factors. It turns out that the optimal policy of the Central Bank may depend on whether the payment services are a free service, but also on the structure of the banking system. According to the model, which will be discussed, the market outcome may be seriously impaired if payment services are offered free to the public as a marketing tool. If the customer needs for services are small, the private sector will invest heavily and vice versa. The incentives for the private sector are completely lop-sided in two extreme cases. The first turns out when the market is highly concentrated, and the other puts focus on the same size of market shares. Central Bank plays an important role, since private investments may be insufficient, and therefore that can affect the behaviour of the private sector. The situation becomes different when the payment services are supplied for the incomes generated by fees

that can be earned. Banks can increase their incomes by improving the system quality. Thus, the development of the system becomes profitable when there are many processed payments. This is socially desirable and stimulating structure. A special case is the option in which the market shares of the banks are equal. In such case there wouldn't be serious market failure, but also the participation of the Central Bank is not as important as in terms of free payment services. In fact, the excessive investments by the Central Bank may even worsen the distribution functions, which could arise from the use of private resources.

## 2. Central Banks and interbank payments in the real world

6. The secure and efficient payment system is essential, both for the effective functioning of the economy and the financial stability of a country. The payment system plays an important role in the financial purview, as it represents the main channel for the liquidity distribution. Reliable operation of the payment system ensures secure and timely execution of financial transactions, and thus supports the functioning of the economy.

7. Maintaining adequate payment system is a key objective of any Central Bank. The payment system provides cash transfers between participants and consists of a set of instruments, banking procedures, agreements and systems for inter-bank transfers to facilitate monetary circulation. The financial transfers between economic agents are fulfilled by the payment system. For the transfer, it can be used different types of payment instruments as paper instruments (such as checks) and electronic instruments (such as credit and debit cards). The basis of such payment activities are agreements on financial transfers between economic agents. Payment systems include network connections between these firms and the rules and procedures for such activity. Clearing (**Clearing** - settlement system, which ensures the implementation of the transactions, equates accounts between clearing members, monitors the volume of transactions, informs the members of the stock exchange and the public about open positions and controls the behaviour of participants in the exchange trading.) and settlement (**Settlement** is the final transfer of funds to the current accounts of participants in the payment process.) of securities may be considered as part of the overall payment mechanisms and obey the same rules as final settlement of monetary obligations arising from transfer of securities shall be made in the payment system.

The purpose of payment systems is to centralize payment transactions between banks, thus facilitating the mutual settlement of obligations between them. With the creation of these systems, the aim is to:

- reduce the processing time for interbank payments;
- increase the security and efficiency of the payment process;
- create opportunities for advance monitoring of incoming and outgoing money transfers in order to optimize banks liquidity management.

Clearing can be defined as the process of referral, consultation and in some cases, confirming payment orders or instructions for the transfer of securities prior to settlement, possibly through netting calculation of receivables and obligations of the parties and the establishment of final positions for settlement. Sometimes the term includes settlement, but this use is incorrect. Clearing is called system that performs daily systematic mutual interceptions of receivables takings and obligations between banks. The settlement is the final adjustment of obligations arising from the transfer of cash or securities between two or more parties. Payment systems are a major channel for the movement of capitals. They are essential for the provision of services to final consumers (final individual or entity who is a part to the payment or transfer). Typically, the end user uses the services of a participant in the payment system. Under the participant we understand the bank (or other institution), which has a relationship with the system operator to perform money transfers.

Payment systems provide safe and timely transfer of funds. They allow customers of one bank to make payments to client of another bank. Not, of course, forget the importance of payment systems for the financial markets, without which it could be carried out the purchase of financial assets. Payment systems are used to settle the monetary part of the transaction. Financial market participants rely on the timely execution of settlement in order to provide liquidity for the last transactions [1]. Central Banks use payment systems and monetary policy purposes. According to the time of debiting/authentication of current accounts of participants in the payment system, there are the following types of payment systems:

1. Systems for gross settlement in real time - RTGS (Real Time Gross Settlement System) - debiting and authentication the accounts of participants in real time and on a gross basis (i.e. immediately and individually for each transaction) - according to the time of their submission to the system and balances). Usually they curry substantial volume amounts;

2. Systems for net settlement at a particular time (Deferred Net Settlement System) - a system that at certain times of the day the system settles on the results of the restatement of multilateral or bilateral mutual claims and obligations of the participants. Most often used for the settlement of small amounts (customer payments).

3. Hybrid systems - combined systems characteristics of net settlement at a particular time and gross settlement systems in real time. They netted payment instructions and the time between cycles net is very small, i.e. almost achieves real-time settlement. With these systems saves liquidity.

There are other classifications of payment systems:

- ✓ payment system based on correspondent accounts - funds transfer is done by requiring the correspondent accounts of banks;
- ✓ interbank payment systems - debiting / authentication of customer accounts is carried out within one credit institution;
- ✓ Quasi-systems - operated by private banks, which are included in the interbank payment system (e.g. gross settlement system in real time) and cover specific geographical area or a specific category of customers [1]. From the perspective of private actors bank settlement agent in which the agent (the person making the settlement) using the access channel to the Central Bank. Debit and credit positions between the operator and the participants' quasi-system balance using the correspondent model.

International comparisons show that there are clear differences between countries in the degree of involvement of the Central Bank in payment systems. In Germany and the United States a significant part of the payment system is operated by the Central Bank. In some other countries, such as Britain, the role of the Central Bank is limited and used only a few major payment systems. Efforts of the Central Bank are often essential for the smooth running of the transactions that are subject to interbank payments. In practice, it is possible to create a reliable and effective system for interbank payments without any involvement by the Central Bank. And the smallest transactions between banks are made money with the Central Bank. There are at least three types of investments that the Central Bank can do, namely:

1) The Central Bank may make purely technical investments and repairs of its own computer hardware and software. For example, it can offer various types of alternative systems for the settlement payments made by Central Bank money, or to organize and automation chain payments in order to facilitate operation in the absence of liquidity. These improvements can reduce the number of errors and speed up processes. The Central Bank can also make direct contributions to systems that are owned and operated jointly by the private and government sector. The Central Bank can adapt its rules and practices, so that the system of interbank payments to function well. For example, the rate of net payments can be

increased, which will enable banks to organize faster payment services. Or another example, if the bank's clients make significantly more payments than they receive, the bank may not be able to pay the amount of these payments to other clearing parties unless the Central Bank does not provide sufficient liquidity.

2) The Central Bank may also arrange combinations of regulations and subsidies that lead to the improvement of payment services. For example, the Central Bank may require at least some payments are processed in real time gross settlement (which could be a major burden for banks) and at the same time subsidize the participating banks by offering free services.

There is a vast amount of literature that analyzes the function of the monetary policy of Central Banks. There are several studies that give theoretical insights on the optimal policy of the Central Bank in regard to payment systems. The functioning of different settlement systems is one of the few topics that were analyzed, but the emphasis here is placed on the differences between the net and settlement systems, not in terms of market failure requiring government intervention. There are several theoretical studies that analyze the specifics of the banking industry as an oligopoly or monopoly. In these studies, the focus placed on the role of banks as financial intermediaries between savers and investors. One of the central topics affects the strengths and weaknesses of intermediated finance compared with direct market financing. These models are often based on asymmetric information. It is assumed that banks can control their debtors better than savers can. In addition, there are several studies that deal with payment systems. In most of these previous studies, the focus is on the trade-off between risk and cost-effectiveness in clearing and settlement. If banks coordinate their actions, they usually maintain an optimal balance for settling such accounts in the Central Bank [2]. The aim here is to analyze the issue from another perspective, namely, considering the situation as a struggle for market share. Risks and costs associated with settlement systems are often ignored.

## 3. The role of Central Bank in international perspective

One of the key issues addressed in this report affects the importance of market concentration for optimal degree of involvement of the Central Bank in the small systems for payment services. So we would be interesting to look closely the situation in different countries. Since the model better describes giro system than the system based on check payments, at the forefront of international comparisons must participate countries where check payments play a minor role in the payment system. The table below describes the situation in nine different Member States. Five Member States (France, Italy, UK, Greece and Ireland) were excluded because their payment systems are based on checks and Luxembourg due to the signally nature of its financial industry. The degree of involvement of the Central Bank in retail payments is based on subjective classification. The degree of market concentration is measured in terms of three companies.

As we see, there are moderate negative correlation between the degree of market concentration and the role of the Central Bank. This should be in conjunction with the conclusions drawn below; if the market is concentrated, the Central Bank should not invest heavily in the system. Of course, due to the small sample size, this evidence must be interpreted with caution. Furthermore, the results may be due to effects that have nothing to do with those analyzed in this study. As regards the countries in which the checks were predominant method of payment, the situation varies widely. In some of them, the Central Bank plays a key role in the payment system, as the market itself is highly fragmented (USA, Italy) or quite concentrated (France). In the UK, checking system coexists with an extremely limited role of the Central Bank.

Table 1. The role of Central Bank in international perspective

| | The role of Central Bank – limited | The role of Central Bank – important |
|---|---|---|
| Concentrated bank industry | Finland, Sweden, Denmark, Nederland | |
| Intermediate degree of concentration | Belgium | Austria |
| A fragmented bank industry | Portugal | Spain, Germany |

*Source: Eurostat, http://ec.europa.eu/eurostat/web/main/home*

McAndrews, Bartolini and Hilton [3] have developed a model of duopoly that describes the processing of checks. In the presence of network externalities the benefits derived from the consumption of goods becomes greater if the number of users using the same product is greater. The more users using the same product, there is the greater the benefit in the presence of network externalities. Phone and email are excellent examples of these types of effects. The phone provides utility to consumers, provided that at least two phones are connected to the same network. Liebowitz and Margolis emphasize the difference between network externalities and ordinary network effects. They have called that net network externalities are not commonplace. In contrast, network effects are widespread. The effect of a network occurs, such as when software improves as the number of users who use it increases [4]. However, this effect is not an external factor in the narrow sense, as the number of users does not enter directly into the utility function of each user. The utility generated by the use of the computer and its software cannot depend on the number of other users on the basis of standard. Impact of network externalities in competition between companies was the subject of several recent theoretical studies. Through the joint efforts of awareness, analysis and assessment of risks and the establishment of minimum standards and best practices of Committee and the Central Banks of the Member States in an increasingly close cooperation with the Central Banks of other countries play a leading role in promoting efficient and secure payment and settlement systems. Their work in this area reflects the need for continuous monitoring and improvement of risk management in these systems, which are essential for the stability of financial markets and the world economy.

## 4. Basic principles for systemically important payment systems

In 2001 the Committee with representatives of 23 other Central Banks (including Bulgaria) adopt the basic principles for systemically important payment systems, which formulate guiding principles that promote the development and design of stable and efficient systemically important payment systems (basic systems in the countries). They contribute to the establishment of basic standards and best practices that are considered essential for strengthening the global financial infrastructure. Fundamental principles provide guidance for Central Banks and international organizations such as the International Monetary Fund (IMF) and World Bank in their efforts to improve the design and efficiency of payment systems on a global scale and contribute to improving and maintaining financial stability.

Ten key principles for systemically important payment systems and four responsibilities of Central Banks in their implementation are outlined.

1. The system must have a good legal framework covering all relevant jurisdictions.
2. The rules and procedures of the system should give participants a clear picture of each of the financial risks that arise from participation in the system.

3. The system should have clearly defined procedures for the management of credit risk and liquidity risk, which define the respective responsibilities of system operators and participants to provide appropriate incentives to manage and contain those risks.

4. The system should provide timely and final settlement on the value date, preferably during the working day and at the latest at the end of the day.

5. A system that enables multilateral netting must ensure timely implementation of the settlement in the event that the participant with the largest commitments to meet them.

6. Assets used for settlement is preferable to be money in the Central Bank. When the payment system used other assets, they should wear little or not carry any credit risk or liquidity risk.

7. The system should ensure a high degree of security and reliability of operations and should have established measures for timely implementation of daily operations in the event of an emergency.

8. The system should provide a means of making payments that are practical consumer-oriented and cost-effective.

9. The system should have objective and publicly disclosed criteria for participation that permit fair and open access.

10. Organization of the corporate governance system must be efficient, accountable and transparent.

In this respect the recommendations to Central Banks applying these principles should be as follows:

1. The Central Bank should clearly define its objectives and should publicly declare their role and policy.

2. The Central Bank should monitor the observance of fundamental principles.

3. The Central Bank must ensure that the basic principles are respected by systems that do not administer it, and must be able to carry out this oversight.

4. The Central Bank should cooperate with other Central Banks and relevant domestic or foreign authorities.

Oversight of payment systems in each Central Bank estimates payment systems based on these basic principles. Different payment systems have varying degrees of importance. Accordingly, the relevance of these general principles varies. For example, gross settlement systems in real time are systemically important payment systems (Systemically Important Payment Systems - SIPSs). Therefore we shall apply to them all the basic principles. Furthermore except systemically important payment systems, there are other types of payment systems - for example, especially important payment systems for small payments (Prominently Important Retail Payment Systems - PIRPS). Last do not expose participants to the systemic risk in the same way as systemically important payment systems. Hence, they do not apply all the basic principles. Definition of the different types of payment systems by categories is the responsibility of the Central Bank.

The system of cashless adjustments of settlement relations between enterprises increases the involvement of banks in the payment process. They become centre through which pass immensely payments related to the settlement of various types of settlement relations. Moreover, most of the payments are made through a minimum of two banks. This leads to the emergence of many settlement relations between the banks themselves, which should be settled as soon as possible. To achieve this effect, with decision of the BNB Governing Council, our country is built Banking Integrated System for Electronic payments (BISERA), codified as unified bank standard BUS 7092 - Settlement Bank. Its purpose is to incur and settle final payments between banks by transferring the amounts on their current accounts in the BNB electronically. It is an organization for the electronic transfer of

payments. Its object is only interbank payments. These are payments between customers, accounts of which are held in various bank addressable units constituting the various banks. Interbank transactions involving payments between customers, accounts of which are held in various bank addressable units (links) of the same bank are governed outside of BISERA by the so-called internal estimates. In computing system servicing bank, each payment document has been introduced. After completion of the working day each bank participating in the system for electronic transfer of BISERA sends by telecommunication or CD arrays of necessary records of payments made by each recipient. Once processed the data and specify serial settlement, information databases for results concerning any bank involved in settlement is created. They represent a complete statement of operations for the day by the respective current account and transmitted to BISERA dispatch of the sender. Payment shall be made by reducing the account balance of the payer's bank and increase the account balance of the bank serving the supplier, without allowing partial payments [5].

## 5. Structure of the payment system in the model

The model describes a payment system consisting of the payment service providers and consumers who use the system for payment transactions between them. There are two commercial banks maximizing their profits, which provide their customers with a variety of banking services, including payment services. The payment services could be giro, check, or debit cards, although the model would best describe giro payments. There are N numbers of customers; n is the very large number (n >> 0). In the real world, it is likely that there will be millions of users. As the number of customers is very large, no customer can affect market shares through their own decisions. Each user is in client relationships exactly with one of the two banks. These customers have to make payments to each other. The same clients are payers and payees. Each client must make a payment to another client in the same economy.

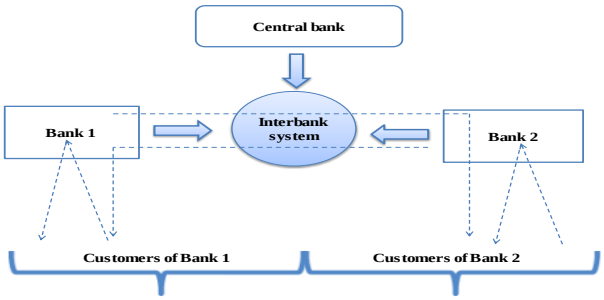



Fig. 1 The role of participants in payment system

Factors that influence to enhance the usefulness of all users in that chain are the following: increasing the personal well-being and improve the quality of payment services. The payment system consists of the following subsystems:

1) Internal systems used for endorsement, checks and debit card payments between customers of the same bank. Both banks have a system of internal bank, the quality of which is exogenous.

2) Interbank money transfer system, checks and debit card payments between customers of different banks. Interbank system is developed and operates between the two commercial banks and the Central Bank. The quality of the interbank system is determined by the investment of both private banks and the Central Bank.

3) Money supply in circulation. Each user in the economy may make a payment in cash to any other user in the economy. Customers do not have to perform all cash payments.

There is no collusion between banks. Exchange of information about the payment is made directly between two banks or indirectly by the Central Bank. Pure clearing services are offered by the Central Bank. It is assumed that there are no restrictions in the capacity of the system. After it is established and developed, all transactions can be processed. In this sense, the model describes the current situation, not the past. Computer systems are usually expensive to establish, but the marginal costs are insignificant. Also tax incentives can be offered on which banks may offer free or discounted services; customers prefer nontaxable benefits such as access to free or subsidized services, compared with taxable interest income [6]. The model represents a game in which all economic agents can always correctly calculate decisions and moves of others and have complete information. Actions are performed in the following order:

1) The Central Bank sets its own investments for the development of the interbank payment system. Commercial banks can immediately observe the level of investment.

2) Commercial banks decide both their investment in the development of the interbank payment system.

3) Customers monitor the quality of the interbank system. Each user selects his bank. These choices are made independently and simultaneously without cooperation. When choosing a bank, the customer does not know to whom to make payment. All other customers are equally likely to be of payees.

4) Each client notes to whom will make his payment.

5) Customers make and receive payments.

Development of a national payment system is a continuous and ongoing process. Planning and implementation of payment system development is a difficult task. These difficulties are due to the complexity and diversity of approaches to reform. Authorities promoting new initiatives in their national payment systems generally look to other countries and to the Committee on Payment and Settlement Systems (CPSS), the International Monetary Fund (IMF), World Bank and other international financial institutions to provide information, advice and help on how best to plan and implement reforms in this system. Development of safe and efficient national payment system is very important for monetary policy, financial stability and the common interests of economic development of the Central Bank. Therefore, Central Banks monitor the development of the payment system to assess its impact on the demand for money, the influence of monetary policy transactions and the efficiency and stability of critically related financial markets.

## 6. Conclusion

National payment system is important for the effective implementation of monetary policy, using payments market transactions to influence overall financial and economic activity. In addition, development of the payment system can affect the speed and predictability of the turnover of cash balances that may affect the overall demand for money in the economy. The Central Bank is a key factor for the development of the payment system. It generally plays a variety of key roles in the payment system. It is the operator, observer of payments and payment service user and a catalyst for reform of the system. Modern Central Banks should consider payment systems as a key area of strategic interest, both in terms of their responsibilities in regard to financial stability and the implementation of monetary policy.

## References

1. Beev I., Mixed economy: quasi-market segments, "Application of information and communication technology and statistics in economy and education" (ICAICTSEE-2014), University of National and World Economy, Sofia, 2014.
2. Goodhart, C., The changing role of central banks, Monetary and Economic Department, Bank for International Settlements Communications, 2010, pp. 23.
3. Bartolini L, Hilton S., McAndrews J., Settlement delays in the money market, Staff Reports 319, Federal Reserve Bank of New York, 2008.
4. Liebowitz S.J., Margolis St., Path Dependence, Lock-in, and History, Journal of Law, Economics and Organization, Oxford University Press, vol. 11(1), 1995, pp. 205-226.
5. Българска Народна Банка, http://www.bnb.bg/AboutUs/AUFAQ/CONTR_PAYMENT_SYSTEM_FAQ
6. Kauko K., Developing an interbank payment system – Efficiency of public versus private investments, Bank of Finland Studies, 2000, p. 37.

# Teaching Information Systems for Bachelor Programs in Computer Science

**Juliana Peneva, Stanislav Ivanov**

Dept. of Informatics, New Bulgarian University

**Abstract.**Today the information systems are common basis to achieve operational excellence, managerial diversity and improved decision making. The study of information systems involves sources from technical and behavioral disciplines. The technical approach includes computer science, management science, and operations research. The behavioral approach concerns psychology, sociology, and economics. This paper focuses on some didactical issues when implementing the technical approach to the study of information systems. The main discussion concerns the choice of the topics and learning activities for a course in information systems delivered to students majoring computer science or information technologies. We assert that students have to achieve information's systems literacy which encompasses how and why the technology is applied in organizations. We adhere to ACM and AIS IS 2010 Curriculum Guidelines.

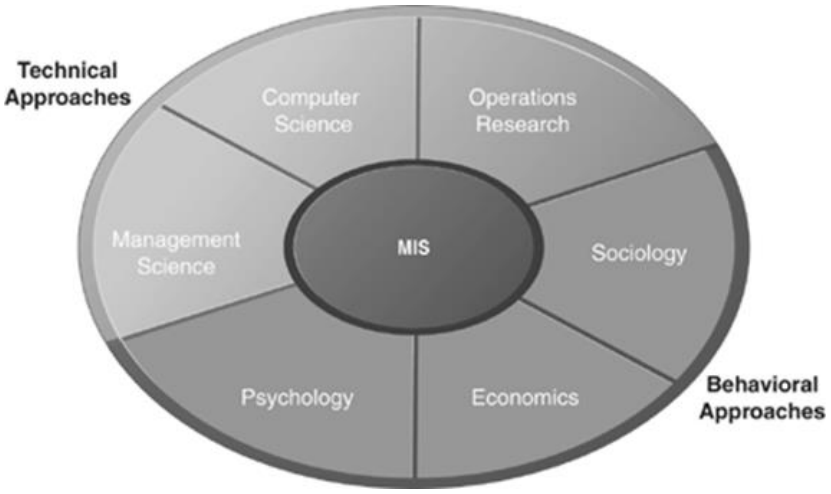
**Keywords:** information systems, curriculum issues

## 1.Introduction

Nowadays information systems have become an important managerial tool that helps organizations to operate in a global economy thus representing a foundation for conducting business. Organizations are trying to become more competitive and efficient by transforming themselves into digital firms where nearly all core business processes and relationships with customers, suppliers, and employees are digitally enabled. Businesses today use information systems to achieve operational excellence; new products and services and an improved decision making. From a business perspective, an information system provides a solution to a problem or challenge facing a firm and provides real economic value to the business.

The study of information systems involves contributions from technical and behavioral disciplines. The technical approach includes computer science, management science, and operations research. The behavioral approach concerns psychology, sociology, and economics (see Fig.1) [1].

In the context of the above the goal of this paper is to focus on some didactical issues when applying the technical approach to the study of information systems. The main issue discussed concerns the choice of the topics and learning activities for a course in information systems delivered to students majoring computer science or information technologies. We assert that students have to achieve information's systems literacy which encompasses how and why the technology is applied in organizations. So, there is a need to find the proper balance between theory and practice. This balance could be achieved through the choice of relevant topics to be covered. We adhere to ACM [5] and AIS IS 2010 [4] Curriculum Guidelines. Finally we briefly discuss the relationship of information systems courses to other courses in computer science and information technology majors.




## 2.The Role of Information Systems in Business Today

An information system represents a combination of management, organization, and technology elements. The management dimension of information systems involves leadership, strategy, and management behavior. The technology dimension consists of computer hardware, software, data management technology, and networking/telecommunications technology (including the Internet). The organization dimension of information systems involves the organization's hierarchy, functional specialties, business processes, culture, and political interest groups.

Any information system supports a series of value-adding activities for acquiring, transforming, and distributing information to improve management decision making, enhance organizational performance, and, ultimately, increase firm profitability. Information technology cannot provide this value unless it is accompanied by respective changes in organization and management. The way the organization executes its business depends on the specific business processes and information systems facilitate their automation.

Information systems support the major functional areas of the business. Manufacturing and production systems deal with the planning, development, and fabrication of products or services, and control the flow of production. Finance and accounting systems manage the firm's financial assets and fund flows. Human resources systems manage employee records and track employee skills e.g. job performance, and training; and take care for employee compensation and career development.

In contemporary organizations information systems are serving operational, middle, and senior management. Systems serving operational management are transaction processing systems (TPS). Management information systems (MIS) and decision support systems (DSS) provide middle management with reports and access to the organization's current performance and historical records. Finally executives support systems (ESS) support senior management by providing data of greatest importance to senior management decision makers, delivered as graphs and charts or other convenient form.

In order to operate efficiently the organization also applies different kind of integrated enterprise applications e.g. supply chain management systems, customer relationship management systems, and knowledge management systems. These systems are designed to support organization-wide process coordination and integration. Supply chain management

systems help the firm manage its relationship with suppliers to optimize the planning, sourcing, manufacturing, and delivery of products and services. Customer relationship management uses information systems to coordinate all of the business processes surrounding the firm's interactions with its customers to optimize firm revenue and customer satisfaction. Knowledge management systems enable firms to optimize the creation, sharing, and distribution of knowledge to improve business processes and management decisions. Enterprise systems integrate the key internal business processes of a firm into a single software system. They span multiple functions and business processes and may be tied to the business processes of other organizations. Due to the use of Internet technology and standards, Intranet facilities assemble information from various systems and present it to the user in a Web page format. Extranets make portions of private corporate intranets available to outsiders.

Information systems and the organizations in which they are used interact with and influence each other. They are to be designed to serve the needs of important organizational groups and conforms the organization's structure, task, goals, culture, politics, and management. In addition information systems help businesses and leverage their core competencies by promoting the sharing of knowledge across business units. They facilitate business models based on large networks of users.

Summarizing, information systems must be employed strategically in order to improve the organization's efficiency, effectiveness, and profitability. They provide most of the communication tools that organizations need for managing businesses on a global scale. Innovative managers and their employees need information system literacy as it concerns basic technologies, applications and development strategies. They need to be familiar with the powerful communication and information tools available to them.

So, what should students majoring computer science or information technologies know? First of all they need a general knowledge on the foundations of information systems so as to take an active role in the future of technology as a key enabler to the effectiveness of business organizations and a truly global society. Next, students are to be familiar with technological innovations in computer hardware and software, data management and networking. Last but not least they have to be acquainted with information system engineering issues.

The course provides focus on managing in a digital world though valuing, securing, building, and ethical comprehension of Information Systems. Taking into account that we apply the technical approach students will gain the experience required to influence technology-enabled growth in whatever industry or pursuit they choose.

## 3.Body of knowledge of the course Information systems

This is a required course in the CS and IT core curricula designed to introduce systems ideas, concepts, and principles of Information Systems (IS) that every student of technology should know. The course should also represent the managerial aspects and to give an understanding of how Information Systems can be used to support the organization's strategic objectives. Implementation of new computer systems must be evaluated for feasibility (Can they be implemented?) and cost effectiveness (Should they be implemented?). The importance of keeping the user involved in the development or purchase process will be emphasized.

In the information age, information technology is widely used strategically to improve the organization's efficiency, effectiveness, and/or profitability. Today, information systems provide most communication tools that organizations need for managing businesses on a global scale. To be effective in a worldwide market, managers need to be familiar with the

powerful communication and information tools available to them. This course offers a framework that can be used to examine technology needs with an emphasis on satisfying business objectives.

The course introduces the fundamental concepts of information systems. It helps understanding issues concerning hardware, software and data management. A classification of the different kinds of information systems is presented. The course comprises topics related to software modeling process, analysis and modeling of data and information systems development. Students have to achieve information's systems literacy e.g. how and why information technologies are applied in organizations.

Upon completion of this course, learners should successfully demonstrate the following generic and specific competences:

1. Evaluate current level of understanding and competence with Information Systems as outlined in the course.
2. Integrate the informational knowledge with personal career.
3. Demonstrate the ability to use critical thinking skills to evaluate an area of Information Systems that is directly related to current career.
4. Illustrate the ability to communicate in an online environment via learning content management system Moodle and create a sense of a learning community.
5. Discuss the role of information systems in enterprises, including the digital firm.
6. Learn and implement the concepts of e-commerce and e-business.
7. Evaluate the suitability of various configurations of hardware, software, database and telecommunications to solve business problems.
8. Evaluate the uses of the Internet and intranets for business.
9. Evaluate the potential impact of new information technology on business processes.
10. Explain the importance of protecting organizational information technology resources and list strategies for providing that protection.
11. Examine the ethical issues pertaining to managing information systems in terms of intellectual property, privacy and security.
12. Discuss information systems security and control.
13. Develop strategies for managing information technology to obtain a competitive advantage.
14. Compare/contrast the activities in each of the phases of the Systems Development Life Cycle.
15. Evaluate types of DSS, MIS and ESS.
16. Explain technical opportunities for global value chains.

A balance between theory and practice is respected following ACM and AIS IS 2010 Curriculum Guidelines. Students will be able to:

1. Compare and contrast the roles of hardware, software, people, and data resources in an information system.
2. Relate a computer user's needs to the appropriate hardware and software resources.
3. Evaluate the effect of data communications on business operations.
4. Compare and contrast the functions and capabilities of Transaction Processing, Management Information systems and Decision Support systems.
5. Relate the level of organizational support to the success of end-user computing.
6. Relate the level of user involvement in a systems development project to the success of the project.

7. Explain the effect of information technology on the managers, structure and activities of organizations.

To achieve the educational goals of this course, the following core concepts are to be mastered:

1. Globalization through Information Systems software.
2. Managing in a Digital World: Information Systems Hardware.
3. Enabling Commerce using the Internet and World Wide Web.
4. Information Systems Classification.
5. Valuing Information Systems Investments: Database Management.
6. Managing the Information Systems Infrastructure: Networking and Cloud technology.
7. Development of Information Systems.

## 4.Learning activities

*Information Technologies Initial Pre self-analysis:* Write a one page assessment of students' current state of understanding Information Systems as it applies to relevant field of study. The students' knowledge of Information Technologies as it applies to the field of study. Review the recommended course schedule or the text book for the topics covered in the class. Include one outside resource. Cite this resource in the body of the text and list it as a reference, using the APA format. Post the assignment on Moodle.

*Globalization Paper:* Globalization has had a significant influence on how companies accomplish tasks over the last 20 years, and globalization has been a major driver of changes in the computer industry. The students will write a five to eight page paper on the influence and impact of globalization on companies, people, projects, technology, and the art of doing business. The students should look at globalization from an Information Systems viewpoint – what communication systems, data sharing systems, political, ethical, legal, geographical and economic factors come into play when a company is considering outsourcing work to another company. The paper will be graded on the degree of research, addressing all the technological and social variables that influence an outsourcing decision including political and economic concerns that can influence the decision to outsource to a company in another country. The students should pay attention to the local and global impact that outsourcing has on organizations and society for both the host and outsource countries and employees. Post the assignment on Moodle.

*The Influence of Technology Paper:* Technology has become part of the everyday fabric in our daily life. Computers and the capabilities programmed into them help provide the basis of our modern society. Examples range from food production and distribution to entertainment, news, games, and just about everything we touch or use. The student will write a ten to fifteen page paper on the differences between the technology in use in 1990 and the technology in use today. The student should critically evaluate the pace of change over the last 60 years and predict where technology will be in 2020. The student will review and identify the job skills in management and technology needed for continually development of professional identity in relationship to how work, play, and life will be lived in 2020. Post the assignment on Moodle

*Quizzes:* There are two short (up to 30') quizzes during the course. They comprise multiple choice, fill in the blanks and yes/no questions. Quizzes can be considered as training for the exams. There are no make-ups! If an absence is excused, the quiz will be given at the soonest possible time, after the student returns.

*Panel Discussion and Topic Presentation:* Students are supposed to take part in different panel discussions. The topic is to be chosen from the learning content. Students are expected to develop a PowerPoint slide show which should demonstrate a high quality design and have at least 10 slides.

## 5.Examinations and grading

The grade comprises the evaluation of the current learning performance and could be composed as follows:

| Overview of Required Assignments | % of Final Grade |
|---|---|
| Information Technologies Initial Pre self-analysis | 10% |
| Globalization Paper | 10% |
| The Influence of Technology Paper | 20% |
| Panel Discussions and Topic Presentations (Participation) | 10% |
| Quizzes (two equally weighted) | 30% |
| Final Exam including Information Systems Final post self-analysis | 20% |
| TOTAL | 100% |

## 6.Conclusion

IT programs should produce graduates who apply information technologies in a wide range of settings. The IS courses for CS and IT specialists should focus on the technical and developmental aspects, but on the organizational and managerial views as well. This is because of the twofold role these specialists play: designers and developers, but managers of their own work. Organizations of all kinds are dependent on networked computing infrastructure and IT people should be prepared to compose, deploy, and maintain this infrastructure, nevertheless in the context of Information Systems or other.

The Information Systems course itself is a specific focus in every Computer Science program around which many other courses can be integrated and expose the value of competences they produce. We can mention here courses like Database systems, Data mining, Network technology, Web design, Software engineering etc. This predetermines its position in the upper level of the bachelor program.

## References

1. Laudon, K. C., & Laudon, J. P. Management information systems: Managing the digital firm (13th Ed.). Prentice Hall. (2014).
2. McLeod R., Schell G. Management Information Systems, X-th Ed., Prentice Hall, 2006
3. O'Brien J., Marakas G. Management information systems (10th ed.). McGraw Hill/ Irwin, 2010.
4. IS 2010 Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems.
5. Computer Science Curriculum 2013: Ironman Draft (Version 1.0)

# New Features and Benefits of the Bulgarian Payroll Software

**Borislav Boyanov**

Department "Accountancy and Analysis", UNWE, 1700 Sofia, Bulgaria
bboyanov@unwe.eu

**Abstract.** This paper assesses the state of the Bulgarian market of software for organization and management of human resources in the enterprises. The mostly commonly used payroll software is presented. The main focus is on traditional and emerging functionalities. Discussed are the advantages and new features for automated relationship management staff. Comparative analysis of the most widely used Bulgarian payroll software is made. The benefits of automated calculation and reporting of salaries, social security contributions, tax liabilities, the preparation of statistical reports and the creation of electronic diaries and electronic files of staff are highlighted.

**Keywords.** Human resources, payroll, accounting, organization

## 1.Introduction

The dynamics and trends of the Bulgarian market of specialized software for human resource organization and management are developed under the influence of a number of direct and indirect factors. The major market law providing for that a product demand defines the requirements to its production and supply applies to the software market as well. To this end, the size and the structure of business entities looking for payroll software defines the requirements to its developers.

The structure of the Bulgarian economy, with view of the number and size of operating business entities during the recent years, is characterized with the presence of 98 % micro- and small enterprises with average number of employees between 0 and 50 people. Enterprises' small sizes suppose the availability of small budgets and allocation of limited resources for auxiliary and administrative activities. The limited financial resources of most of the enterprises in the country turned out to be favourable base for the development of business services (financial, accounting, administrative) sector. In parallel, the scientific-technological revolution and the development of information and communication technologies allowed the establishment and the stormy growth of the new economic sector – ICT. The most innovative and the most realistic strategies of the software companies enabled them to become leading stakeholders on the payroll software market, and at the same time they managed to conquer the biggest market shares, to win competitive advantages and to develop short- and mid-term market trends.

## 2.Bulgarian Market of Payroll Software – Retrospective, Status and Trends

During the last 10-15 years, along with the extremely fast growth of ICT, the market of software applications for human resource management follows the global trend for continuous improvement of traditional opportunities and functionalities of the products, by introducing new software applications operating on totally different base that comprises

much wider scope of activities and processes. For example, at the beginning of the 21st century, the payroll software was developed to work under MS DOS or on the basis of MS Access – when working with the operating system MS Windows, now the MS SQL platform is widely used.

Researches on the Bulgarian market of payroll software show the following status and trends:

**A. Sustainable number and structure of offered payroll products**

- The market of automated software applications for human resource management in Bulgaria takes a share of 15-20 % of the whole range of offered business software;

- The number of offered standard employee management software applications during the last decade remains stable;

- New products that permanently stay on the market appear continuously. The main reason for the difficult penetration on the market is the presence of already established market leaders, the high level of user's satisfaction, as well as the significant potential expenses for purchase of new equipment, staff training and requalification;

- Opportunities for successful and permanent establishment of new products are available when one and the same developer offers a new generation of already offered software.

**B. Main criteria for purchase of new software**

- Extremely strong price competition exists. The product price has the highest relative share upon making the decision for purchase of payroll software;

- In a situation of "price war" product developers face serious financial challenges. On one hand, users are looking for cheap products, with the lowest possible price for service and subsequent maintenance, and on the other hand – the human resource price for IT specialists on the labour market is one of the highest prices. These two opposite trends extremely expand the "price scissors" and eliminate the smaller market stakeholders. Thus, a few dozens of developers remained on the payroll software market in Bulgaria;

- During the last decade all users who are looking for payroll software realise that the price should not be the single factor. When introducing new products, the importance of factors, such as possibility to process large information sets, timely update of products, compliance with applicable legislation, abilities for specific generation of references and automatic creation of documents to be submitted to different state authorities and institutions, greatly increases.

**C. Provision of technical means (hardware)**

- Moderate application of new programming platforms and introduction of new technologies can be seen, as most of the users are still changing their hardware within 3-5 years, i.e. they work with technical means with outdated parameters and limited capacities for introduction of the latest applications;

- Certain restrictions for updating and for offering completely new products exist, due to fully subjective factors, defined by the users' abilities and attitudes to work, which are characterised with pessimistic and cautious attitude to new products, accompanied by nostalgia and favouritism of payroll products used for a long period of time.

**D. Characteristics of offered payroll software**

- The Bulgarian independent payroll software prevails on the market;

- The application of ERP modules for employee management is more limited;

- The reasons for the high relative share of the independent software in comparison to the ERP modules are not only the modest financial capacities of the enterprises, but there restricted needs too;

- The small sizes of Bulgarian business entities suggest the availability of small number of employees, production of one or several products, provision of a limited scope of services, respectively. This supposes smaller sizes and structure of the remunerations, with limited calculation of additional remunerations and social schemes, lack of the need to account the working clothes and uniforms, food coupons provided, etc.;

- The opportunities for reviving the market are due to the European funding programs, and due to the penetration of foreign companies needing development of payroll software according to individual project;

- Payroll product's functionalities and flexible capacities for user setting and for generation of a great number of references are main factors when selecting such products.

**E. Trends in the development and application of payroll software**

- We are witnessing trends of integrated human resource organisation and management, with automated comprising of more and more key factors for human capital management;

- The interest to introduction of ERP systems allowing the creation of information both for salaries and remunerations, social securities and taxes, leaves, compensations, generation of payrolls, schedules, social security declarations, and for the education and qualification of staff, for continuous assessment of employees, for access control and working hours, for binding additional remuneration and social incentives with economic results from the work of each employee or group of employees.

## 3.Leading Payroll Products on the Bulgarian Market

Author's research shows that the Bulgarian market offers not more than 100 human resource management systems. Table 1 shows the 25 products most used for human resource management by Bulgarian micro- and small enterprises. The information systemised in table 1 shows data about the name of the software product, its developer and the operating system where it can faultlessly operate.

**Table 1. Leading Payroll Products**

| No | Product | Developer | Operating system |
|---|---|---|---|
| 1 | **ALADIN** | Microcomplex International - LTD | Windows 2000 / XP / 7 / 8 / 8.X |
| 2 | **Atlas Payroll** | ASS - LTD | Windows 98 / ME / NT / 2000 / XP / 7 |
| 3 | **BulMar ТРЗ и ЛС** | Kreston Bulmar - LTD | Windows 2000 / XP / Vista / 7 |
| 4 | **Business Navigator Payroll** | Commerce Finance - LTD | Windows XP / Vista / 7 |
| **No** | **Product** | **Developer** | **Operating system** |
| 5 | **Fact-Payroll-Staff-NSSI** | RANCOM`91 - GP | DOS, Windows 95 / 98 / ME / NT / 2000 / XP |
| 6 | **FIX** | FIX - Radev and Co - GP | DOS, Windows 95 / 98 / ME / NT / 2000 / XP |
| 7 | **HeRMeS** | TechnoLogica - JSC | Windows 2000 / XP / Vista / 7 |
| 8 | **КТК TRZ** | Kontico - LTD | Windows 98 / Me / 2000 / NT / XP / 2003 / Vista / 7 |

| | | | |
|---|---|---|---|
| **9** | **Microinvest TRZ and LS PRO** | Microinvest - LTD | Windows XP / Vista / 7 / 8 / 8.X (32 bit/64 bit) |
| **10** | **Mirag TRZ and LS** | Brain Storm Consult - Sole LTD | Windows 2000 / XP / Vista / 7 |
| **11** | **Omex 2000** | Omega Soft - LTD | Windows XP / Vista / 7 / 8 / 8.X |
| **12** | **PASOSS - salaries 5** | ZEN Electronix - LTD | DOS, Windows 95 / 98 / ME / NT / 2000 / XP / Windows 7 (32 bit) |
| **13** | **Payroll NT** | Multimedia - LTD | DOS, Windows 95 / 98 / ME / NT / 2000 / XP / Windows 7 (32 bit) |
| **14** | **PERFECT TRZ** | DAN Electronic - LTD | Windows 95 / 98 / Me / 2000 / NT / XP / 2003 / Vista / 7 |
| **15** | **Plus Minus Payroll** | Plus Minus - LTD | Windows 2000 / XP / Vista / 7 |
| **16** | **Policont - human resources** | C+C Automation - LTD | Windows 2000 / XP / Vista / 7 / Linux |
| **17** | **REGATA** | KSoft - LTD | DOS, Windows 95 / 98 / ME / NT / 2000 / XP / Windows 7 (32 bit) |
| **18** | **RZWIN, HONWIN, LSWIN** | Omega-Team Bulgaria - Sole LTD | Windows 2000 / XP / Vista / 7 |
| **19** | **SANTA - TRZ** | GRAVIS Bulgaria - JSC | Windows 2000 / XP / Vista / 7 |
| **20** | **STYLE - TRZ and LS** | Em Soft - LTD | Windows 98 / Me / 2000 / NT / XP / 2003 / Vista / 7 / 8 / 8.X |
| **21** | **TaSaT - Payroll** | TaSaT - LTD | Windows 95 / 98 / ME / NT / 2000 / XP / Windows 7 (32 bit) |
| **22** | **TERES - TRZ and LS** | ADA Soft - LTD | Windows 2000 / XP / Vista / 7 |
| **23** | **TEREZA** | Tonegan - LTD | Windows 2000 / XP / Vista / 7 |
| **24** | **TRZ (contracts and salaries)** | Kodar - LTD | Windows 2000 / XP / Vista / 7 |
| **25** | **ZAPY** | ComByte - LTD | Windows XP / Vista / 7 |

## 4.Benchmarking Analysis of New Capacities, Functionalities and Benefits from the Use of Bulgarian Payroll Software

This benchmarking analysis is carried out on the basis of author's subjective assessment of objective indicators and data, as a user of payroll software and as a representative of one of the developer companies, specified in table 1. The name of the company and its product are not mentioned intentionally, with view of the research objective and characteristics of this publication, as well as for the purpose of avoiding accusations in lobbying performance of hidden advertising campaign.

The benchmarking analysis will analyse the products "Microinvest TRZ and LS Pro", "STYLE - TRZ and LS", and "Omex 2000".

**A. Microinvest TRZ and LS Pro**

The software application ***Microinvest TRZ and LS Pro*** is developed by the company "Microinvest" LTD. It appears on the market at the end of 2011 and at the beginning of 2012. It is a successor of the product ***Microinvest TRZ and LS***, developed at the end of the 1990s. The program is developed on the basis of MS SQL Server, which suggests security and continuous protection of information and opportunities to work with large amounts of data. The application gives the opportunity to process staff data of many companies, regardless of the number of employees. Furthermore, it ensures access control with different administrator rights, work from remote places, creation of electronic employment dossiers for the staff, automatic enumeration of issued documents and electronic document registry. According to the price factor, the product ***Microinvest TRZ and LS Pro*** falls in the lower price range – 99 EURO, VAT exclusive. Subscription service of the product is according to the principle of purchasing a specific number of update codes, where a package of 1 code costs 29 EURO, VAT exclusive, and the maximum package of 10 update codes costs 199 EURO, VAT exclusive. Product's simplified interface, standard configuration, in combination with affordable price, timely update and good maintenance of the product, are among the main reasons to be preferred by more and more users in comparison with other payroll products on the market.

The program is developed on the basis of the following menus – "File", "Operations", "Edit", "References" "Other", Windows" and "Help". Working sequence is as follows:

1. Creation of new company and input of corporate data – name, address, manager, identification number, code and name of the economic activity.
2. Input of staff data (personal cards).
3. Staff data editing and setting.
4. Calculation of monthly remunerations, deductions and compensations.
5. Generating standard and customised references.

**The new functionalities of the application are as follows:**

1. Introducing the economic activity of the enterprise during database configuration allows every employee's personal card to generate information about the minimum social security income according to the position taken and the amount of the social security contribution to Occupational Accident and Occupational Disease (OAOD) Fund.

2. When generating a personal card, the program allows to attach a picture of the employee.

3. For users' convenience, when entering the zip code of the city, town or village where the person lives, there is an option to automatically retrieve information for the district, municipality and city, town or village where such person lives.

4. Each personal card has a different section "Absences Data" where different types of absences – sick leaves, annual leaves, absence periods, and the respective remunerations or compensations payable to the worker, are entered.

5. The following sections are created:

- "Documents", where forms of employment contract, job description, requests, orders, declarations and applications related to the occurrence, amendments and termination of employment relation, are automatically generated;
- "Bank accounts" – to enter information about the worker's bank account that is used for the transfer of remunerations, or for announcing it to the state social security (SSS), respectively, with view of receiving compensations and social aid.

6. Menu "Operations" allows automatic calculation of monthly remunerations under employment contracts, service contracts and management contracts, as well as automatic calculation of obligatory social security declarations, form № 1 and form № 6 with option for

data export and printout. Furthermore, this menu gives the opportunity for generation and export of data to external carrier required by the National Revenue Agency (NRA) for concluded employment contracts, files to banks for transfer of remunerations and social security contributions, as well as creation of statutory files for sick leaves and compensations for temporary working incapacity submitted to the National Social Security Institute (NSSI). There is an option for automatic generation of "Reference under article 73 of the Individual Income Tax Act" for annual incomes paid to individuals.

7. Menu "References" allows the retrieval, printout and export of the following types of references in files in *.pdf, *.doc, *.xls format. It further allows to retrieve the following types of references, both for all employees of the company, and by departments, units, workshops and individual:

- Payrolls for employment contracts, management contracts, for self-employed persons;
- Pay-slips, certificates and reports for paid amounts;
- Reconciliations, references for social security contributions and taxes due to the budget;
- Generation of accounting references, show/ no-show schedules, registers for sick leaves and annual leaves;
- The references generating the documents required for retirement – form UP-2 and form UP-3, as well as the automatic generation of income certificates, quarterly and annual "Reports for employed persons and wage and salary expenses" required by the National Institute of Statistics (NIS) are among the new functionalities of the product;
- The function "Customised References" is of great benefit for the product user. It allows to customise the required reference by selecting the necessary data and indicators among all possible variables of the program.

8. Other functionalities and benefits of the product – here we can mention the options for automatic generation and export to the accounting product ***Microinvest Delta Pro*** of accounting entries for reporting salary and social security expenses and for the company's accounts with its staff. The application has an integrated functionality to verify the correctness of the VAT number, BULSTAT (UIC), PIN and IBAN. The option for generation and printout of banking documents for transfer of social security contributions and employee benefits is among the new functionalities of the application.

**B. STYLE - TRZ and LS**

***STYLE - TRZ and LS*** is a specialized software product for reporting the company's labour resources. The product is developed by the company "MSoft" LTD and is offered in two versions:

a. "TRZ and LS" – developed on the basis of MS Access. It is appropriate for companies with up to 200 workers and employees.

b. "TRZ and LS Pro" – it uses MS SQL Server and is appropriate for use by all companies, regardless of the number of employees.

The product ***STYLE - TRZ and LS*** is offered by granting different types of licenses, depending on the number of employees for whom information will be created and processed, and whether information for one or many companies will be processed. Depending on these criteria, direct costs vary within the range of 200 EURO to 7300 EURO, VAT exclusive. The annual maintenance of the product varies between 50 and 1200 EURO, VAT exclusive.

**The program has the following structure and working sequence:**

1. Entry of main data for the company – name, EIK, economic activity, officers, bank accounts.

2. Main form of the program – this is a hierarchic structure where information about the company's organisation chart, its departments and units, positions and specific personal data of each employee, is entered. The main form includes three groups of fields - "Permanent position chart", "Personnel" and "Remunerations".

3. "Permanent position chart" – the company's structure by departments, units and the positions in them are initially configured.

4. "Personnel" – The second step is to enter personal data of each worker. This products works on the principle of creation of worker's personal card with entry of personal data, official data, information about education and qualification, as well as data for remote access of the worker to the Web module of the program. The field "Personnel" gives the opportunity to enter obligatory data for each employment contract, supplementary agreement, order for termination, leaves and sick leaves. In addition to the entry of information for the employment contract, there is an option for automatic generation of file for submission of Notice under article 62 of the Labour Code to the NRA.

5. "Remunerations" – it contains three elements "Pay-slips", "Charges" and "Deductions". Here users can enter information about the basic salary, additional remunerations and deductions from thee worker's salary, and to set parameters for generation of social security declarations form № 1 and form № 6. The periods of paid and unpaid leave, and data for each sick leave of the worker can be entered in this section of the program.

In this menu the user creates information for the work schedules of the company, the working days and the working shifts, which allows automatic calculation of the remunerations for day, night and overtime hours worked-out by the staff. There is an option for import of personal schedules from MS Excel. The program allows to work with "Aggregate reporting of working hours" by making the respective settings, to enter the reporting period and the persons working under these conditions.

With view of compensations payable to the staff upon resigning pursuant to the provisions of the Labour Code, STYLE TRZ and LS requires to add the respective variable for the type of compensation and it automatically calculates the amounts, social security contributions and taxes due, and correctly generates the required files to be submitted to the NRA.

The program automatically generates file for large-scale payments of salaries and social security contributions by means of bank transfer. There is a function for creation and printing of bank payment documents for settlement of due social securities and taxes and for transfer of salaries.

There is an option "Accounting" where data for salaries and social securities are imported to accounting software, upon pre-configuration of accounting entry templates.

***STYLE TRZ and LS* has rich functionality of references and documents in the following areas:**

- Personnel – users can generate a list of permanent positions, application for leave, order for allowing leave, register of used and unused leaves, register of sick leaves, customised references by employees and by positions;

- Salary – it allows users to print references, such as: pay-slips, payrolls, attendance forms, reconciliations by items and by funds, charges, deductions and leaves for specific period, unused leaves, provisions for unused leaves, customised references;

- Employment contracts – it allows users to extract and printout ready forms of employment contract, supplementary agreement, order for termination of employment contract, income certificates, certificates for income under article 45 of the Individual Income Tax Act, certificates upon retirement – form UP-2 and UP-3;

- Service contracts – here users can generate references for paid remunerations by assignments, contributions to be paid by the principal and the contractor, certificates under article 45 of the Individual Income Tax Act, Reference under article 73 of the Individual Income Tax Act for amounts paid to individuals, and customised references.

**C. OMEX 2000**

The program ***OMEX 2000*** is developed by "Omega Soft" LTD – a leading company for development, introduction and integration of software solutions for employee management, salary calculation and client relations management. ***OMEX 2000*** supports the following types of database: Interbase/Firebird and MS SQL Server. The product comprises the following software elements: a) Database server; b) Omex 2000 Server, and c) Omex 2000 Client.

The price of the product, depending on the purchased number of database, licenses and modules, varies within the range of 95 and 495 EURO, VAT exclusive, and the guarantee service and maintenance vary between 25 and 185 EURO, VAT exclusive.

**Upon installation of *OMEX 2000*, the user works in the following sequence:**

1. Initial settings – the user sets the general parameters for work with the system, such as: initial month and year of processing, entry of enterprise's names, addresses, officers, EIK, code of the economic activity. At this stage, the user configures the organisation chart and the positions of the enterprise, as well as the main variables for information generation and processing.

2. Entry of data for workers and employees under employment and service contracts – a personal card is generated for each employee, comprising the following sections: main data, additional data, social security data, employment contracts, absences, charges and deductions, previous amounts, amounts and documents for the NSSI.

After the entry of employee's data, the program allows automatic generation and printing of form documents – employment contracts, supplementary agreements, requests and orders, declarations and protocols for NRA and NSSI. The user is allowed to modify form documents and to add new forms. Similarly to other payroll products, OMEX 2000 provides opportunity to create templates for export and accounting of employment expenses to accounting software.

3. Monthly activities – this group of functions comprises the following activities carried out in the enterprise:

- Appointment of a new employee;

- Dismissal of employee;

- Editing data for remunerations, bonuses, deductions and compensations. The program makes calculations for standard working hours and for overtime, night work, hour rates, shift work, as well as for aggregate reporting of working hours. It allows to add a number of variables stipulated in the labour and social security law, such as bonuses or salary deductions;

- Generating and printing monthly reporting documents – it allows to generate pay-slips and payrolls, references for sick leaves and leaves, certificates under article 45 of the Individual Income Tax Act, reconciliations and tables for social security contributions and taxes due by funds, printing bank payment documents for social securities, taxes and salaries;

- Monthly closing operations – the product automatically generates files with declaration form 1 and declaration form 6, cover letters, and files for sick leaves and short-term statistical reports.

In general, the program is widely applied in a wide range of economic activities – production, sales, services, operations of state organisations and institutions, and therefore it comprises more menus and settings for a number of specific cases, it requires more time and efforts for configuration and entry of initial information, thus needing more serious education and training for the program users.

## 5.Conclusion

On the basis of the benchmarking analysis of the above products, we can conclude that the independent Bulgarian payroll software follows the progressive growth of ICT in general. New functionalities and benefits for the users are due to the application of new, flexible and adaptive programming platforms that allow quick, correct and precise generation of references, registers, declarations and applications required by the labour and social security legislation. During the recent years functions that allow generation of files and printing of income certificates – UP-2 and UP-3, "Reference under article 73 of the Individual Income Tax Act for amounts paid to individuals", short-term and annual "Reports for employees and employment expenses" for NIS, automatic export of data to accounting software and options for configuration and retrieval of a wide range of customised references, have the biggest contribution.

## References

1. Microinvest LTD, Microinvest TRZ and LS Pro - User's Guide, 2013.
2. MSoft LTD, Quick Guide to STYLE - TRZ an LS, 2012.
3. Omega Soft LTD, Instructions for the User - OMEX 2000, 2012.
4. 12 Systems for HRM and Payroll, Technologies and Concepts, issue 3, 2009.

# Utility and Optimal Allocation of Financial and Material Resources in a Small Factory: Preferences Based Approach

**Yuri Pavlov, Valentina Terzieva, Petia Kademova-Katzarova**

ICCS - BAS, Acad. G. Bonchev Str., Bl. 2, 1113 Sofia, Bulgaria
yupavlov15@isdip.bas.bg, valia@isdip.bas.bg, petia@isdip.bas.bg

**Abstract.** The paper describes an application of a mathematical approach to decision-making in tasks with significant human involvement. The difficulty is the complexity of the problem, the formal description of the objectives, considering the empirical knowledge and modelling the behaviour of decision-maker. We use the estimates of decision-makers' subjective knowledge and forecasts as subjective preferences and probabilities combined with existing mathematical models. The purpose is to allocate optimally the resources in a factory based on the expert's evaluation of assorted furniture's production. To realize this we build a utility function that analytically represents the preferences for the key production factors.

**Keywords.** Management, Preferences, Optimal allocation, Utility function.

## 1.Introduction

Planning and allocation of material and financial resources in a factory nowadays is a complicated and elaborated task. It is required the implementation of advanced mathematical theories and models to address them. Powerful computing resources are employed for their fulfillment. In recent years, almost all companies use Enterprise Resource Planning (ERP). They are computer-based systems with various functions for unified management of all processes across an entire company – resource planning, accounting, control and analysis, etc. in order to achieve high efficiency [1]. ERP is intended to enable information sharing, processes coordination and decision making on basis of an integrated set of corporate data. This type of system allows easy planning the workload of the resources (raw materials, financial, human, goods, etc.) at a wide variety of activities and enables the monitoring and flexible adaptation to changing external factors. The provided services can be applied both to an entire company (large or small) and to a particular department/ activity. Such systems allow resources allocation and planning based on mathematical principles in order to comply with the possibilities and resources of the company and, if necessary (in case of changed circumstances) to be updated.

Besides planning, ERP systems are designed also to give prognoses. According to the specific activities respectively anticipate the need of raw materials, labour, type and volume of production and others. The prognoses are often based on subjective judgments related to market trends in complex situations. One of the important functions of ERP systems is to support the management team in decision making process within the organization. This requires the provision of reliable information on the overall processes in the company. As a result of the modelling and according to various possible alternatives different predictions are obtained. They assist the decision-making and give a reasoned choice for future development. The core of ERP resource planning systems includes different mathematical and economic theories and methods.

This paper presents a possible approach based on Utility and Measurement theory for applications in management information systems with a focus on microeconomics. In the considered case an approach for analytical description of complex problems and expert evaluation regarding the assortment of furniture and corresponding materials is applied. The first stage of analytical processing of such qualitative subjective information reveals the main objective and sub-objectives. The purpose is expressed based on the empirical knowledge and subjective level of experience. The second stage is to choose and develop specific algorithms and methodology in accordance with the measurement scales of the sub-objectives (criteria). The aim is to build a utility function that analytical represents an expert's preference in terms of important production attributes. We examined the value / utility independence of the sub-objectives. This part of the investigation determines the structure of the generalized utility function, giving the relationships between the one-dimension sub-functions in explicit form [2]. Mathematically such complex investigations are based on the Measurement theory, Utility theory, Mathematical statistics and pattern recognition, stochastic approximation, etc.

## 2.Theoretical Basis of Utility Function and Preferences

We aim analytical representation of the individual's preferences as utility function for value based decision making. Measurement is an operation in which a given state of the observed object is mapped to a given denotation [3]. In the case when the observed property allows us not only to distinguish between states but to compare them by preference we use a stronger scale, the ordering scale. The preference relation in the ordering scale ($x$ is preferable to $y$) is denoted by ($x \succ y$). If there are incomparable alternatives, then the scale is called *partial ordering*. A "value" function is a function u(.) for which it is fulfilled [2, 4]:

$$((x, y) \in \mathbf{X^2}, x \succ y) \Leftrightarrow (u(x) > u(y)).$$

Under this scale we cannot talk about distance between the different alternatives. Here only ordinal evaluations may be used. If with the ordering of the alternatives we can evaluate the distance between them we talk about interval scale [3]. For these scales the distances between the alternatives have the meaning of real numbers. The transition from one interval scale to another is achieved with precision up to affine transformation:

$$x = ay + b,\ (x, y) \in \boldsymbol{X^2},\ a > 0,\ b \in \mathbf{R}$$

For example, if we say that a body is twice as warm as another in Celsius, this will not be true if the measurements were in Kelvin. The distance between them is always the same. Among these type of scales is the measurement of the utility function through the so called "*lottery approach*".

Let $\mathbf{X}$ be the set of alternatives ($\mathbf{X} \subseteq \mathbf{R}^m$). The DM's preferences over $\boldsymbol{X}$ are expressed by ($\succ$). Let $\mathbf{A_{u^*}}$ and $\mathbf{B_{u^*}}$ be the sets: $\mathbf{A_{u^*}} = \{(x,y) \in \mathbf{R}^{2m} / (u^*(x)) > u^*(y)\}$, $\mathbf{B_{u^*}} = \{(x, y) \in \mathbf{R}^{2m} / (u^*(x)) < u^*(y)\}$. If there is a function $\mathbf{F}(x,y)$ of the form $\mathbf{F}(x,y) = \boldsymbol{f}(x) - \boldsymbol{f}(y)$, positive over $\mathbf{A_{u^*}}$ and negative over $\mathbf{B_{u^*}}$, then the function $\boldsymbol{f}(x)$ is a value function equivalent to the empirical DM's value function u*(.). The construction of such functions of two variables is one of the ways for evaluating the value functions in ordinal aspect [5, 6]. Such approach needs stochastic "pattern recognition" for solving the problems. In the deterministic case it is true that $\mathbf{A_{u^*}} \cap \mathbf{B_{u^*}} = \varnothing$. In the probabilistic case it is true that $\mathbf{A_{u^*}} \cap \mathbf{B_{u^*}} \neq \varnothing$ [5]. Let $\boldsymbol{P}$ be a set of probability distributions over $\boldsymbol{X}$. A utility function $u(.)$ is any function for which the following is fulfilled: $(p \succ q, (p, q) \in \boldsymbol{P^2}) \Leftrightarrow (\int u(.)dp > \int u(.)dq)$.

In keeping with von Neumann and Morgenstern [4] the interpretation of the above formula is that the integral of the utility function $u(.)$ is a measure concerning the comparison of the probability distributions $p$ and $q$ defined over $\boldsymbol{X}$. The notation ($\succ$) expresses the

preferences of DM over $\boldsymbol{P}$ including those over $\boldsymbol{X}$ ($\boldsymbol{X} \subseteq \boldsymbol{P}$). There are different systems of mathematical axioms that give sufficient conditions of a utility function existence. The most famous of them is the system of von Neumann and Morgenstern's axioms [4]. The following propositions are almost obvious [4].

**Proposition 1.** The DM "*preference*" relation ($\succ$) defined by the utility function $u(.)$ is "*negatively transitive*" ( $\neg(p \succ t) \wedge \neg (t \succ q)$ ) $\Rightarrow \neg(p \succ q)$ ).

**Proposition 2.** In the case of a "*negative transitivity*" of ($\succ$) the "*indifferenc*e" relation ($\approx$) is transitive ( $((x \approx y) \wedge (y \approx t)) \Leftrightarrow (x \approx t)$ )).

**Corollary:** Let the relation ($\succ$) be "*irreflexive*" ($\neg(p \succ p)$) and "*negatively transitive*", then the "*indifference*" relation ($\approx$) is an "*equivalence*" (reflexive, symmetric and transitive).

We need next two definitions. A "*weak order*" is an asymmetric and "negatively transitive" relation. The transitivity of the "*weak order*" ($\succ$) follows from the "asymmetry" and the "negative transitivity". A "*strong order*" is a "*weak order*" for which is fulfilled $(\neg(x \approx y) \Rightarrow ((x \succ y) \vee (x \prec y))$. It is proven in [4] that the existence of a "*weak order*" ($\succ$) over $\boldsymbol{X}$ leads to the existence of a "*strong order*" ($\succ$) over the factorized set $\boldsymbol{X}/\approx$. *The presumption of existence of a utility function u(.) leads to the existence of: asymmetry* $(x \prec y) \Rightarrow (\neg(x \succ y))$, *transitivity* $(x \succ y) \wedge (y \succ z) \Rightarrow (x \succ z)$, *and transitivity of the "indifference"* ($\approx$).

So far we are in the ordering scale. The assumption of equivalence with precision up to affine transformation has not been included. In other words we have only a value function. We underline that the mathematical expectation is included in the definition of the utility function. To achieve this effect, measurements are made using the lottery (gambling) approach [2]. A "*lottery*" is called every discrete probability distribution over X. We mark as $<x, y, \alpha>$ the lottery: $\alpha$ is the probability of the appearance of the alternative $x$ and $(1-\alpha)$ – the probability of the alternative $y$. The most used evaluation approach is the following assessment: $z \approx <x, y, \alpha>$ , where $(x, y, z) \in \boldsymbol{X}^3$, $(x \succ z \succ y)$ and $\alpha \in [0,1]$. Weak points of this approach are violations of the transitivity of the relations and the so called "certainty effect" and "probability distortion" [7]. Additionally, the determination of alternatives $x$ (*the best*) and $y$ (*the worst*) on condition that $(x \succ z \succ y)$ where z is the analyzed alternative, is not easy. Therefore, the problem of utility function evaluation on the grounds of expert preferences is a topical one [7, 8]. The violations of the transitivity of the relation ($\approx$) also leads to declinations in the utility assessment. All these difficulties could explain the DM behaviour observed in the famous Allais Paradox that arises from the "*independence*" axiom [9].

The determination of a measurement scale of the utility function $u(.)$ originates from the previous mathematical formulation of the relations ($\succ$) and ($\approx$). It is accepted that ($\boldsymbol{X} \subseteq \boldsymbol{P}$) and that $\boldsymbol{P}$ is a convex set ($(q, p) \in \boldsymbol{P}^2 \Rightarrow (\alpha q + (1-\alpha)p) \in \boldsymbol{P}$, for $\forall \alpha \in [0,1]$). Then the utility function $u(.)$ over $\boldsymbol{X}$ is determined with the accuracy of an affine transformation [4]:

**Proposition 3**. If ($(x \in X \wedge p(x)=1) \Rightarrow p \in \boldsymbol{P}$) and ($((q, p) \in \boldsymbol{P}^2) \Rightarrow ((\alpha p + (1-\alpha)q) \in \boldsymbol{P}$, $\alpha \in [0,1]$)) are realized, then the utility function $u(.)$ is defined with precision up to an affine transformation $(u_1(.) \approx u_2(.)) \Leftrightarrow (u_1(.) = a u_2(.) + b, a>0)$.

Now we are in interval scale and here the mathematical expectation is feasible. That is to say, this is a utility function [3, 4]. This property is essential for the application of the utility theory, since it allows a decomposition of the multi attribute utility functions into simple functions [2]. Starting from the gambling approach we use the following sets motivated by Proposition 3:

$\boldsymbol{A}_{u^*} = \{(\alpha, x, y, z) / (\alpha u^*(x) + (1-\alpha) u^*(y)) > u^*(z)\}$, $\boldsymbol{B}_{u^*} = \{(\alpha, x, y, z) / (\alpha u^*(x) + (1-\alpha) u^*(y)) > u^*(z)\}$.

The notation $u^*(.)$ is the DM's empirical utility assessment. The approach we are using for the evaluation of the utility functions in its essence is the recognition of these sets. Through stochastic recurrent algorithms we approximate functions recognizing the above

two sets [6]. We propose the next stochastic approximation procedure for evaluation of the utility function. In correspondence with the Proposition 3 it is assumed that ($\boldsymbol{X}\subseteq\boldsymbol{P}$), $((q,p)\in\boldsymbol{P}^2\Rightarrow(\alpha q+(1-\alpha)p)\in\boldsymbol{P}$, for $\forall\alpha\in[0,1]$) and that the utility function $u(.)$ exists. The following proposition is in the foundation of the used stochastic approximation approach:

**Proposition 4.** We denote $\boldsymbol{A}_u=\{(\alpha, x, y, z)/(\alpha u(x) + (1-\alpha)u(y)) > u(z)\}$. If $\boldsymbol{A}_{u1} = \boldsymbol{A}_{u2}$ and $u_1(.)$ and $u_2(.)$ are continuous functions than is true ($u_1(.) = au_2(.) + b$, $a > 0$) [6].

The approximation of the utility function is constructed by pattern recognition of the set $\boldsymbol{A}_u$ [6]. The proposed assessment process is machine-learning based on the DM's preferences and is a probabilistic pattern recognition ($\boldsymbol{A}_{u*}\cap\boldsymbol{B}_{u*}\neq\varnothing$). The utility evaluation is a stochastic approximation with noise (uncertainty) elimination. The evaluation procedure is: The DM compares the "*lottery*" $\langle x, y, \alpha\rangle$ with the simple alternative $z$, $z\in Z$ ("*better* – $\succ$", $f(x,y,z,\alpha)=1$", "*worse*– $\prec$", $f(x,y,z,\alpha)=(-1)$" or "*can't answer or equivalent*– $\sim$, $f(x,y,z,\alpha)=0$", $f(.)$ denotes the qualitative DM's answer). This determine a *learning point* ($(x,y,z,\alpha)$, $f(x,y,z,\alpha)$). The following recurrent stochastic algorithm constructs the utility polynomial approximation:

$$u(x)=\sum_i c_i\Phi_i(x), \qquad c_i^{n+1}=c_i^n+\gamma_n[f(t^{n+1})-\overline{(c^n,\Psi(t^{n+1}))}]\Psi_i(t^{n+1}),$$

$$\sum_n\gamma_n=+\infty,\ \sum_n\gamma_n^2<+\infty,\ \forall n,\gamma_n>0.$$

In the formula are used the following notations (based on $\boldsymbol{A}_u$): $\boldsymbol{t}=(x, y, z, \alpha)$, $\boldsymbol{\psi}_i(t)=\boldsymbol{\psi}_i(x, y, z, \alpha)=\alpha\Phi_i(x)+(1-\alpha)\Phi_i(y)-\Phi_i(z)$, where $(\boldsymbol{\Phi}_i(x))$ is a family of polynomials. The line above the scalar product $\bar{v}=(c^n,\Psi(t))$ means: ($\bar{v}=1$), if ($v>1$), ($\bar{v}=-1$), if ($v<-1$) and ($\bar{v}=v$) if ($-1<v<1$). The coefficients $c_i^n$ take part in the polynomial presentation $g^n(x)=\sum_{i=1}^N c_i^n\Phi_i(x)$ and $(c^n,\Psi(t))=\alpha g^n(x)+(1-\alpha)g^n(y)-g^n(z)=G^n(x,y,z,\alpha)$ is a scalar product.

The learning points are set with a pseudo random sequence. The mathematical procedure describes the following assessment process: The expert relates intuitively the "learning point" $(x, y, z,\alpha)$ to the set $\boldsymbol{A}_{u*}$ with probability $D_1(x,y,z,\alpha)$ or to the set $\boldsymbol{B}_{u*}$ with probability $D_2(x,y,z,\alpha)$. The probabilities $D_1(x,y,z,\alpha)$ and $D_2(x,y,z,\alpha)$ are mathematical expectation of $f(.)$ over $\boldsymbol{A}_{u*}$ and $\boldsymbol{B}_{u*}$ respectively, ($D_1(x,y,z,\alpha)$ = $M(f/x,y,z,\alpha)$) if ($M(f/x,y,z,\alpha)>0$), ($D_2(x,y,z,\alpha)$ = (-)$M(f/x,y,z,\alpha)$) if ($M(f/x,y,z,\alpha)<0$). Let $D'(x,y,z,\alpha)$ be the random value: $D'(x,y,z,\alpha)$ = $D_1(x,y,z,\alpha)$ if ($M(f/x,y,z,\alpha)>0$); $D'(x,y,z,\alpha)=(-D_2(x,y,z,\alpha))$ if ($M(f/x,y,z,\alpha)<0$); $D'(x,y,z,\alpha)=0$ if ($M(f/x,y,z,\alpha)=0$). We approximate $D'(x,y,z,\alpha)$ by a function of the type: $g(x)=\sum_i c_i\Phi_i(x)$, where $G(x,y,z,\alpha)=(\alpha g(x)+(1-\alpha)g(y)-g(z))$. The coefficients $c_i^n$ take part in $G(x,y,z,\alpha)$:

$$G^n(x,y,z,\alpha)=(c^n,\Psi(t))=\alpha g^n(x)+(1-\alpha)g^n(y)-g^n(z) \qquad g^n(x)=\sum_{i=1}^N c_i^n\Phi_i(x)$$

The function $G^n(x,y,z,\alpha)$ is positive over $\boldsymbol{A}_{u*}$ and negative over $\boldsymbol{B}_{u*}$ depending on the degree of approximation of $D'(x,y,z,\alpha)$. The function $g^n(x)$ is the approximation of the utility $u(.)$ and is used the decomposition:

The procedure and its modifications are machine learning [10]. The computer is taught to have the same preferences as the DM and the evaluation is recurrent stochastic programming for value based decision making in the framework of the normative approach.

$$f(t^{n+1})=\left[D'(t^{n+1})+\xi^{n+1}\right]$$

## 3.Practical Implementation – Optimal Allocation of Resources in a Small Factory (Example)

We apply the utility theory for planning the product range depending on the various available materials in a small enterprise. Our main purpose is the choice of the best option for furniture production in order to achieve best market realization and profit. To solve this task an examination of the customer's preferences have to be preformed. This research determines the range of products according to market demand, which is represented as a utility function. An optimization of the production process could be made based on this information. The basis for the practical decision is co-implementation of the measurement and utility theories. The utility is used through the prescriptive approach.

Figure 1 shows the task structure and Table 1 and Table 2 give particulars concerning possible materials for particular assortment of furniture products.

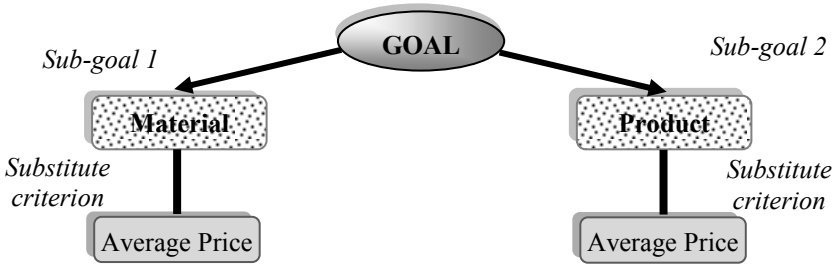



Fig. 1 Defining goal and sub-goals in the practical task

| Material | Average price (BGNs/m$^2$) |
|---|---|
| Chipboard | 50 |
| MDF | 150 |
| Pine | 200 |
| Beech | 400 |
| Oak | 600 |
| Walnut | 750 |

Table 1 Types of production materials

| Type | Product | Price (BGNs) | |
|---|---|---|---|
| | | Min | Max |
| 1 | Shelf | 30 | 70 |
| 2 | Chair | 40 | 130 |
| 3 | Table | 80 | 200 |
| 4 | Desk | 90 | 250 |
| 5 | Cupboard | 120 | 300 |
| 6 | Commode | 150 | 350 |
| 7 | Bed | 250 | 450 |

Table 2 Product range

It is established a priori an independence between the utility of the material and the type of the product. This interrelation defines the structure of the aggregated utility function:

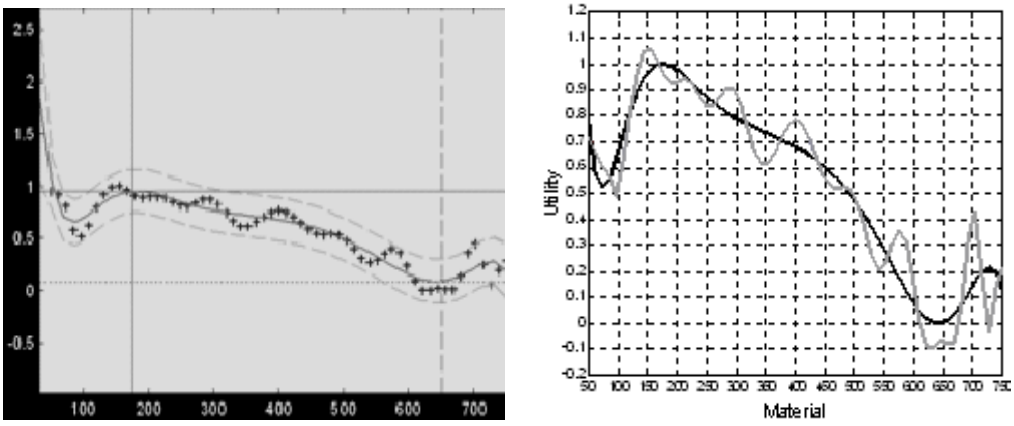
$$u(y, z) = u(y, z_0)\,[1 - u(y_0, z)] + u(y, z_1)\,u(y_0, z) \qquad (1),$$

where $y$ is the product type and $z$ is the material type. The following initial conditions are accepted: $u(y_0, z_0) = 0$, $u(y_1, z_1) = 1$ and the private function is normalized in the interval [0, 1] in accordance with the theory. The values of $z_0$ and $z_1$ are selected in order to satisfy $u(y_0, z_1) > u(y_0, z_0)$.

The semantic interpretations are determined as percentages allocated resources for a product. The appropriate attribute (model parameters) in this practical investigation are:

- Production material – not dependent on the product;
- Product range – dependent on the material.

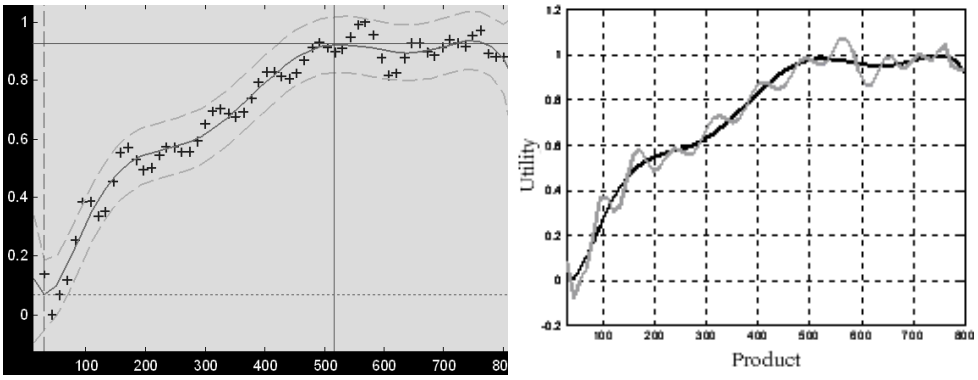
Based on the surveys about the customer's preferences we built three one-dimensional utility functions. Func1 $u(y_0, z)$ describes the DM's preferences regarding the used material (Fig. 2). It also determines the interval scale for all types of the material (price 50 – 750 BGNs/ $m^2$) for product type 1 ($y_0$ is a shelf).




a. minimum & maximum b. approximation

Fig. 2 General presentation of Func1

Func2 $u(y, z_0)$ reflects preferences regarding product at a fixed material (Fig. 3). The minimum of the preference in Func1 determines value of 650 BGNs for a product, i.e. the type of product on which we would least want to spend this amount of money. For unification of the measurement scales we use the fact that Func1 and Func2 share a common point – cheapest product using the cheapest material (30 BGNs for a shelf). As a second fixed point a point in the range between the minimum and maximum of Func2 is selected.

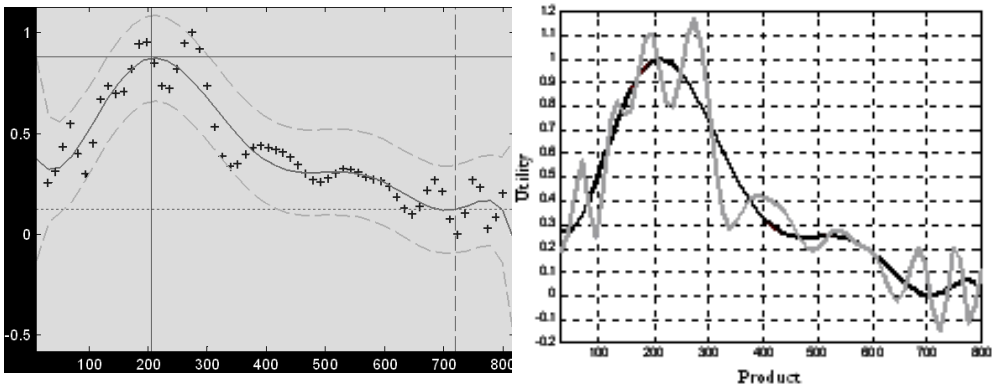



a. minimum & maximum b. approximation

Fig. 3 General presentation of Func2

Func3 $u(y, z_1)$ reflects preferences regarding a product at a fixed material (Fig. 4). The maximum of the preference in Func1 gives value 175 BGNs of a product, i.e. the product of the type we want most to give this amount for.

For unification of the measurement scales we cannot use shared common points of Func1 and Func3. So we are looking for a point where utilities are equivalent. As a second fixed point we select a point in the range between the minimum and maximum of Func3.




a. minimum & maximum b. approximation

Fig. 4 General presentation of Func3

Using the general formula (1) the surface corresponding to the main purpose is built. According to this aggregate utility function we can conclude that:

- preference is given to the type of product (i.e. what we need) rather than the material it is made of (1) ⇔ (2)
- the more expensive products (1) are preferably to be made of high quality materials (if the product is expensive, it should be at least of high quality!)
- less expensive materials are not rated (3)

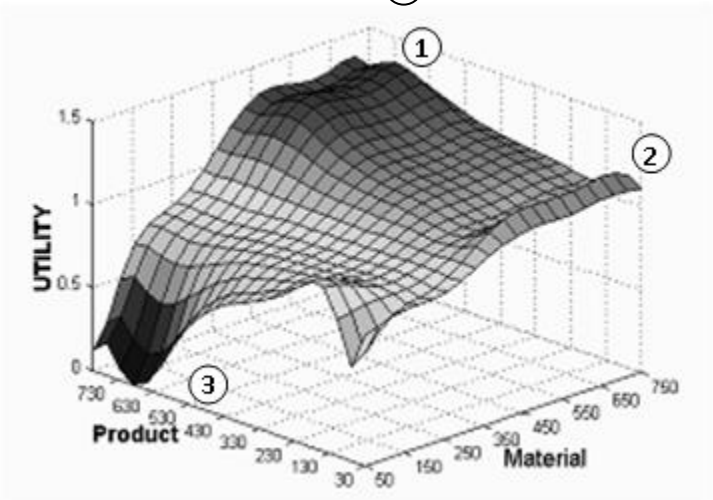



Fig. 5 Construction of the aggregate utility function corresponding to the main purpose

## 4.Conclusion

The paper presents mathematically sound approach to decision-making in complex tasks (resource allocation), where it is difficult to build a logically correct formal decision, considering the complexity of the problem, decision-makers objectives, as well his empirical knowledge level and professional behaviour. It is advantageous to evaluate human subjective expectations for the concrete task as subjective utility functions based on their own experience. Utility approach is used in situations with reflection of the human factor and is based on the Theory of measurement (Utility theory) and Stochastic approximation theory.

Methodology and algorithms for this step should be of an adequate level of generalization and ability to handle primary (marginable) conceptual information. Such an approach has to be mathematically sufficient to allow processing of conceptual information. The provided innovative method for determination of subjective expectations (probabilities) has many advantages, because it allows:

- Utility function evaluation that represents subjective expectations analytically;
- An mathematical expression of the utility of a type of manufactured product to be made;
- Probability distributions of practice or subjective preferences to be found;
- Precise analysis in terms of risk and uncertainty to be made.

Because of these reasons the described methodology for subjective preferences utility representation is to be promising for implementation in ERP systems supporting resource allocation and permitting economical prognosis in companies of all industries.

**ACKNOWLEDGMENT:**

The paper is partially supported by the project **№ BG161PO003-1.1.06-0086-C0001** in the program „Development of the competiveness of the Bulgarian economy 2007-2013"

## References

1. ERP Systems form the Information Management Backbone for Companies of all Sizes and in all Industries – http://www.aptean.com
2. Keeney, R., H. Raiffa: *Decisions with multiple objectives: preferences and value tradeoffs*, Cambridge Univ. Press, 1993.
3. Pfanzagl, J. *Theory of Measurement*. Physical-Verlag, Wurzburg-Wien, 1971.
4. Fishburn, P. *Utility Theory for Decision-Making*, New York, Wiley, 1970.
5. Pavlov Y. A Recurrent Algorithm for the Construction of Value Function. Comptes Rendus de L'Academie Bulgare des Sciences, vol. 42, № 7, p.p. 41-43, 1989. Zentralblatt fur Mathematik, vol. 688, September, section 90A06, N 90009, 1990.
6. Pavlov Y., Preferences Based Stochastic Value and Utility Function Evaluation, Proceedings of Informing Science & IT Education Conference (InSITE) 2011, InSITE 2011, Volume 11 , 2011.
7. Cohen, M., J-Y., Jaffray Certainty Effects versus Probability Distortion: An Experimental Analysis of Decision Making Under Risk, in Journal of Experimental Psychology: Human Perceptions and Performance, vol. 14 (4). pp. 554-560, 1988.
8. Kahneman, D., A. Tversky Prospect theory: An analysis of decision under risk, in: Econometrica, vol. 47, pp. 263-291, 1979.
9. Allais, M. Le Comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. Econometrica, V. 21, 503-546, 1953.
10. Aizerman, M. A., E. Braverman, L. Rozonoer *Potential Function Method in the Theory of Machine Learning*, Moscow, Nauka, 1970. (in Russian).

# Information Technology: Methods and Tools in Energy Saving Calculation for Building Renovation and Industry

**Hristo Dinkov[1], Nanko Bozukov[2], Marko Dimitrov[3]**

[1] UFT, 4002 Plovdiv, bul. Maritsa 26, Bulgaria
hristo.d@dir.bg

[2] UFT, 4002 Plovdiv, bul. Maritsa 26, Bulgaria
bozukovnanko@abv.bg

[3]ERC - Zenit Ltd, 1000 Sofia, bul. Bratya Bukston 40, Bulgaria
dimitrovmarko@abv.bg

**Abstract.** This paper discusses the algorithm of an energy efficiency evaluation program using developed and validated methodologies. Energy savings are calculated by measuring and/or estimating consumption before and after the implementation of an appropriate energy efficiency measure. The algorithms of two energy savings calculation sub-routines are presented: one for industry and one for the renovation of insulation and heating systems in residential and public buildings.

**Keywords.** information technology, energy efficiency

## 1.Introduction

Industrial facilities are the largest energy consumers. They are integrated systems comprising buildings and production technology processes for the realisation of products or services, hence they consume significant amounts of energy from the total energy requirements of the country. Their energy efficiency is achieved by optimising the production technology process and using energy management, intelligent power management systems and renewable energy sources.

Energy audits establish the energy consumption level, identify the reduction options and give recommendations regarding energy efficiency improvement [7]. Existing methodologies for energy saving evaluation use the principle of consumption measurement and/or estimation before and after the implementation of an energy efficiency measure. The objective of the methodologies used is to establish the implementation of individual measures or a set of measures aimed at improving the energy efficiency in residential and public buildings. [6] The methodologies developed cover the following areas:

-replacement of existing appliances with energy efficient equipment;

-equipment or building renovation;

-acquisition of energy efficient equipment or building

The difference in energy used before and after the implementation of the energy efficiency improvement measure or program is the amount of energy saved [6].

## 2.Algorithm for energy efficiency evaluation

An algorithm and a program for intelligent data processing in energy efficiency evaluation were developed (Figure 1).

After the program was started, the number (1, 2, 3 ... 12) of the energy savings calculation methodology used was entered. Each methodology corresponded to a sub-routine

which was executed and the results of its operation were displayed. Upon completion of the assignment, the program was used for displaying the final results. Each sub-routine could be used separately, as an individual program.

## 3.Algorithm for energy savings in industry

An algorithm was developed for the calculation of the energy savings estimate in industry. The algorithm was described by means of a flowchart. A program was created using the MatLab programming language for scientific and engineering calculations [1, 2].

MatLab is open, flexible and easily adaptable to the solution of a user's particular problem. It is based on a matrix representation of the data. MatLab is applied to a large number of scientific areas since it contains tools which include data processing and operations on matrices, vectors, polynomials, complex numbers, signal analysis, digital filtering, two- and three-dimensional graphics, etc. [3, 8].

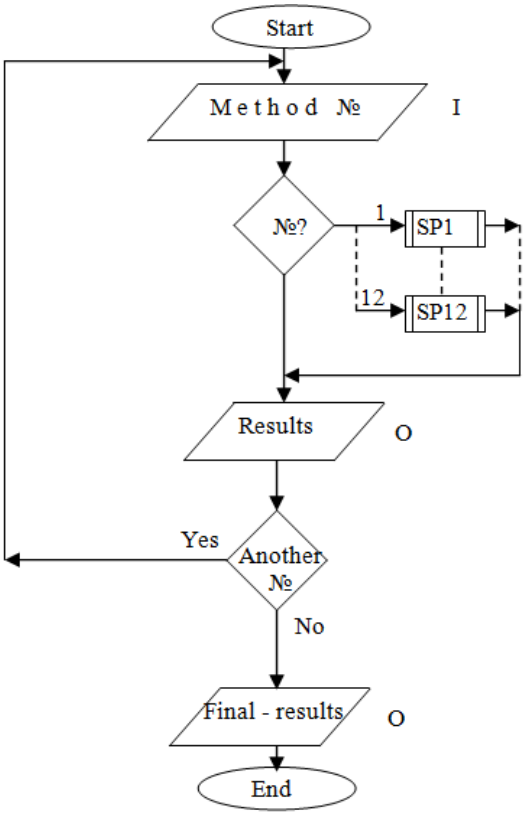



**Fig. 1.** Data processing program flowchart, I-input, O-output

The algorithm of the data processing program for energy savings in industry was presented using a flowchart (Fig. 2).

The parameters introduced were:

SECIC: specific energy consumption per product unit before the implementation of energy saving measures, kWh / product unit / year;

SECIH specific energy consumption per product unit after the implementation of energy saving measures, kWh / product unit / year;

W: production output in the respective year (t, m2, number, etc.);

The amount of energy saved in the manufacture of a particular product in industry for one year (FES [kWh / year]) was calculated and displayed.

The program can be re-executed or terminated.

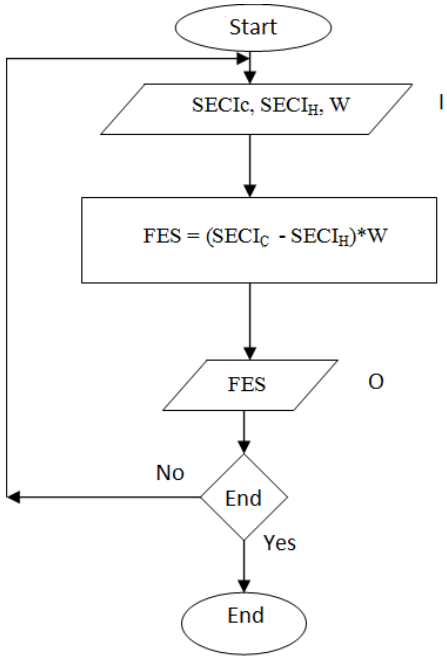



**Fig. 2.** Industrial energy saving program flowchart: I-input, O-output

After the program was started, the steps described in its algorithm were performed. A methodology for energy savings in industry was used. Upon completion of the assignment, the program displayed the final result, which was the amount of energy saved in the manufacture of a particular product for one year, using the following formula:

$$FES = (SECI_C - SECI_H) * W, kWh / year \quad (1)$$

The whole session, all variables and data used in the program were stored in a separate file that could be opened, viewed and printed.

## 4.Data processing algorithm for the renovation of insulation and heating systems in existing residential and public buildings

The algorithm of the data processing sub-routine for the renovation of insulation and heating systems in existing residential and public buildings was described using the flowchart presented in Figure 3. The parameters introduced were [6]:

FPЗП: gross floor area of the building (or part of the building subject to renovation) [m2];

SHDC: specific annual heat consumed prior to renovation [kWh/m2/year]

SHDH: specific annual heat consumed after renovation [kWh/m2/year]

ηC: efficiency of the heating system prior to renovation (seasonal);

ηH: efficiency of the heating system after renovation (seasonal);

The amount of heating energy saved per year (FES [kWh / year]) was calculated and displayed.

If the sub-routine is to be re-executed, it has to be started from the beginning; if it has to be terminated, the user needs to return to the main program.

## 8. Advantages of the algorithm for energy saving evaluation

The following advantages and disadvantages can be associated with the information technology for energy saving evaluation and its algorithm:

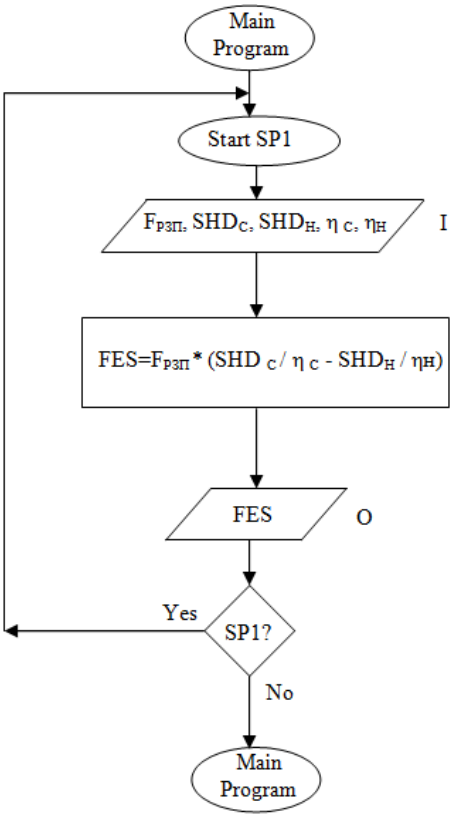



**Fig. 3.** Flowchart of the sub-routine for processing data when updating insulation and heating systems of existing residential and public buildings

- the accumulation of a sufficiently large database can result in the evaluation of the real energy saving effect in industry with considerable accuracy;
- using the database information, a complete analysis can be made, and the projected energy characteristics can be compared to those achieved during actual operation;
- the energy efficiency certificate issued within the period stipulated by law would be supported by reliable factual data;
- this information technology allows for quick and easy processing and calculation of the data needed for the application of the methodology developed to demonstrate the

implementation of individual measures or a set of measures aimed at enhancing energy savings in industry.

## 5.Conclusion

The information technology for data processing, evaluation and calculation of energy savings is an accurate and convenient means of control and objective analysis of the effect of energy efficiency measures applied in industry. This energy saving calculation information technology can ensure high reliability, easy data transfer and processing.

The possibility of collecting and processing available data is essential for the precise and fair assessment of energy savings according to the methodologies developed. It is therefore of utmost importance to have accurate information and data prior to any assessment Information technology should be used in each energy saving calculation methodology. Energy saving systems focus on the efficient energy use through innovative technologies. This will guarantee the adaptability, functionality, aesthetics, ecological compliance and energy efficiency of modern industry.

## References

1. Дьяконов В.П., MATLAB 6: учебный курс, Санкт Петербург, Питер, 2001.
2. Чен Ч., П Джиблинг, А. Ирвинг, MATLAB в математических исследованиях, Москва, Мир, 2001.
3. http://www1.bpo.bg/images/stories/buletini/ binder-2012-09.pdf
4. http://www.eulaw.egov.bg/DocumentDisplay.aspx?ID =382929
5. http://www.climamarket.bg/
6. http://www.mi.government.bg/files/useruploads/files/microsoft_word-metodiki.pdf
7. Microsoft Office 2007, Софт Прес, 2007
8. www.mathworks.com

# Information Approach to Computer Based Piece Wise Pressure Response of Electric Sphygmomanometer

**Lubomir H. Lahtchev**

Institute of Systems Engineering and Robotics, Bulgarian Academy of Sciences,
No 2, Bl. 2, Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
E-mail: lahchev@iser.bas.bg

**Abstract**. The duo sensor control concept applied in the electric sphygmomanometer of Hartmann accounts information of wave oscillation sensor and tension sensor. The computer based exploration of the piece wise blood pressure (BP) response of the tension sensor clarifies the corresponding process of measurement and its peculiarities. The information approach consists of program processing the data about real piece wise signal of the tension sensor and computation the results, obtained with computer based system from the sphygmomanometer. The results about computer based processing of a group of signal data files demonstrate the high computation quality of modelling the measurement knowledge about tension sensor in particular and the sphygmomanometer as whole.

**Key words:** sphygmomanometer, pressure, response, computer, signal.

## 1.Introduction

The tension sensor about slow blood pressure (BP) processes in the human body is implemented in the electric sphygmomanometer of Hartmann [1]. When the display of the sphygmomanometer shows the air pressure within the cuff at every time moment, it is result of accumulation the information of data, collected from the tension sensor. The algorithm of information processing of tension sensor signal is focus of the paper.

Examples of statistic processing the information about BP and its analysis from different points of view are described in several author's publications [2, 3, 4].

The actuality of the explorations of such scientific topics is arisen in the corresponding scientific books. "The fundamental problem of comprehensive theory of automatic control bears a number of important problems, tasks and methods of their solution. Before everything the "optimization in large" must be concerned and its work in real time scale requires full use of available a priori information as model of the controlled process or of object, given in one or other form." [5]. Hence, the discovering algorithmic peculiarities of the BP measurement process in the sphygmomanometer represents a special scientific domain and is a step toward to appointed trend.

The aim of the paper consists of exploration the piece wise blood pressure response from tension sensor of sphygmomanometer with use of information approach to algorithmic data processing at known graphics, built on analogue to digital converter, connected to a personal computer. The scheme, example of which is shown in [4], consists of electric sphygmomanometer, preliminary analogue convertor, analogue to digital convertor (mode 500 Hz) and personal computer and it corresponds to the general scheme of digital measurement unit, described in [6]. In the terms of monograph [6], it belongs to time to discrete quitting and voltage to discrete quitting multichannel, multiprocessor system.

## 2.Information Approach to Piece Wise Pressure Response of Electric Sphygmomanometer

The tension sensor of Hartmann's sphygmomanometer [1] reacts on the cuff air pressure since the moment the air pump starts work. The linear piece wise graphics of the signal, obtained on the personal computer, is shown on fig. 1. It contains several linear sections, corresponding to different stages of measurement: air charge (segment AB), air dynamic shrink (piece wise BCD), self control air discharge (segment DE), systolic BP range and beginning controlled air discharge (p.E), diastolic BP range (p.F), auto stop the measurement (segment FG) and range of static sensor level (p.G).

From the general piece wise graphics is seen, that stages are divided by extremes, and the pressure signal brings information about the local velocity of the process during one time discrete:

$$\Delta x/\ \Delta t\ = (x_n - x_{n-1}) \qquad (1)$$

Moreover, the stages are clean defined by time domains. Hence, settings about levels of the signal and settings about corresponding to it local time domains will define the points of the extremes:

$$S_{ei} = L_i \cap T_i \qquad (2)$$

where $S_{ei}$ – multiplied range of the i-th signal extreme, characterized by function value and time moment; $L_i$ – range of function levels of the i-th signal extreme, given by upper and

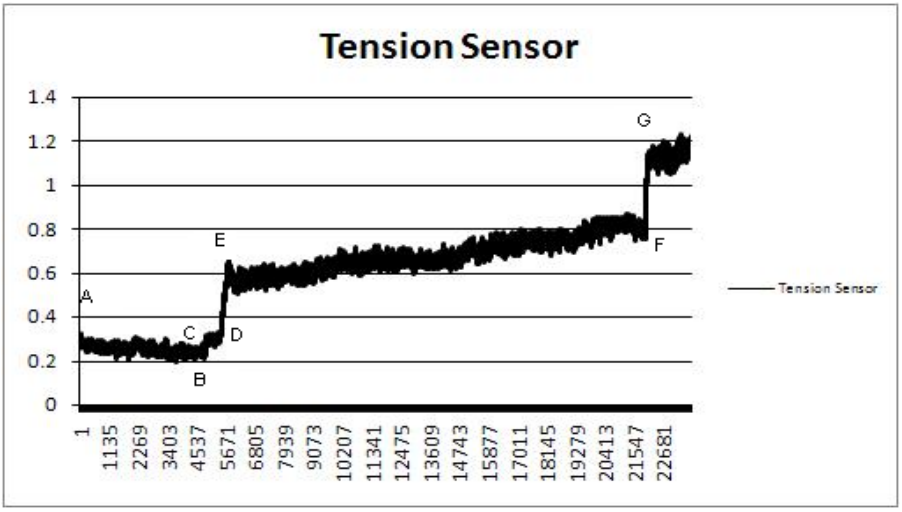



Figure 1. Piece wise graphics of the tension sensor signal about BP.

lower values; $T_i$ – range of time domain of the i-th signal extreme, defined by left and right boundaries; $i \geq 1, 2, \ldots 7$. The extreme level value can be recognized through the inequality:

$$\Delta x\ /\ \Delta t\ \geq \delta_{pr} \qquad (3)$$

where threshold $\delta_{pr}$ = int[(Round($x_0$ . 10), 0)+1] / 100 ; $x_0$ – beginning argument drop value provides information about level of the signal at working air pump. When the signal record

begins at non-working air pump, then beginning argument is on the preliminary sensor static level line. In this case, the first drop value after the pump begins work will represent the $x_0$ in the above corresponding equation.

The character of the piece wise pressure line of the sphygmomanometer witness about inverse process, fulfilled with drop vibrations. Such a signal may have different static levels $L_i$ at different measurement processes, but general shape of the signal during the process airing – measurement – air discharge remains the same. It can be represented by the following discrete function:

$$\neg\mathbf{x} = (1/N) \cdot \sum_{n=1}^{N} (L_s - x_n) = L_s - (1/N) \cdot \sum_{n=1}^{N} x_n \qquad (4)$$

$x_n$ – function discrete value at measurement during time discrete *n* of the analogue to digital converter and $0 < x_n < L_s$. In order to receive real results about pressure the inverse inputs must be kept during the whole process of measurement. After the final stop at balanced air discharge the trial static level line is also computed. The algorithm for the piece wise BP function is shown on fig. 2.

The computation the parameters of stages about linear process of tension sensor data provide numeric description of the corresponding piece wise line of the BP – function. Hence, the ranges of the linear BP function, where the systolic and diastolic BP are located, can be early addressed by ranges of points, defined of eq. 2 on the graphics of fig. 1. This is a task of the "Program settings" – block. The corresponding arrays are preliminary calculated, but a cycle of piece wise extreme procedure reveals all existing extreme points, and program settings can be verified, when the program does not produce results.

The "Read" – block contains a brief program procedure for computation the $N_{pr}$ – number of data about tension sensor signal. Test of condition, described by eq. 3, often locates the character point of next extreme. Whole list of extremes is gradually recorded by the "Record" – block.

The end trial static level lies on the line segment after p. G of the graphics. It is expressed from the tail segment of the high frequency drop amplitude process of the tension sensor, when the cuff becomes empty after auto stop measurement of the BP. The computer program accounts the number K of local extreme drop waves:

$$L_s = 1/K \cdot \sum_{n=1}^{K} x_{Fn} \qquad (5)$$

Hence the inverse inputs of eq. 4 will suppose computation of the addition of a function value to the end trial static level, which corresponds to full charge power supply of sensor at rest. The procedure, which computes the systolic and diastolic BP has the algorithm block-scheme in fig. 3. The block of "Program settings" here concerns pressure constant C, corresponding extreme of p. B (see fig. 1), and the range of static level. The program cycle of blocks about additions and character extremes **x** for systolic and diastolic ranges repeats the same one of the algorithm in fig. 2, but with already known key values. For systolic BP the key value is average of extremes at p. D and p. E of fig. 1. On the other side, it can be mean of range. For the diastolic BP key value is extreme of p. F and its preliminary range till 0.3 s.

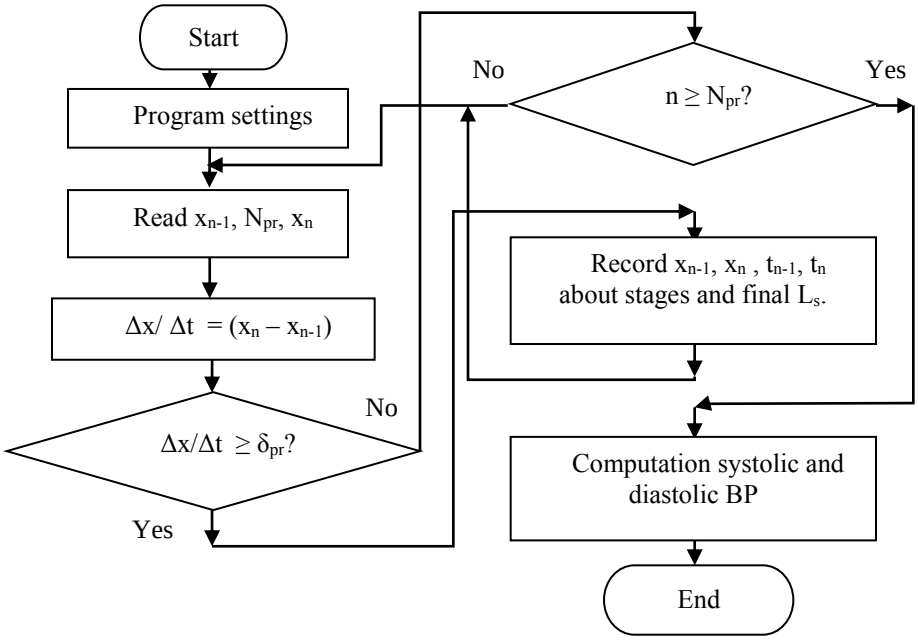



Figure 2. Algorithm block-scheme of recovering the piece wise pressure line.

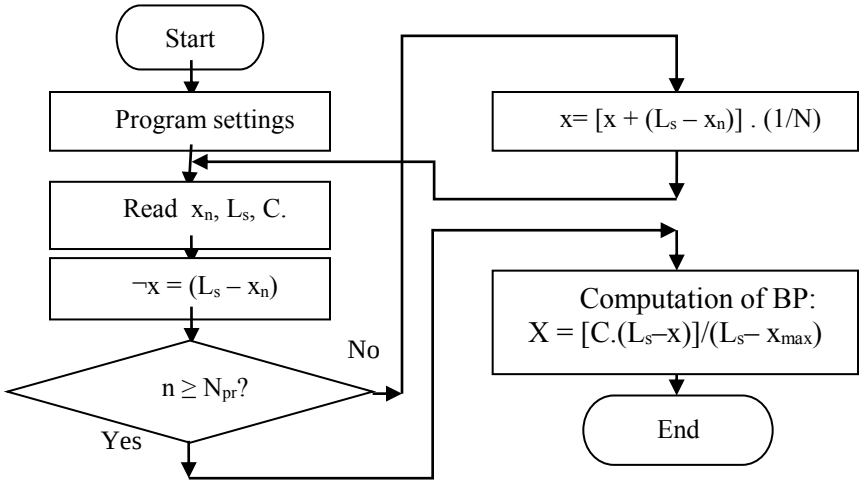



Figure 3. Algorithm block-scheme of the procedure for computation of the BP.

The equation of the BP is:

$$X = [C . (L_s - x)] / (L_s - x_{max}) \quad (6)$$

where C is the measured air pressure at moment $x_{max}$ at stop the air pump or the measurement constant in mmHg. If the computation of the diastolic BP does not produce adequacy result, then an amendment is applied.

## 3.Computed and Measured Results of Blood Pressure

The piece wise process line, example of which is shown in fig. 1, is described by list of points, represented in table 1. The rows of table 1 set points of the graphics and results of computations, accomplished on their foundation. The columns define numeric description of a file or of a BP measurement. For every one line segment separated program cycle completes the computations of function points, whose ranges of existence are defined, preliminary. Another program segment computes the results about BP also with use of preliminary program set and with some amendment condition. It is seen that all expected moments of systolic BP lie in the time range near the extreme of p. E (row 5 of table 1). Thus the most important piece wise transition corresponds to the systolic BP. On the other hand, the time range of the diastolic BP lies on the end extreme at p. F after controlled air discharge and shortly before it. Often it is average of the start and end values of balanced air discharge. In order to assess the quality of computations a pair scaled inverse functions about ranges of BP of expression:

$$\mathbf{x}_{cSD} = C \,.\, (L_s - \mathbf{x}_{nSD}) \qquad (7)$$

is introduced. In the eq. 7 the $\mathbf{x}_{nSD}$ – key values of ranges about systolic BP (row 7 of table 1) diastolic BP (row 9 of table 1) by measurements, $L_s$ – static levels of measurements, C – scale pressure of measurements.

The pairs of graphics in fig. 4 have different line styles and colors and are ordered from up to down as follows: systolic BP; range of systolic BP; diastolic BP; range of diastolic BP; relative deviation of systolic BP; relative deviation of diastolic BP. The double closely situated positive or negative peaks of BP and its range must be met, when a non-compensative deviation exists. There are not such phenomena on the graphics.

The direct comparison of the computed and measured results can be executed by their relative deviations:

$$\mathbf{E} = [(\mathbf{X_c} - \mathbf{X_m})/\ \mathbf{X_c}] \,.\, 100\ \% \qquad (8)$$

represented in table 2. The systolic BP and its range on the measurement No 6 have some similarity to above rule, but the peaks are distanced and deviation is small unlike the same of measurement No 7. The most important relative deviations of diastolic BP, seen in table 2, of measurement No 1, No 6, No 10 do not also meet contain of the above rule despite of the important values of relative deviations. Hence, the deviations of the computed and measured BP bear an accident nature and they can be compensated.

The probability of work of hardware module of Hartmann's sphygmomanometer has one physical nature, and the probability of work of the computer program, which computes the results from the sensor signal – another one physical and computation background. From one side, every one of those probabilities is a source of inaccuracy. From another side, the results of program solution and hardware solution can overlap, what happens at several measurements. The corresponding conditions of such event would be better explored about physic properties of the signal sources.

Table 1. Numerical descriptions of several measurements the BP.

| Parameter | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| | Function / time values | Function / time values | Function / time values | Function / time values | Function / time values | Function / time values |
| First value of array | 0.264/0 | 0.391 / 0 | 0.154 / 0 | 0.164/0 | 0.188/0 | 0.159/0 |
| Maximal aired cuff | 0.239/ 9.834 | 0.386/ 9.356 | 0.115/ 9.756 | 0.154/ 9.96 | 0.181/ 8.456 | 0.156/ 7.39 |
| Start of cuff elastic discharge | 0.291 / 9.836 | 0.442/ 9.358 | 0.149/ 9.758 | 0.186/ 9.962 | 0.22/ 8.458 | 0.186/ 8.392 |
| End of cuff elastic discharge | 0.278/ 10.824 | 0.444/ 10.482 | 0.137/ 10.738 | 0.176/ 10.356 | 0.207/ 9.53 | 0.222/ 8.488 |
| Self controlled air discharge | 0.652/ 11.622 | 0.876/ 11.392 | 0.344/ 12.698 | 0.383/ 11.146 | 0.491/ 10.99 | 0.397/ 9.656 |
| Start of controlled air discharge | 0.634/ 11.624 | 0.847/ 11.394 | 0.337/ 12.74 | 0.378/ 11.148 | 0.478/ 0.992 | 0.386/ 9.658 |
| Range of systolic BP, [V] | 0.544/ 12.122 | 0.831/ 11.892 | 0.337/ 13.198 | 0.374/ 11.646 | 0.461/ 11.49 | 0.388/ 10.156 |
| End of control-led air discharge | 0.769/ 44.052 | 1.023/ 38.238 | 0.452/ 41.532 | 0.527/ 36.46 | 0.644/ 41.588 | 0.872/ 43.296 |
| Range of diastolic BP,[V] | 0.803/ 43.794 | 1.038/ 37.94 | 0.452/ 41.234 | 0.478/ 36.162 | 0.669/ 41.29 | 0.627/ 42.998 |
| Static level | 1.094 | 1.322 | 0.648 | 0.775 | 0.912 | 1.047 |
| Drop extremes of static level | 40 | 85 | 20 | 36 | 31 | 162 |
| Systolic BP | 132 | 129 | 121 | 118 | 123 | 135 |
| Diastolic BP | 88 | 75 | 75 | 68 | 76 | 75 |
| Scale pressure | 179 | 182 | 158 | 150 | 160 | 159 |
| Expected moments of BP | 10.824/ 43.794 | 10.482/ 37.94 | 10.738/ 41.234 | 10.356/ 36.162 | 9.53/ 41.29 | 8.488/ 42.998 |
| Duration, [s] | 34.258 | 28.882 | 31.776 | 26.5 | 33.132 | 34.866 |

The rows of the table 2 correspond to a series of a measurement and its results. The columns contain the computed and measured results for BP, the corresponding relative deviations and computed static sensor level with measured maximal air pressure off the cuff at working air pump. The table 2 reveals the dynamics of the relative deviation about BP.

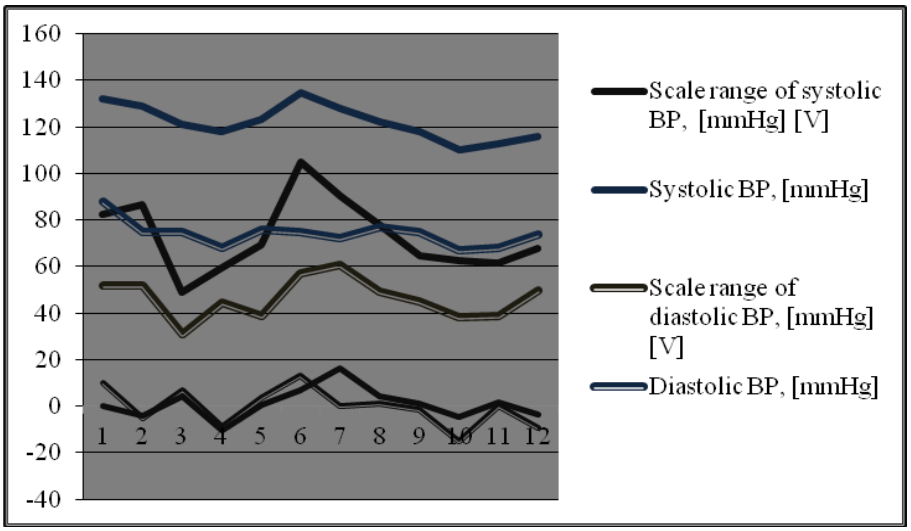



Figure 4. Relations of functions about systolic and diastolic BP, their corresponding ranges, and relative deviations.

Table 2. Comparative analysis of computed and measured results of BP.

| No file | Computed systolic BP, [mmHg] | Computed diastolic BP, [mmHg] | Measured systolic BP, [mmHg] | Measured diastolic BP, [mmHg] | Relative deviation of systolic BP, % | Relative deviation of diastolic BP, % | Static level of the sensor, [V] | Scale value, [mmHg] |
|---|---|---|---|---|---|---|---|---|
| 1 | 132 | 88 | 132 | 79 | 0.000 | 10.227 | 1.094 | 179 |
| 2 | 129 | 75 | 134 | 79 | -3.876 | -5.333 | 1.322 | 182 |
| 3 | 121 | 75 | 116 | 70 | 4.132 | 6.667 | 0.648 | 158 |
| 4 | 118 | 68 | 130 | 74 | -10.169 | -8.824 | 0.775 | 150 |
| 5 | 123 | 76 | 122 | 73 | 0.813 | 3.947 | 0.912 | 160 |
| 6 | 135 | 75 | 126 | 65 | 6.667 | 13.333 | 1.047 | 159 |
| 7 | 128 | 72 | 107 | 72 | 16.406 | 0.000 | 0.999 | 159 |
| 8 | 122 | 77 | 117 | 76 | 4.098 | 1.299 | 1.051 | 158 |
| 9 | 118 | 75 | 117 | 76 | 0.847 | -1.333 | 0.930 | 162 |
| 10 | 110 | 67 | 115 | 77 | -4.545 | -14.925 | 0.940 | 150 |
| 11 | 113 | 68 | 111 | 68 | 1.770 | 0.000 | 1.001 | 150 |
| 12 | 116 | 74 | 120 | 81 | -3.448 | -9.459 | 1.037 | 159 |

The largest ones are at measurement No 6 for the diastolic BP, at measurement No 7 for the systolic BP, and at measurement No 10, already pointed above. The important relative deviation at measurement No 1 for the diastolic BP has the same background. These relative deviations are permissible, less than 20%. From one side of view, obviously, the more accurate computation of a computer program depends on the refinement of the program

settings, and especially of the program procedures and their dynamics. This phenomenon declaratively describes the difficulties, the program engineer is faced. On the other side, the information approach with use of computer programs at data processing of sphygmomanometer tension sensor signal demonstrates the advantages of computer based exploration and of the corresponding high computation quality of modeling the measurement knowledge. As whole, the results also discover a positive assessment of the electric sphygmomanometer, and the scheme about it exploration.

## 4.Conclusion

The information approach to computer based piece wise pressure response is built on comparatively simple program procedures. The exploration of data readings from tension sensor requires exceeded attention at measurements with electric sphygmomanometer, special electric scheme about inputs to computer and powerful software. The piece wise graphics of the tension sensor function recognizes specific ranges, defined by points of separation. These points are revealed and highlighted with an algorithm. The relative deviations of the computed and measured results, which are negligible for the majority of measurements, sometimes increase over 10%. The refinement of program settings and program procedures improve the piece wise pressure response of the sphygmomanometer. As whole, the information approach to computer based exploration discovers a positive assessment of the electric sphygmomanometer and the scheme about it. The data processing and results of program procedures are produced on Excel.

## References

1. P. Hartmann. *Tensoval Duo Control. Instructions for use.*
2. L. Lahtchev *Cloud Computing Individual Assessment of a Medical Device for the Education.* Conf. Proc. of the Intern. Conf. on Appl. of Inf. and Comm. Tech. in Economy and Education, December – 2011, UNWE, Sofia, Bulgaria, pp. 503 – 512.
3. L. Lahtchev *Individual Assessment of a Medical Device for the Education by Relative Features.* Proc. of the Intern. Conf. on Application of the Information and Communication Technology in Economy and Education, December 2011, UNWE, Sofia, Bulgaria, pp. 281 – 289.
4. L. Lahtchev. *Cloud Computing Statistics of Individual Alternative Tests with a Sphygmomanometer*. *Conf.* Proc. of the Intern Conf. on Application of Inf. and Comm. Tech. and Statistics in Economy and Education. October 2012, UNWE, Sofia, Bulgaria, pp. 322-330.
5. *A Reference Book on the Automatic Control Theory*. Under Editorship of A.A. Krassovskii. Moscow, Science, 1987, 712 p. In Russian.
6. N.N. Vostroknutov. *Digital Measurement Units. Theory of Erroneousness, Tests, and Verification.* Moscow, Energy Atom Publishing, 1990, 208 p. In Russian.

# Implementation of Contemporary Technologies in Virtualization and Construction of an Information Cloud of Systems for University Needs in the Field of Transport Education

**Dimitar Dimitrov**

Todor Kableshkov University of Transport, 1574 Sofia, Bulgaria
ddimitrov@vtu.bg

**Abstract.** Current publication treats issues related to the implementation of contemporary technologies in virtualization and construction of an information cloud of systems. A technological problem solution, used in organizing and carrying out of university transport education, is presented, based on up-to-date tendencies. The paper presents conceptions and model of the general packet of information systems, technical security provision as well as organization initiatives concerning system development and expansion. Sustainability system performance analysis has been carried out.

**Keywords.** Virtualization, Cloud, Computing, Virtual Machine, Cloud Services, Hardware and Software Infrastructure, University, Education.

## 1.Introduction

Contemporary model of information services provision is related to these services granting on subscription. This model provides services to clients in accordance with their requirements and needs without their knowing of service hosting location and the way it reaches the client. This issue is also related to the dropping of the necessity for the clients to build a complicated hardware and software construction for their own information servicing, as well as with the issues concerning protection and archiving of data related to their work.

All these peculiarities have altered the paradigm of information services delivery and use. Hardware computing power is achieved through the construction of huge centers containing computer clusters in compliance with the so-called *Grid computing* technology, commonly referred to as *cloud technology* or simply as *Cloud*. This practically allows the users access to information services on subscription from each world location through users' access devices (computers, mobile devices, etc.), computing procedure performance remaining in the cloud. Powerful hardware resources allow separate operational systems installation, on the principle of sharing of resources, and formation of the so-called *Virtual Machines* (VMs).

Information resource centers construction is a paramount importance task, which requires serious capital investments and fast means liquidity in compliance with economic life of computer equipment. It is advisable to build the so-called *Academic Cloud*, to be used in a not really commercial manner to serve academic clients (students, lecturers and scientists), in the sphere of education and science, which is not typically commercialized, and has its own particular features as to the implemented software.

Current publication presents the approach and principles of the construction of a base element pertaining to the common academic information cloud of information resources at the Todor Kableshkov University of Transport. Principles of its planning and construction

are also commented; information structure achievement approach is likewise presented and analyzed.

## 2.General Conception and Principles

Virtualization basic purpose is to accomplish the model of many-platform work environment of operation systems and applications, located on a given server or a group of servers to efficiently use integrated information power and computer resources. This model is usually achieved through the construction of Data Centers, equipped with high capacity internet connectedness and providing system functioning under specific conditions (temperature, dust, etc.), where computer resources are located.

Depending on the type and amount of used hardware, construction of virtualization information structures can be achieved in two options, as shown on fig.1 and fig. 2.

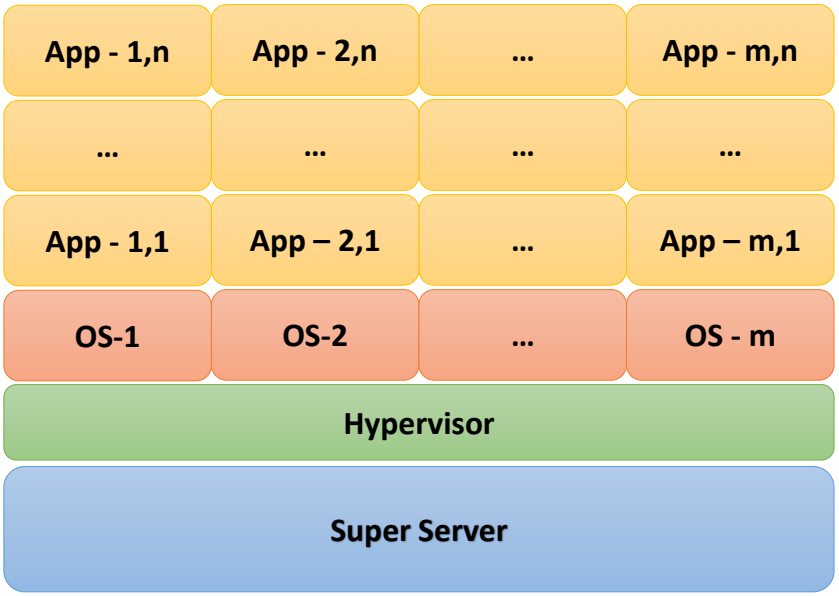



Fig. 1 Typical scheme of virtualization while using one hardware unit (Super Server)

The first option requires the provision of a super powerful server, equipped with the capacity to assume separate operational systems as well as their working applications. It is primarily related to server resources shared use, such as processor power, used memory and disc space. Static and shared resources use is also possible (processor power, memory, disc space). It all happens through the so-called software layer *Hypervisor*, representing the main work environment by means of which virtual operation systems are installed. In some solutions *Hypervisor* functions as a superstructure layer of an installed base operation system.

Individual virtual operation systems possess characteristics and functionality of ordinary operation systems, communicating through their virtual hardware.

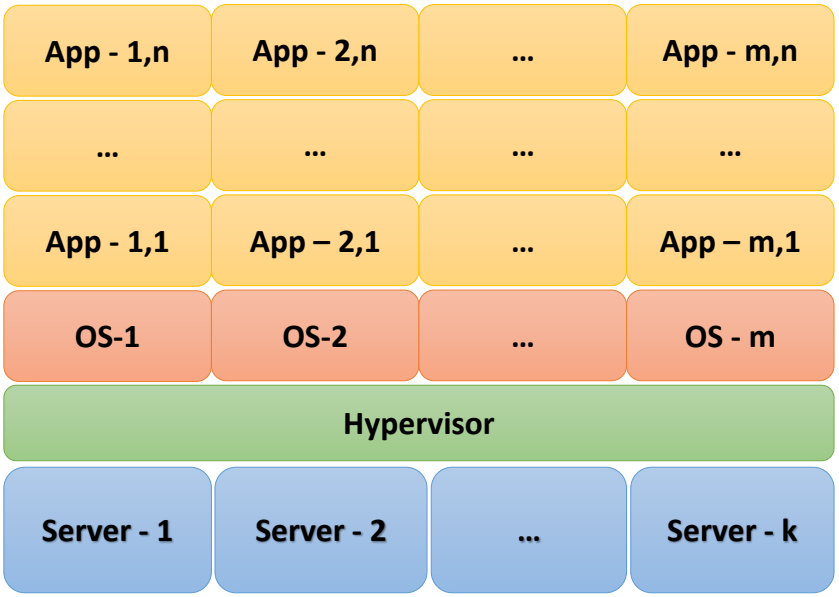



Fig. 2 Typical scheme of virtualization while using many hardware units (Multi Servers)

Figure 2 illustrates an alternative way of solving virtualization problem, though through the use of a greater number of hardware resources (servers). This model is also known as work in a cluster, that is, separate servers functioning in a parallel mode and representing an enlarged server unit. This solution advantage consists in the opportunity to cover a server failure, as well as in a dynamic change or rather hardware resource expansion.

There is a variety of works supporting the achievement of these virtualization models and allowing the construction of big centres with huge numbers of servers, operation systems and applications.

## 3.Model for the development of information systems and services

The common model for information services development is based on the following linguistic mathematical description:

$$iCloud_i = f\{iComp_i, iOS_i\lceil iApps_{ij}\rceil\}, \quad (1)$$

where,
$iCloud_i$ – information cloud *i*;
$iComp_i$ – information hardware in cloud *i*;
$iOS_i\lceil iApps_{ij}\rceil$ – operation systems and applications *j* functioning in cloud *i*;

Common capacity of cloud *i* is determined and dimensioned through:

$$CAPACITY(iCloud_i) = \sum iOS_i\lceil iApps_{ij}\rceil \leq \sum iComp_i, \quad (2)$$

Common computer resource capacity in the cloud must be greater than or equal to the needed capacity used by virtual machines.

It is also necessary to achieve the so-called nominal cloud loading, that is server platform (cluster) functioning without resources overloading, particularly when a server cluster model is concerned and a cluster machine fails. This fact is planned based on the following inequality:

$$RESOURCES(iCloud_i) = iComp_i^{(HDD,CPU,MEM)} > iOS_i\left\lceil iApps_{ij}\right\rceil^{(HDD,CPU,MEM)} \quad (3)$$

The last formula allows us to calculate the cloud common resources and can be presented by an inequality between present hardware resources and used resources in virtual operation systems and applications, that is, in processor power, memory and disc space use. Nominal loading in practice is considered use of resources in the range of 70-80%.

## 4.Presentation of information structure in the process of building

Information structure building on the territory of the Todor Kableshkov University of Transport [1, 2, 3] is characterized by the following peculiarities and requirements:

- Hardware infrastructure building and development.
- Server systems virtualization.
- Information systems and services development.
- Accomplishment of a business-oriented model of work and use of systems.
- Solution of problems related to the security of systems and data.

Figure 3 shows the general scheme of information structure in process of building.

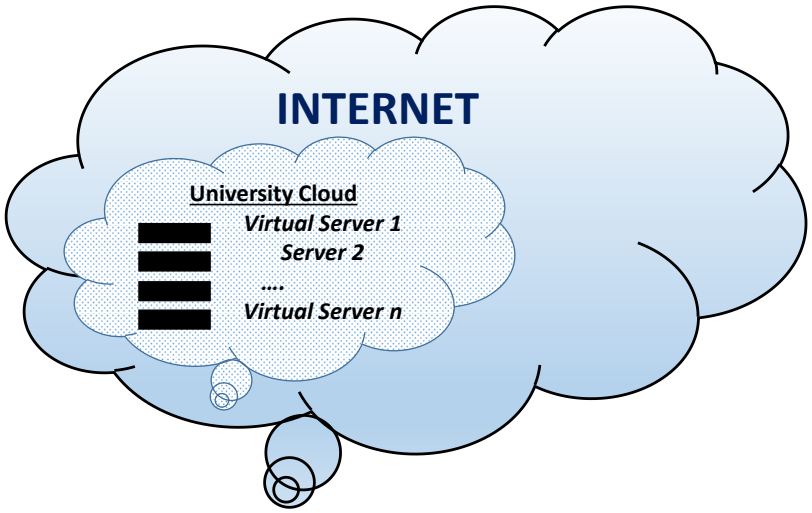



Fig. 3 General scheme of presented university information cloud as part of the Internet web

Thus information structure in the process of building can be classified as information cloud with relatively small dimensions. This cloud includes 3 super-servers with sets of virtual machines installed through which main university systems and platforms are serviced. More than 10 physical servers function in a parallel mode, along with communication cupboards with equipment for telephony and other appliances allowing the functioning of another group of information systems and securing information channels. The entire

equipment is located in three separate server and communication premises with secured uninterrupted power supply, air-conditioning and maintenance.

One super-server is used to provide the complete security of Internet connection between the Internet web, university server cloud and university networks in study buildings and campus. This super-server possesses 10 Gbps network interfaces and is the main routing component of the entire university information system. There is an integral packet of virtual servers constructed on the same machine, through which it performs its functions of a virtual supporting router, as well as of a virtual proxy server and virtual system of monitoring. This super-server resource loading has been below 20% so far, though as a result of development and expansion of network and university cloud in the forthcoming 1-2 years, this loading is expected to reach 50%.

Some small percentage of cloud loading is used for the provision of information services and virtual servers to university business partners. These information services have been related so far mainly to web hosting, mail services, video observation, supply of internet services to clients, students, etc.

## 5.Conclusion

Information technology development towards the construction and use of Cloud infrastructures is considered not only new technology, but rather, a new business model by means of which contemporary communications and information exchange are carried out. New technologies and business model provide substantially new type of information services. Access to information and services is becoming centralized, which facilitates consumers, and increases usability of information resources.

This business model use for university needs is an investment of crucial significance, which is valid to a greater extent in case of education satisfying the needs of transport sector with its typical dynamics and highly technological security requirements. In practice this matches the requirements of creating not only systems and services, promoting contemporary training process and services provision to students, but also the use of similar systems and simulators in the field of transport.

To sum up, implemented approach as well as technical and technological information solutions have been oriented in the right direction, which will lead in the future to substantial improvement of work conditions and performance at the Todor Kableshkov University of Transport. Efforts must be also invested in the expansion of contemporary information systems and in the university cloud complete construction, part of which has already been built at university.

## References

1. Dimitrov D., Proektirane i izgrajdane na avtonomna sistema za kompleksno razprostranenie na internet uslugi v universitetski kampus, Nauchno spisanie “Mehanika, Transport, Komunikacii”, ISSN 1312-3823, broj 3, 2011 г., statia № 0641.
2. Dimitrov D., Logichesko razvitie na mrejata v universitetski kampus, Sbornik dokladi na mejdunarodna nauchna konferencia "Prilojenie na informacionnite i komunikacionni tehnologii v ikonomikata I obrazovanieto", ICAICTEE-2011, UNSS, Sofia.
3. Dimitrov D., System for monitoring of the university computer network performance, Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education – ICAICTSEE-2012, UNWE, Sofia

# Structural Settings of Financial Mechanisms in Bulgarian Forestry after Reformation in 2011

**Anna Dobritchova[1], Nikolay Neykov[2], Ilko Dobritchov[3], Yavor Poryazov[4]**

[1] University of Forestry, Bulgaria
ani_dobrichova@abv.bg
[2] University of Forestry, Bulgaria
nkneykov@gmail.com
[3,4] University of Forestry, Bulgaria

**Abstract.** The paper interpolates concepts of understandings about substance of financial mechanisms in the context of forestry. There are determined the levels of existence and management of financial flows in forestry enterprises and their facilities. Overall system is structured and responsibilities are pinned to the levels of management.

**Keywords.** Financial mechanisms, Forestry, definitions.

## 1.Introduction

To the current moment is not easy task to find appropriate definitions of financial mechanisms in forestry systems. This insists inducing or own definition. In the same time financial mechanisms in Bulgarian science are rarely discussed. They are not exactly defined and their elements are not distinguished in a useful matter.

The main purpose of this paper is to define the financial mechanisms in Bulgarian forestry ant to determine their elements and levels of management.

## 2.Definitions and considerations on financial mechanisms

Financial mechanism is not exact defined in economic and financial researching. In western literature they are discussed in the context of their tasks in any financial systems, but faintly have particular interpretations of their nature. In one of the widest spread definitions is given by Flores, M. at al. [1] It claims that financial mechanisms are instruments for funding different costs. Similar definitions can be found in many business dictionaries and specific sites [7, 8]. Definitions like discussed one are accepted in this research like "narrow definitions". In Russian literature financial mechanisms are well defined. Their usages are universal ones. Vorobev and Osipov [6] claim that financial mechanism is a set of methods, resources and approaches for distribution, redistribution and exact allocation of financial resources. Here for is given an additional part for the overall revealing the nature of such mechanisms- allocation. David Jackson [3] reviews and rounds out the understanding of financial mechanisms with costs. He claims that public costs are integral part of these mechanisms.

To the current moment researching in financial mechanisms of Bulgarian forestry are mainly conducted in University o Forestry and Forest Institute of Bulgarian Academy of Science. Petrin at. al [5] sided appreciates mechanisms like economic assessment of potential forest size. This is core approach in many papers. In some investigations the most static method is used- monetary tables. Many methods like regression analysis are implemented

but never in integration in overall financial system. Particular characteristics of financial mechanisms are discussed in researching of Georgieva [2], Paligorov, Dobricthov [4]. They reveal some of market based ways for economic assessments and financing forest activities. Summarizing the most of the researching in Bulgarian scientific literature, financial mechanisms are not well known and dynamic features of financial system of enterprises are not covered enough.

## 3.Current conditions of Financial Mechanisms in Bulgarian Forestry

Bulgarian Forestry is not the same after the reform made in 2011. The whole principals were changed. Forestry units before served to regional structures. Today these structures are so called Enterprises, six in number. They have to work in a completely different manner- like private entrepreneurial organizations. This means that each of them has to look after its own profitability. Each enterprise is composed of many strategic business units, called Governmental Forest Departments (GFD) and Governmental Hunt Departments (GHD). Each of them makes several basic activities like foresting, timber producing, administration and so called "other activities". Financing different activities should be conducted trough market transactions. The new Law Forestry states some of the rules about transferring financial resources, but there is a fundamental principle involved, that sates preliminary determined inefficiency. The main management department like executive manager etc. at the end of each year makes a decision of how many financial resources to distribute from one GFD or GHD to another.

Until now nobody has summarized the exact structure of financial mechanisms in new system. The main reason is that there is no exact definition, like it was mentioned above. So for the purpose of current and future researching on new Bulgarian Forestry financial system is vital such a definition to be given:

Under the Financial Mechanism in Bulgarian Forestry should be understood a set of approaches which provide coverage of the costs made by the departments and enterprises with appropriate financial resources, needed to equate or make positive the budgets of these organizations.

Definition above gives an answer about the role of local budgets and in the enterprises as well. They have to provide profitableness. Consequently every step toward optimization and improvement should be done like that. Claims like "the forest departments should work for the purpose of forest not for profits" are inconsistent. If the current legislation determines the main producing units like "Enterprises", their financial mechanisms have to ensure positive budget balances.

The next step is to define the structures of the mechanisms. At first there is a managerial structure, built of departments and enterprises. During the research was discovered, and for first time mentioned here, that there two main types of mechanisms can be distinguished. Mechanisms First Level (MFL) and Mechanisms Second Level (MSL).

The MFLs are all enterprises and controlling institutions like regional directorates. They form indirect cash flows. These cash flows are composed of net profits, gained by GFDs and GHDs. The first level mechanisms determine tactical and strategic parameters of financial system. They also realize allocation function of financial mechanisms. Enterprises plan the overall budget and consequently boundaries of the local budgets of departments. All these features of the mechanisms are not determine in the legislations so in detail. At the same time their role is to work after- whatever the results are they have to correct the negative ones.

MSLs are the "front line soldiers". They immediately connect the system to the markets. All the financial recourses they use and invest are direct results of the prices they give to the customers. They are firsthand responsible for the budget covering. Here are the major weaknesses of the mechanisms and the system at all. They don't have the exact rules on how to react permanently if any ineffectiveness has appeared.

## 4.Conclusion

This paper reveals the main elements of forest system that gain the economic efficiency and resources for development. It places the foundations for improvement of the financial mechanism's management in the context of contemporary forestry circumstances. The research reveals that current legislation and scientific literature do not determine and consequently implement the potential capabilities of financial mechanisms. Definition and structure given here support analysis and future developments of the system. Many questions still remain unanswered. Future researching would give answers about the way of effective funding allocation, mechanisms simplification and functions optimization.

## References

1. Flores, M. at al., Financial Planning for National Systems of Protected Areas: Guidelines and Early Lessons, Arlington, Virginia, US., 2008, p.29.
2. Georgieva, D., I. Markov Organized lumber markets in the transitioning to market economy, MSD, 2001, pp.91-95
3. Jackson, David, Effective Financial Mechanisms at the national and local level for Disaster Risk Reduction, United Nations Capital Development Fund, 2011.
4. Paligorov,. I., I. Dobritchov, D. Vasilev, Concept of Information System for Bulgarian Forest Management, MSD, 2000, 1-2, pp.56-61
5. Petrin, R., K. Bogdanov, I. Markov, Multifunctionality analusis in forest resources usage in Regional Directorate Kyustendil, MSD, 2009, 1, pp.116-122
6. Vorobev, Osipov, Theory of Finance, MIU, Minsk, 2005.
7. http://www.businessdictionary.com/definition/financial-mechanism.html
8. http://www.wisegeek.com/what-is-a-financial-mechanism.htm

# Improving Current Condition of Bulgarian Forest Enterprises after the Reform in 2011 by Optimization of Financial Resources

**Anna Dobritchova[1], Nikolay Neykov[2],**

[1] University of Forestry, Bulgaria
ani_dobrichova@abv.bg
[2] University of Forestry, Bulgaria
nkneykov@gmail.com

**Abstract.** The paper presents the model developed during a research, conducted in University o Forestry, Bulgaria. It is a mathematical model for optimal resource allocation trough distribution of revenues and costs between different activities being made by Forestry Enterprises.

**Keywords.** Financial mechanisms, Forestry, optimization, allocation.

## 1.Introduction

Today's legislation system and scientific research do not provide methodology of optimal allocation of financial resources, which is one of the main tasks of their own. All kind of allocation activities are concerned on distribution of budget surplus in the end of the year or any month.

## 2.Optimization and its place in the Financial Mechanisms of Bulgarian Forestry

In the conditions of modern economic development of forestry in Bulgaria tend exists of increasing the volume of information. This leads to complications of making management decisions. Budgeting in forest enterprises requires improvement of traditionally established methods, as well as those in the legislative base with new scientifically founded ones. The need to obtain the best results in the field of financial planning mechanisms requires the application of methods leading to optimization. In essence optimization is purposeful activity to obtain the best result in a sense, and under certain conditions. [3] The optimization is to make something that works perfectly as possible or more efficient. [1] These definitions understand optimization in a broader sense. When talking about financial mechanisms optimization should be considered in the strict sense. In a narrow sense the optimization of financial mechanisms should be leading mathematical model to quantify its structure in order to maximize the economic efficiency of the financial system of the enterprise. Here is the place to summarize the understanding of economic efficiency, which in this study will contribute to the adequate choice of optimization criteria. As stated Eddie McLaney [2], business objectives, and in the context of the survey these optimality criteria could be the following few:

- Maximization of profit.
- Maximization of return on capital.
- Survival.

- Long-term stability and others.

Problem Solving optimization performance of financial mechanisms includes the development of a number of alternative options. They can vary greatly in terms of pre-accepted criterion for assessing the effectiveness of the plans. The option that achieves previously accepted criterion is optimal. In essence this is the detection of maximum or minimum of a mathematical function. The points at which developed function reaches a maximum or minimum value are called extremes. In the development and practical implementation of the modeled relationships, there are significant difficulties in the area of planning. Sometimes the biggest error is to be developed alternatives on yearly basis. By this way the influences of permanent acting negative factors remain missed and are impossible systemic failures to be corrected. Much more effective method is to develop a universal model in which the principal basis does not change in the long run. Differences will arise only from the values of the output information. In real terms the determination of the optimal structure and size of financial mechanisms is a complex task. This arises from the following circumstances:

- Financial reporting has to undergo a series of transformations for the purpose of mathematical modeling;
- Many different situations appeared that require adaptability of the developed mathematical model;
- Conversion of models in management tool requires training of those who use it and sometimes resistance on their part.

Given number of optimized tasks related to planning, obtained by mathematical method allows generation of a large economic impact. At the same time enterprises do not need to spend great amount of money to extend and improve financial management. The achievement of the goal by solving problems in budgeting is measured by the developed universal criterion for optimality. This issue is so important that the question of its own effectiveness. It should be cumulatively fulfilled the following requirements:

- Be quantifiable;
- After being placed economic task, the criterion should only;

Management of financial mechanisms in the practice of forestry requires consideration of many factors. Here are revealed the imperfections of the mathematical modeling in economics. They stem from the uniqueness of the criterion of optimality for each system. This means that the mathematical model should be a tool to support the process of decision making, but not mandatory rule. If the model is developed effectively, the latter should be adapted to other suboptimal decisions without prejudice to the principle of its functioning.

Mathematical model which optimizes the performance of financial mechanisms is a source of new information previously unknown to researchers and managers dealing with finance in the forest sector in Bulgaria.

In most - general mathematical models are mathematical description of some processes, using the equations and inequalities that describe the processes studied. The mathematical model is a starting point for the application of the mathematical methods for solving problems of the practice of forestry.

## 3.The Model and empirical data

As it was stated in the first chapter of this work, the optimization model must achieve a goal of corporate governance and to provide additional information, that is not available at the time of calculations. In the context of the purpose of the present study, the optimization model is to form optimal specifications of financial mechanism and help enterprise managers in their current activity. The main indicator to be chosen as a modeling element of the

objective function is the relative economic efficiency. It is very useful for formulating analytical model function, because its maximum value is a major characteristic criterion for optimizing the economic activity of any business organization. Therefore, the derivation of a maximum of target function which describes the economic efficiency does not contradict the logic of the management of each process running in the surveyed sites.

This study uses a model of the "general distribution task." The purpose of such models is the optimal attachment of resources to needs. The model is characterized by the following objective function:

$$F_x=\sum\sum c_i Ef_j \rightarrow \min, \quad (1)$$

where $C_i$ are the allocatedcosts of type i and $Ef_j$ are the revenues efficiency of type j,

in the constraints of $\sum c_i=C_i$ and $\sum r_j =R_j$

The economic logic of the proposed model is as follows: To minimize revenue efficiency for each business and / or their divisions by allocating the costs from one to another, so that a large part of the costs in unprofitable activities to shift to the more profitable ones. Specific implications of such optimization in practice are reflected in the many activities, such as the inclusion of allocated costs in cost of products created in unprofitable activities, as evidenced by the annual reports of enterprises. The purpose of such an optimal plan is to avoid the delay of the annual financial results and hence to absorb losses. It's necessary in operating matter, even at the level of revenue measures to be taken.

The main modeling elements are economic efficiencies in operations. Attachment of resources - financial mechanisms to costs - needs is implemented cross. Financial mechanisms specific to each activity may be attached to the same costs or to other activities. Revenue efficiencies are the following:

$$Ef=1/R_j, \quad (2)$$

where $R_j$ are revenues of type j.

Empirical data are taken from the Northwestern Forest Enterprise for 2012 in bulgarian curency. The activities are:

- Timber production
- Hunting
- Forestation
- Administration
- Other

The revenues and costs are differentiated in the matter of activities.

## 4.Results

The model presented above pushes costs to the most profitable activity. If come costs are placed into other activity, this means that the revenues taken in it during the year have to cover these costs also. The new efficiencies include the sum of the new revenue efficiency and 1, which means that distributed costs have to increase the initial efficiency of that activity. The results are shown in table 1

Table 1.

| | Timber production | Forestation | Hunting | Other | Administration | Overall |
|---|---|---|---|---|---|---|
| initial | 2.634 | 0.187 | 0.921 | 0.944 | 0.080 | 0.953 |
| optimal | 1.372 | 1.792 | 0.940 | 1.941 | 1.200 | 1.449 |

Optimization states the exact values of the desired economic efficiency. It is visible that the most profitable activity doesn't have to be as well but to give resources to other activities. The bigger the difference is the harder the decisions are. For example administrative taxes should be risen due to provide bigger profit per unit services. The overall efficiency has to be increased by 0,5 lv per each cost. Costing and pricing are the most powerful tools for gaining the desired profits.

## 5.Conclusion

Optimization presented in the paper is only small part of the entire research on Bulgarian Forestry after y.2011. It only marks the path for economies. But if it takes it worthy place in financial planning the results will be much better than now. Many of the activities will provide enough resources for cover problems in other not only timber production. The future of Bulgarian Forestry belongs to optimality and scientifically justified approaches.

## References

1. Bettinger, P., K. Boston, J.Siry, D.L. Grebner, Forest management and planning, Elsevier, 2009.
2. McLaney, E. J. Business finance : theory and practice, Prentice-Hall, 2009.
3. http://www.businessdictionary.com/definition/optimization.html

# Estimation of Freshwater Sensitivity to Acidification Using Different Approaches

**Sonya Damyanova**

[1] University of Forestry, 10 Kliment Ohridsky Blvd., 1756 Sofia, Bulgaria
sonya_damyanova@abv.bg

**Abstract.** The paper reviews the some popular models and indicators for estimation of freshwater acidity. Acid neutralize capacity is widely used as a measure of freshwater acid status. Three methods for calculating critical loads of acidifying N and S deposition are described. Steady-State Water Chemistry (SSWC) model, empirical diatom model and linear regression model are discussed.

**Keywords.** Freshwater acidity, Critical load of acidity, Acid deposition, Aquatic ecosystem

## 1.Introduction

The ecological condition of running waters is determined by complex of physical and chemical events. These are controlled by climatic factors, such as storm events and droughts. Storm events affect streams directly through increased discharge, and indirectly through changes in water quality; in semi-natural catchments, storm events typically lead to increased flow through surface soil horizons, diluting base cation inputs from weathering and increasing inputs of dissolved organic carbon and/or nitrate from shallow soils. These processes can lead to ecologically damaging episodic acidification, particularly when long-term catchment acidification associated with anthropogenically-derived sulphur (S) and nitrogen (N) deposition [6]. Additionally, in coastal areas, heavy storms can result in pulses of sea salt deposition, due to entrainment of marine aerosols from breaking waves at high wind speeds. The major impact of marine ion deposition, generally termed the "sea-salt effect", is the alteration of cation-exchange equlibria between soil and soil water. This is a transient process, in which episodic inputs of marine base cations, primarily sodium (Na) and magnesium (Mg), cause the displacement of adsorbed acid cations, protons (H) and inorganic Aluminium. This in turn can lead to the episodic acidification of runoff [4]. In many areas, the legacy of base saturation depletion due to historic acid deposition means that soils, and therefore surface waters, are likely to remain highly susceptible to sea salt episodes for many years to come [5]. Climate change induced increases in the future severity of high-flow events in general, and sea-salt episodes in particular, therefore have the potential to reduce, or even negate, the benefits of reduced S and N deposition for many aquatic ecosystems.

Prevention of acid rain has started by emission control of nitrogen and acidity given in the Convention of Long Range Transboundary Air Pollution (CLRTAP) from 1979 year. Sites seriously affected by eutrophication has not almost left recently as a result of that international environmental strategy. The main idea was to create an integrated indicator for assessment of the whole ecosystem and its functions because of the complex relationships existing between individuals. Anthropogenic impact on ecosystems has expressed by calculation of critical loads of some air pollutants. It has started by estimation of acidity by

calculation of critical loads of main acidifying gaseous such as nitrogen oxides and sulfur dioxide. After successful implementation of that approach to restrict acidity, it has been applied to heavy metal pollution since 90's years of the previous century.

Long-range transported air pollution contains nitrogen and sulphur compounds contributing to acidification of soils and freshwaters [9]. Acidification of acid sensitive freshwater systems provided some of the earliest evidence of the damage caused by sulphur and nitrogen emissions [19]. The sensitivity of these systems suggested that they were ideal for studying the effects and response to changes in pollution deposition.

The aim of this paper is to summarize some popular methods for surface waters assessment in term of acidification .

## 2.Sensitivity Classes

Practice has shown the most of the available water data contained information limited to properties such as pH, alkalinity and/or calcium. Lake sensitivity to acidic deposition has traditionally assigned based on classes developed by the National Research Council of Canada [14] and the Saskatchewan Research Council [16], Swain [18]. The methodology assigns a sensitivity rating to each lake of high, medium, or low based on the known chemical parameters. High, medium and low sensitivity classes correspond to critical loads of 0.25, 0.50 and 1.00 keqH.ha$^{-1}$.yr$^{-1}$, respectively (Table 1). If a particular lake had measurements of more than one chemical parameter, then the lowest sensitivity class was chosen to offer more complete protection, following Swain [18] and in accordance with First Nations' priorities.

Table 1. Assessment of acidification sensitivity based on single lake chemistry parameters [10]

| Sensitivity | Alkalinity ($mgCaCO_3.dm^{-3}$) | pH | Ca ($mg.dm^{-3}$) | $HCO_3$ ($mg.dm^{-3}$) | Conductivity ($\mu S.cm^{-1}$) | CL(A) $keqH.ha^{-1}.yr^{-1}$ |
|---|---|---|---|---|---|---|
| High | 0 – 10 | <6,6 | <4 | 1 – 12 | 0 – 35 | 0.25 |
| Moderate | 11 – 20 | 6,7 – 7 | 4 – 8 | 12 – 24 | 22 – 78 | 0.5 |
| Low | 21 – 40 | >7 | >8 | >24 | >60 | 1 |

## 3.Acid Neutralize Capacity (ANC)

Acid neutralize capacity is defined by Reuss and Johnson as the potential of a solution to neutralize additions of strong acids to a given level [15]. ANC is calculated according to:

$$ANC = [HCO_3^-] + [A^-] – [H^+] – [AL^{n+}] \quad (2.1)$$

For lake with negligible content of organic matter, $[A^-]=[AL^{n+}]$ and approximately zero,than:

$$ANC = [HCO_3^-] – [H^+] \quad (2.2)$$

Using ion balance of water:

SUM (charge cations) = SUM (charge anions) in $meq.m^{-3}$:

$$[H^+]+[AL^{n+}]+[Ca^{2+}]+[Mg^{2+}]+[K^+]+[Na^+]+[NH_4^+]=[HCO_3^-]+[Cl^-]+[SO_4^{2-}]+[NO_3^-] \quad (2.3)$$

And replacing in eq. 2.2:

$$ANC = [Ca^{2+}] + [Mg^{2+}] + [K^+] + [Na^+] + [NH_4^+] = [Cl^-] + [SO_4^{2-}] + [NO_3^-] \quad (2.4)$$

ANC is calculated as the difference between base cations and acid anions. This parameter is widely used as a measure of freshwater acid status, and an indicator of biological conditions [9]. Unlike pH and alkalinity, ANC is conservative with respect to $CO_2$ degassing and reactions with aluminium or organic species.

## 4.Steady-State Water Chemistry (SSWC) Model

The purpose of critical loads for aquatic ecosystems is to estimate the maximum deposition below which 'significant harmful effects' on biological species do not occur. Similar to terrestrial ecosystems, the links between water chemistry and biological impacts cannot be modelled adequately at present as such, water quality criteria are generally used to derive critical loads for aquatic ecosystems.

The critical load of a lake or stream can be derived from present day water chemistry using annual mean values. It assumes that all sulphate ($SO_4^{2-}$) in runoff originates from sea salt spray and anthropogenic deposition (no adsorption or retention). The model uses Acid Neutralize Capacity (ANC) as the variable linking water chemistry to sensitive indicator organisms in freshwaters. The SSWC model [8, 17] has calculated critical load of acidity (CL(A)) based on the principle that the acid load should not exceed the non-marine, non-anthropogenic base cation input and sources and sinks in the catchment minus a buffer to protect selected biota from being damaged, i.e.:

$$CL(A) = BC^*_{dep} + BC_w - Bcu - ANC_{limit} \quad (3.1)$$

where BC*dep (BC=Ca+Mg+K+Na) is the sea-salt corrected (with Cl as a tracer) non-anthropogenic deposition of base cations, BCw is the average weathering flux, $Bc_u$ (Bc=Ca+Mg+K) is the net long-term average uptake of base cations in the biomass (i.e., the annual average removal of base cations due to harvesting), and $ANC_{limit}$ the lowest ANC-flux that does not damage the selected biota. Since the average flux of base cations weathered in a catchment and reaching the lake is difficult to measure or to compute from available information, a critical load equation that uses water quality data alone has been derived.

In pre-acidification times the non-marine flux of base cations from the lake, $BC^*_0$, is given by (all parameters are expressed as annual fluxes, e.g. in $eq.m^{-2}.yr^{-1}$):

$$BC^*_0 = BC^*_{dep} + BC_w - Bc_u \quad (3.2)$$

$$CL(A) = BC^*_0 - ANC_{limit} \quad (3.3)$$

$$CL(A) = Q \,.\, ( BC^*_0 - ANC_{limit}) \quad (3.4)$$

where the second identity expresses the critical load in terms of the catchment runoff Q (in $m.yr^{-1}$) and concentrations ([X]=X/Q).

To estimate the pre-acidification flux of base cations the present flux of base cations, $BC^*_t$, have to be calculated:

$$BC^*_t = BC_{dep} + BC_w - Bc_u - Bc_{exc} \quad (3.5)$$

where $BC_{exc}$ is the release of base cations due to ion-exchange processes. Assuming that deposition, weathering rate and net uptake have not changed over time, $BC_{exc}$ is obtained by subtracting eq. 3.2 from eq. 3.5

$$Bc_{exc} = BC^*_t - BC^*_0 \quad (3.6)$$

This present-day excess production of base cations in the catchment is related to the long-term changes in inputs of non-marine acid anions by the so-called F-factor:

$$BC_{exc} = F \,.\, \Delta SO_4^* \,.\, \Delta NO_3 \tag{3.7}$$

For the pre-acidification base cation flux we thus get from eq. 3.6 ($DX = X_t - X_0$):

$$BC^*_0 = BC^*_t - F \,.\, (SO_{4,t}^* - SO_{4,0}^* + NO_{3,t} - NO_{3,0}) \tag{3.8}$$

The pre-acidification nitrate concentration, $NO_{3,0}$, is generally assumed zero.

The F-factor describes the degree to which anthropogenic sulfate in the lake had been neutralized by soil processes, i.e. the neutralization ratio:

$$F = (BC^*_t - BC^*_0) / (SO_{4,t}^* - SO_{4,0}^* + NO_{3,t}) \tag{3.9}$$

The SSWC model has developed for waters of Fennoscandia, Scotland, Canada and Ireland. In such areas, surface waters are generally more sensitive to acid inputs than soils. The model assumes that all sulphate in runoff originates from deposition alone, except for a small geologic contribution. In areas where the geological conditions lead to more alkaline waters, the SSWC model has to be modified, since significant amounts of sulphate from geological sources can be present in the runoff water.

The critical level of ANC varied among fish species, with Atlantic salmon being the most sensitive, followed by brown trout [7]. Henriksen et al. concluded that Atlantic salmon is a good indicator of acidification of rivers, and trout is a useful indicator for acidification of lakes. Based on an evaluation of fish and invertebrate populations, a critical lower limit of $[ANC] = 20$ meq.m$^{-3}$ was suggested as the tolerance level for Norwegian surface waters [11]. This limit has been widely used [3, 13]; however, it has been set to zero in the United Kingdom (CLAG 1995) and to 40 meq.m$^{-3}$ in south-central Ontario, Canada [8]. Lydersen et al. [12] suggested to correct the $ANC_{limit}$ with the amount of organic acids present in the lake. They replaced 'standard' ANC with an 'organic acid adjusted' ANC ($[ANC]_{oaa}$). They define this quantity as:

$$[ANC]_{oaa} = [ANC] - 1/3 \,.\, m \,.\, TOC \tag{3.10}$$

where m·TOC is the total organic carbon expressed in meq.m$^{-3}$ (m being the charge density). Such a correction leads to a lower ANC-limit, i.e., higher critical loads. The lakes studied receive very low to very high (for Norway) levels of deposition, thus including a wide range of affected lakes. This implies that for a given ANC-value lakes of varying sensitivity exist, receiving varying amounts of deposition.

The simplest functional relationship with this feature is a linear relationship between $[ANC]_{limit}$ and the critical load CL:

$$[ANC]_{lim} = k \,.\, ANC.$$

This gives the following implicit equation for the critical load (eq. 3.4) and after re-arranging for CL:

$$CL = Q \,.\, ([BC^*]_0 / (1 + k \,.\, Q) \tag{3.11}$$

$$[ANC]_{lim} = k \,.\, Q \,.\, ([BC^*]_0 / (1 + k \,.\, Q). \tag{3.12}$$

For the constant [ANC]limit used earlier, the proportionality constant k should be derived from data. If we assume that for CL=0, the $[ANC]_{limit}=0$ and for a critical load of 200

meq.m$^{-2}$.yr$^{-1}$ the ANC-limit should not exceed 50 meq.m$^{-3}$, as has been assumed in Sweden, we arrive at a k-value of 50/200 = 0.25 yr.m$^{-1}$.

In addition, for CL-values above 200 meq.m$^{-2}$.yr$^{-1}$ we set the $[ANC]_{limit}$ to the constant value of 50 meq.m$^{-3}$. The value of k is derived from experience in the Nordic countries and reflects the geology, deposition history and biological diversity (fish species) of that region. For different regions other k-values may be more appropriate.

## 5.Empirical diatom model

The empirical diatom model is an alternative approach to the SSWC model and is developed from paleolimnological data [1]. The point of acidification is indicated by a shift towards a more acidophilous diatom flora. Diatoms are amongst the most sensitive indicators of acidification in freshwater ecosystems. The acidification status (as defined by diatom analyses) of 41 sites in the United Kingdom (UK) was compared to site sensitivity (defined by lake-water calcium concentrations) and current deposition loading. The optimal separation of acidified and non-acidified sites is given by a [Ca]:Sdep ratio of 94:1 [1], acidified sites having a ratio less than 94:1. This critical ratio, determined by logistic regression, can be used to define critical sulphur loads for any site, including streams.

Critical load values are calculated from pre-acidification calcium concentrations using the F-factor [2]. For example, the critical sulphur load for a lake with a $[Ca]_0$-value of 40 meq.m$^{-3}$ is approximately 0.43 keq.ha$^{-1}$.yr$^{-1}$. The diatom model has been adapted to provide critical loads, and critical load exceedances, for total acidity (sulphur and nitrogen). Exceedance values for total acidity require a measure of the fraction of deposited nitrogen leached to the surface waters. This is calculated from the differences between the ratios of sulphate/nitrate in the water and in the deposition at the site. In this way the fraction of the nitrogen deposition contributing to acidification, $f_N$, is added to the value of sulphur deposition to provide a total 'effective' acid deposition:

$$f_N = S^*_{dep} / N = [SO_4^*] / [NO_3] \quad (4.1)$$

This model assumes equilibrium between sulphur deposition and sulphate in water, and only applies to sites with no additional catchment nitrogen inputs. The diatom model has been re-calibrated for total acidity loads by substituting total effective acid deposition for sulphur deposition. The resulting critical ratio is 89:1, slightly lower than when considering sulphur alone. The basic equation for the critical load of total acidity in the empirical diatom model is therefore as follows:

$$CL(A) = [Ca^*]_0 / 89 \quad (4.2)$$

where CL(A) is in keq.ha$^{-1}$.yr$^{-1}$ and $[Ca^*]_0$ in meq.m$^{-3}$. The pre-acidification Ca-concentration is calculated as:

$$[Ca^*]_0 = [Ca^*]_t - F_{Ca} \cdot ([SO_4^*]_t - [SO_4^*]_0 + [NO_3]_t - [NO_3]_0) \quad (4.3)$$

$$F_{Ca} = \sin (\pi/2 \cdot [Ca^*]_t / [S_{Ca}]) \quad (4.4)$$

where $[S_{Ca}]$ is the Ca-concentration at which $F_{Ca}$=1. It can vary between 200 and 400 meq.m$^{-3}$, depending on location. In the UK critical loads mapping exercise a value of $[S_{Ca}]$=400 meq.m$^{-3}$ has been used, and in waters with $[Ca^*]t>[S_{Ca}]$, $F_{Ca}$ was set to 1. The pre-acidification nitrate concentration, $[NO_3]_0$, is assumed zero. The pre-acidification sea-salt corrected sulphate concentration, $[SO_4^*]_0$, is estimated according to [2].

The diatom model has been calibrated using sites and data from the UK. However, a major advantage of the approach is that predictions for any lake can be validated by

analysing diatoms in a sediment core. In this way the applicability of the model to sites outside the UK can be tested.

## 6.Linear Regression Model

J. Krzyzanowski et al. 2010 [10] have analyzed relationships between individual chemical characteristics as mass fluxes in $eq.ha^{-1}.yr^{-1}$ and critical load of acidity. Linear relationships with a coefficient of determination ($r^2$) greater than 0.85, were found for Ca, alkalinity and conductivity only:

$$CL(A) = 1.55 \,.\, [Ca] \,.\, Q / 100 \quad (5.1a)$$

$$CL(A) = 1.52 \,.\, [Ca] \,.\, Q / 100 \quad (5.2b)$$

$$CL(A) = 1.04 \,.\, [ALK] \,.\, Q / 100 \quad (5.3a)$$

$$CL(A) = 1.02 \,.\, [ALK] \,.\, Q / 100 \quad (5.4b)$$

$$CL(A) = 10.10 \,.\, [CON] \,.\, Q / 100 \quad (5.5a)$$

$$CL(A) = 9.88 \,.\, [CON] \,.\, Q / 100 \quad (5.6b)$$

where CL(A) is lake sensitivity to acidification in $keq.ha^{-1}.yr^{-1}$ estimated using SSWC, [Ca] is lake water calcium concentration ($meq.m^{-3}$), [ALK] is lake water alkalinity measured as CaCO3, [COND] is conductivity measured in $\mu S.cm^{-1}$ ($meq.m^{-3}$), Q is runoff ($m.yr^{-1}$), and the factor of 100 is used to convert fluxes to units of $keq.ha^{-1}.yr^{-1}$. Equations labelled 'a' used an $[ANC]_{lim}$ ¼ 50 $meq.m^{-3}$ and 'b' an $[ANC]_{lim}$ ¼ 100 $meq.m^{-3}$ offering assessment options at two different levels of protection.

## 7.Conclusion

Steady-State Water Chemistry (SSWC) model is wide used for estimation of surface freshwater acidity state. As the critical load of acidity reveals a sensitivity of aquatic catchments to acidifying S and N contaminants, it is a reliable indicator for water state assessment and protection of biota in rivers and lakes. Difficulties and uncertainties in calculation of $[ANC]_{lim}$ and pre-acidification ions limit an application of the model. Linear Regression Model suggested by J. Krzyzanowski et al. facilitates assessment of CL(A) in case of information limited of properties such as alkalinity, Ca concentration and conductivity.

## References


1. Battarbee R.W., T.E.H. Allott, S. Juggins, A.M. Kreiser, C. Curtis, R. Harriman, R., 1996. Critical loads of acidity to surface waters: an empirical diatom-based Palaeolimnological model, *Ambio*, 25 (5), 366–369,1996.
2. Brakke D.F., A. Henriksen, S.A. Norton. A variable F-factor to explain changes in base cation concentrations as a function of strong acid deposition, *Verh. Internat. Verein. Limnol*, 22, 692–698, 1990.
3. Duan L., J. Hao, S. Xie, K. Du. Critical loads of acidity for surface waters in China, *Science of Total Environment*, 246 (1), 1–10, 2000.
4. Evans C.D., D.T. Monteith, R. Harriman. Long-term variability in the deposition of marine ions at west coast sites in the UK Acid Waters Monitoring Network: impacts on surface water chemistry and significance for trend determination, *Science of Total Environment*, 265, 115-129, 2001.
5. Evans C.D. Modelling the effects of climate change on an acidic upland stream, *Biogeochemistry*, 74, 21-46, 2005.
6. Evans C.D., B. Reynolds, C. Hinton, S. Hughes, D. Norris, S. Grant, B Williams B. Effects of decreasing acid deposition and climate change on acid extremes in an upland stream, *Hydrology and Earth Systemt Science*, 12, 337-351, 2008.
7. Henriksen A., P.J. Dillon, J. Aherne. Critical Loads of Acidity for surface waters and south-central Ontario, Canada: region al application of the steady-state water chemistry model. *Canadian Journal of Fisheries and Aquatic Science*, 59, 1287–1295,2002.
8. Henriksen A., M. Posch. Steady-state models for calculating critical loads of acidity for surface waters, *Water Air Soil Pollutution Focus*, 1, 375–398, 2001.
9. ICP Waters, Norwegian Institute for Water research, Report 105/2010.
10. Krzyzanowski J., John L. Innes. Back to the basics – Estimating the sensitivity of freshwater to acidification using traditional approaches, *Journal of Environmental Management*, 91, 1227–1236, 2010.
11. Lien L, G.G. Raddum, A. Fjellheim, A.A. Henriksen. A critical limit for acid neutralizing capacity in Norwegian surface waters, based on new analyses of fish and invertebrate responses, *Science of the Total Environment*, 177, 173–193, 1996.
12. Lydersen E, T. Larssen, E. Fjeld. The influence of total organic carbon (TOC) on the relationship between acid neutralizing capacity (ANC) and fish status in Norwegian lakes, *Science of the Total Environment*, 362, 63–69,2004.
13. Moiseenko, T.I. Assessment of critical loads of acid fallout for surface waters, *Water Research*, 29 (3), 292–298, 2002.
14. NRCC - National Research Council of Canada, 1981. Acidification in the Canadian Aquatic Environment: Scientific Criteria for Assessing the Effects of Acidic Deposition on Aquatic Ecosystems. National Research Council of Canada, Associate Committee on Scientific Criteria for Environmental Quality, NRCC No.18475, 369, 1981.
15. Reuss J. O., D.W. Johnson D. W. Effect of soil processes on the acidification of water by acid deposition. Journal of Environmental Quality, 14, 26-31, 1986.
16. SRC – Saskatchewan Research Council. Evaluation of the Status of Surface Water Sensitivity Mapping for Acidic Deposition in Western Canada. Coordinating Committee on Surface Waters, SRC No. C-805-11-E-82, 90, 1982.
17. Sverdrup H., W. De Vries, A. Henriksen. Mapping Critical Loads. Miljo¨ rapport, 1990:14. Nordic Council of Ministers, Copenhagen, 124, 1990.
18. Swain L.G. Second Report on Chemical Sensitivity of B.C. Lakes to Acidic Inputs. British Columbia Ministry of Environment and Parks, Victoria, B.C., 31, 1987.
19. UNECE-CLRTAP. Manual on Methodologies and Criteria for Modelling and Mapping Critical Loads and Levels, and Air Pollution Effects and Trends, 202 pp. URL. United Nations Economic Commission – Convention on Long-range Transboundary Air Pollution, Berlin, Germany, 2004.

# Main Principles of Managing Big Data Analitics

**Dorina Kabakchieva[1], Kamelia Stefanova[2]**

[1] UNWE, 1700 Sofia, Bulgaria
dkabakchieva@unwe.bg

[2] UNWE, 1700 Sofia, Bulgaria
kstefanova@unwe.bg

**Abstract.** The effects of big data are enormous in a practical scale, as the technology is applied to find solutions for everyday problems. Big data is set to reshape the way of traditional working and thinking. The knowledge, that for years was devoted to understand the past, now is directed to predict the future. The amount of data will continuously grow, as well as the technologies to process it. The paper describes the substance of Big Data and updates the main characteristics. The types of Big Data analytics are discussed with their specific tools and outcomes.

**Keywords.** Business Intelligence, Big Data

## 1.Introduction

The business environment today main characteristics focus on the severe competition and high dynamics, on the other hand, the expansion of volumes of data deepen the challenges and risks for any enterprise. In these conditions each opportunity that could provide competitive advantages should be well accepted and meaningfully applied. One of the indisputable alternatives in this direction is the availability and use of innovative methods for data collection and management - the Big Data approach.

Big Data becomes a challenge and strong direction for the IT infrastructure development of all companies' types and sizes. Companies need to introduce new approaches, methods and tools for processing data and intelligently to manage what data, how to be analyzed in order to bring new values to organizations. Big Data is not only getting expressed in volumes of data, this is a new philosophy in management, analysis and use of data for the decision taking processes at all management levels of the companies.

Organizations are facing every day the growing volumes of data - terabytes, petabytes approaching already the new measures – exabytes, zettabytes, yottabytes. Data comes from different sources: operational and transactional systems, management systems, contact with customers, from media, web, mobile.

The explosion of data is not a new trend. The data growth started in late 1970s. What has changed is the speed of growth of diverse and complex data. The new challenge today requires innovative approaches for improving the analysis and use of information that could support and transform businesses.

According to the International Data Corporation, the volume of data worldwide is expected to increase 44 times in the period by 2020 and to reach about 35.2 zettabytes. Although for 75% of the information in the digital world people are responsible, 80% of the data at some point in their digital life is established and produced by organizations. Amount of information that people create personally - documents, photos, music, etc., is much less than the amount of information created for them by companies and institutions.

The success of many businesses in the coming years will depend largely on the ability to support facing the challenges of the new digital universe.

## 2.Big Data definition and characteristics

Big data is a new philosophy that enables organizations to gather, store, manage, and manipulate huge amounts of data at the right speed, at the right time, to gain the right insights.

Perhaps the term sounds and could be connected to the amount of data, but this should not be the only focus. The term Big Data refers to the technology and includes processes and tools.

It is believed that the term comes from companies engaged in Internet research on many different sources of data and which need to make applications with large aggregations of poorly structured data.

*Big data* defines the massive volumes of structured and unstructured data being difficult to be processed by using traditional databases and software techniques. Data is too big in volume, changes very quickly and exceeds the typical enterprise processing capacity. Big data has the potential to help large companies improve operations and support fast and intelligent solutions taking.

Big Data is a set of techniques and technologies that require new forms of data integration in order to investigate the hidden values within the large datasets that consist of diverse, complex and large-scale data.

Businesses have passed through long experiences to find the right pragmatic approach to identify and collect information about their customers, products, and services. During the last century, companies and markets have grown globally and become very complicated. In order to survive and get competitive advantages, companies need to innovate in new product lines and introduce diversified channels to get to their clients.

Data expansion is not generated only by businesses. Research organizations develop and extend their computing capacity in order to introduce and examine new sophisticated models and learn how to process all heterogeneous types of data coming from many and different sources of scientific data.

Some data is structured and stored in traditional relational databases, other data, including documents, customer records, additional materials – images, videos and audios, is unstructured. Companies concentrate efforts how to organize, process and analyse all the available data to take informative on-time decisions.

Another data challenge comes from the new types of data generated by machines and sensors, from human generated sources - data from social media, click-stream data of website interactions. In addition, the availability and implementation of mobile devices, equipped with global access to networks create new sources for data.

Each type of data can be managed and searched for independently. The main challenge today grows with the need companies to integrate all the different types of data and take decisions based on multidimensional analysis of all the data.

The traditional data management ways could not deliver a solution how knowledge could be extracted from so much information in so many different forms. It appeared obvious to introduce new approaches and technologies in order to manage data differently and this is the opportunity, and on the other hand, the challenge of Big Data.

The Gartner analyst Doug Lane identified in late 2001 the challenges for the growth of data and the ability this trend to be presented in three main dimensions - increasing in *volume* (amount of data), *velocity* (speed of data processing), and *variety* (range of data types

and sources) (Fig 1). Today, this model of "3Vs" is still vital and mainly used for description of big data.

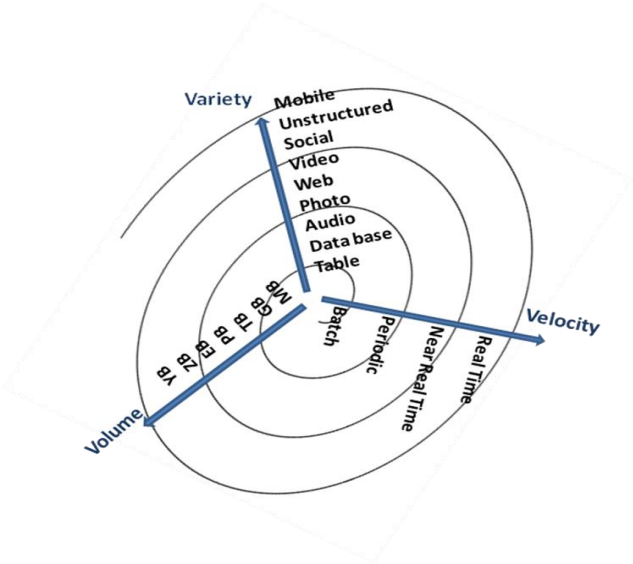



Figure 1: Characteristics of Big Data

In 2012, Gartner updated the definition as follows: "Big Data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." So, the 3 main Vs remain to be the substance of Big Data.

- Volume - the amount of generated data is a very important characteristic. This is the amount of data that requires the new approaches, models and technologies collecting, integrating and processing the data. Volume describes the aspect of data whether could be classified as Big Data. The name "Big Data" itself contains a term that is related to the data size and focuses on the substance. It is reported that in 2012 about 2.5 exabytes data have being created every day, and this number is getting doubled every 40 months. More information is crossing Internet per second than that was stored in all internet sites 20 years ago.
- Variety - this next aspect of Big Data refers to diversity. It means that the Big Data philosophy should be known by data analysts. This approach supports people who are closely related to analysis of data and assists them in producing effective data to be used for getting insights and understanding better the company's picture that is drawn with diverse pieces of information. Many of the most important sources of big data are relatively new. Big Data in the form of messages, updates and images posted on social networks, records from sensors, GPS signals from mobile phones and others, needs to be collected and appropriately integrated.
- Velocity - the term "velocity" is considered in the context of how fast the data is generated and processed in order to meet the needs and challenges that companies should focus on the way of their growth and development. For many applications, the speed for data creation is more critical and

important than the volume. Real-time or near real-time produced information enables companies to be flexible in opportunities analysis and problems solving, creates competitive advantage among its competitors.

Within the experience and expansion of Big Data philosophy additionally new Vs are introduced by different authors in order to better focus on the specificity on the growing data challenge.

Very important appeared to be the fourth V - *veracity*. This characteristic is addressing the accurateness of data and preciseness in predicting the business value. The requirement within this characteristic refers the results of the big data analysis and if the insights are accurately derived. It is very critical this characteristic to be not underestimated. Data should be verified in both aspects - accuracy and context. Businesses need to be supported by technologies that are able to analyse massive amounts of data in real-time and quickly assess the value of opportunities for gaining additional competitive advantages. In this respect, it is extremely necessary to identify the right scope and types of data to be analysed and considered for decisions that would impact the business outcomes.

Organisations should be well aware of the different aspects that emerge when developing a big data strategy and consider the following additional V's that are becoming important:

*volatility* – considers the period that data should be stored;
*validity* - the data should be correct and accurate for the intended usage;
*viability* - uncovering the latent, hidden relationships among the variables;
*variability* - means that the meaning is changing and is very relevant in performing sentiment analyses;
*visualization* - means complex graphs that can include many variables of data to become understandable and readable;
*value* - all the available data to create a lot of value for organisations, societies and consumers.

The substance of Big Data varies depending on the organization's ability to manage the big data sets and the capabilities of the applications to process and analyse all relevant data received from different sources. Big Data is a constantly growing approach – something that is considered "Big" today, could not be so big in a few years.

## 3.Managing Big Data Analytics

What's most important about big data analytics is the new mindset that is emerging. Most data is considered as something to be explored by anyone who needs to explore it. The management is moving away from a limited view of the concrete own business data. Moreover, the analytical models provide better results because the data comes from different sources and is more complete.

Big data analytics is the process of examining large data sets containing a variety of data types to identify hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to support management in different business perspectives - competitive advantages, revenue opportunities, effective marketing, better customer service, improved operational efficiency, and other benefits.

The primary goal of big data analytics is to help companies in making informed business decisions by enabling data analytics professionals to analyze large volumes of

transaction data, and other forms of data that is not embraced by conventional business intelligence (BI) programs. That could include Web server logs and Internet clickstream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call records and machine data captured by sensors connected to the Internet of Things. There are several types of Big data analytics:

- Basic analytics for insight – reporting, simple visualization, basic monitoring;
- Advanced analytics for insight – complex analysis performed by predictive modeling and other pattern matching techniques;
- Operationalised analytics – analytics is incorporated with the business processes;
- Monetized analytics – analytics is directly monitoring the revenue.

*Basic analytics* can be used to explore the data, and support discovering the valuable business aspects. This could include simple visualizations or simple statistics. Basic analysis is often used when exist large amounts of disparate data. Basic monitoring is appropriate to monitor large volumes of data in real time.

*Advanced analytics* provides algorithms for complex analysis of either structured or unstructured data. It includes sophisticated statistical models, machine learning, neural networks, text analytics. Advanced analytics can be deployed to find patterns in data, prediction, forecasting, and complex event processing. Predictive modeling is one of the most popular big data advanced analytics. A predictive model is a statistical or data-mining solution consisting of algorithms and techniques that can be used on both structured and unstructured data. In the big data environment large numbers of predictive attributes across huge amounts of observations could be analyzed. Text analytics has become an important component of the big data environment. The analysis and extraction processes used in text analytics take advantage of techniques from computational linguistics, statistics, and other computer science disciplines. Text analytics is used in all sorts of analysis - predicting churn, fraud, and social media analytics. Advanced analytics could accomplish also advanced forecasting, optimization, cluster analysis for segmentation or even micro segmentation, or affinity analysis.

The main objective of *operationalized analytics* is to become part of business processes. The analytics model along with some decision rules could be included in the company's business processes to flag critical situations, problems or increased opportunities. The business agents might not even know that a predictive model is working behind the scenes of the process performance and according to the activities results to be guided to a concrete action.

*Monetized analytics* can be used to optimize the business to create better decisions and drive top-line revenue. Big data analytics can be used to derive revenue above and beyond the insights. The idea is that various sources of data, such as billing data, location data, text-messaging data, or web-browsing data can be used together or separately to make inferences that could be useful. Even the model could be built in compliance with legislation and privacy policies.

## 4.Conclusion

Traditional BI is based on structured data, but these insights often fall short in predictive and indicative analytics, due to data sets that are too old or limited in scope.

Structured data is only a small portion of stored business data. Indeed, many analysts estimate that structured data is only about 5 percent of total enterprise data.

Big data analytics enables better analytical insights by integrating massive amounts of data of varying complexity, formats, and timeliness into one structured output.

Combining text, voice, streaming data, and unstructured data analytics into one structure allows businesses to integrate different views of related information into dynamic analytical models. These models support multidimensional metrics that can be leveraged with traditional analytics.

BI tools are still evolving to support the approach to big data analytics that provides better data visualization capabilities taking into account the use of near-real-time information and the widening range of structured and unstructured data.

## References

1. Big Data - What it is & why it matters (2014), Available at: http://www.sas.com/en_us/insights/big-data/what-is-big-data.html.
2. Big Data in 2014 – A Necessary Update (2014), Available at: http://www.datashaka.com/blog/big-data-in-2014-a-necessary-update.
3. The Missing V's in Big Data (2013): Viability and Value, Available at: http://www.wired.com/2013/05/the-missing-vs-in-big-data-viability-and-value/.
4. Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety", Gartner, Retrieved 6 February 2001.
5. Why The 3V's Are Not Sufficient To Describe Big Data (2014), Available at: https://datafloq.com/read/3vs-sufficient-describe-big-data/166.
6. Big Data Analytics Deep Dive (2014), Available at: http://staticworld.net/media-resource/ast-0073561_big_data_ibm
7. JISC (2011), Business Intelligence: Monitoring Performance and Planning Improvement. Available at: http://www.jisc.ac.uk/media/documents/publications/briefingpaper/2011/businessintelligence.pdf.
8. Big Data (2013), Available at: https://books.google.bg/books?id=XPkAEFXo7VgC&pg=PT170&lpg=PT170&dq=Advanc%D0%
9. Big Data Analytics, (2014), Available at: http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics
10. Wiley Taming the Big Data Tidal Wave, Finding Opportunities in Huge Data Streams with Advanced Analytics (2012)
11. Big Data Analytics Calls for Agile Minds, (2014), Available at: http://searchbusinessanalytics.bitpipe.com/fulfillment/1413483011_102
12. A survey on platforms for big data analytics, (2014), Available at: http://www.journalofbigdata.com/content/2/1/8
13. Big data survey reveals why IT is -- or isn't -- on board, (2014), Available at: http://searchdatacenter.techtarget.com/feature/ utm_medium= EM&asrc=EM_NLN_36787178&utm_campaign=20141124_Gartner%20survey%

# Version Control in the Cloud

**Emil Tsanov[1], Asen Bozhikov[2]**

[1] D. A. Tsenov Academy of Economics - Svishtov, Bulgaria
emil@uni-svishtov.bg

[2] D. A. Tsenov Academy of Economics - Svishtov, Bulgaria
asen@uni-svishtov.bg

**Abstract.** Using version control systems is an indispensable part of the software development process. A version control system is a tool for tracking the changes in source code files made by members of the development team. At the same time it is a foundation for team collaboration because nowadays it is common to see the members of a team geographically separated at different locations. This paper reviews the possibility of a software project to utilize cloud-based version control services so that developers will always have access to their source code from any location and any device connected to the Internet. Alongside this, the authors compare four of the widely used cloud services for version control – GitHub, Bitbucket, Assembla and Google code.

**Keywords.** Version control, centralized version control, distributed version control, cloud-based version control, software as a service.

## 1.Introduction

Modern software has become increasingly complex while trying to fully address the fast growing requirements of customers. The size and complexity of software projects have grown at a serious pace and now it's normal for a project to include millions of lines of source code which is written by hundreds, even thousands of software developers. In many cases the developers (alone or in teams) are scattered across different geographic locations where they work on one or more modules of the whole software system which is developed at the end of the project. All of this implies that software development is a complex process with many phases and there is a need for a tool to facilitate effective collaboration between programmers which is the key to improving the quality of the software and development time. This kind of tool is a version control system. Its use in the software development process is a necessity. It helps to maintain a history of all the changes made to source code files and at the same time allows better coordination between the developers in a team.

The two types of version control systems that are used today while developing new software are centralized and decentralized. Centralized version control systems have a central repository where all the source code files are located and their architecture is based on the client/server model. Decentralized version control systems utilize the peer-to-peer model and allow the developers to have the whole project (including a history of changes) on their local computers. The existence of an in-house version control system means that someone should configure it for using and maintaining it, and create regular backups of the repositories [6].

With the increasing rates of adoption of cloud computing worldwide [5] we should pay attention to the possibility of working with a version control system which can be provided to us as a cloud service. There are many providers which offer these kinds of

services. Some of them offer a version control system following the model of software as a service (SaaS) while others provide their customers with an entire software development environment in the cloud, thus following the model of a platform as a service (PaaS). GitHub, Assembla and Bitbucket are typical examples of the first group while Heroku and CloudForge represents the second group of providers.

## 2.Cloud-based Version Control Services

The main point of cloud services for version control is that the version control system is located on a server in the data center of the cloud provider (where the customers' repositories are saved). The installation and the maintenance of the system are undertaken by the cloud provider. Thus, the customer only uses the version control system. This eliminates the need for their own servers or IT infrastructure for maintaining this kind of system i.e. it saves money, which can be seen as a significant advantage for the company.

Together with providing access to some of the famous version control systems like Git, Subversion, Team Foundation Server, Mercurial and Perforce, cloud-based version control services usually include integration with tools for project management, issue tracking, bug tracking, code review, wikis etc. All of them are very useful in the whole software development process and their presence can be beneficial to the customers of these services [7].

One section of cloud services for version control offer access to public repositories while others offer private ones. There are also services that offer both variants. Public repositories are free to use. Using private repositories requires a certain amount of money to be paid to the cloud provider every month. The amount depends on the number of collaborators working on the project and the number of private repositories used.

Currently cloud services for version control are extensively used for open-source projects which rely on public repositories so that many contributors can join and work on the project. As an example we could mention the following projects - MySQL, Ruby on Rails, jQuery, and AngularJS.

All of the tasks connected with the maintenance of the version control system, scheduling regular backups of the repositories and the preparation and testing of a disaster recovery strategy should be undertaken by the cloud provider. However, this should be defined clearly and precisely in the service level agreement of the provider. When choosing a cloud-based solution it is always good practice to interview the provider about the security policy which is implemented. This includes information about the geographic location of the provider's data centers, the type of servers, the encryption of the data, protection against denial of service attacks, protection against cyber threats, and results of previous audits. In most cases the provider doesn't want to provide this information to customers, which is a serious obstacle when choosing a cloud solution.

It is important as to whether the cloud-based version control service is widely used or not as this works as an indicator for the size of its community. This in turn, implies that there is a possibility for solving similar problems considerably faster if the community is bigger.

## 3.Comparison of Cloud-based Version Control Services

In this part of the paper we present a comparison of some of the widely used cloud-based version control services – GitHub [3], Bitbucket [2], Assembla [1] and Google code [4]. For the purpose of our research we propose a group of eight criteria for assessing these services. They are explained below. Our summary is not comprehensive because there are

many other cloud services similar to the ones we've studied but it provides an overview of what they can offer to their users.

**a. Supported version control systems**

Version control in the cloud means that every provider should support at least one version control system. It is important exactly which centralized or decentralized version control system is offered because usually the users choose the one that they know and that they've used before the relocation of the project's repositories to the cloud.

**b. Repository type**

Depending on the publicity of the repository's code and the level of public accessibility to the project, cloud-based version control services offer access to public and/or private repositories. Public repositories are open and publicly accessible to anyone while private ones are used for developing copyrighted software.

**c. Plans and pricing**

As the name implies it is connected with the available plans for using a cloud-based version control system. Important parameters which are included in every plan and determine its price are: the number of repositories, the number of projects, the number of collaborators (users working on the project), the size of cloud disk space etc.

**d. Bug and issue tracking**

The main purpose of bug and issue trackers is to manage the issues (tickets) related to the project. Usually an issue is related to defects, changes and requests for support. These systems are crucial because they provide a possibility for the public dissemination of information about the occurrence of a software bug and its further debugging, and also keeps track of the software development process with the many issues created by customers and resolved by programmers.

**e. Code review tools**

These tools are especially useful for developers and they are mostly related to the method of representing code using a suitable user interface. Code review tools have features such as presenting a block of code, hierarchical structures, nesting, the designation of branches, tracking and comparing commits in the repository and illustrating the difference between them. Some cloud services for version control offer embedded applications while others provide integration with third party tools for code review.

**f. Third-party tools integration**

Most of the cloud providers which offer version control systems enable integration with third-party tools, plugins and bridging modules. In this group of tools we can mention collaboration tools, applications for group communication, synchronous chat, issue tracking, reaction time management, continuous testing and deployment, wikis and many more.

**g. Security**

This is the most important part of all cloud services because customers will upload their projects in the cloud data center. That's why they need to be informed about the type of authentication and communication protocols which are used, and the possibilities for: role-based permission management, group-based permissions, IP-based restriction, restrictions on different physical or logical levels of connectivity, maintaining log files, archive, backup and disaster recovery.

**h. Desktop/mobile applications**

The presence of desktop or mobile applications which allow remote access to the hosted version control service is an advantage. This is mostly regarding the possibility of mobile access to cloud-based version control service functionality and viewing files in the repository, reviewing the history of commits, the communication between team members etc.

Table 1 Comparison of Cloud-based Version Control Services

| Criteria | GitHub | Bitbucket | assembla | Google code |
|---|---|---|---|---|
| **Supported version control systems** | Git Subversion | Git Mercurial | Git Subversion Perforce | Git Subversion Mercurial |
| **Repository form** | Public/Private | Public/Private | Public/Private | Public |
| **Plans and pricing** | Public repositories are free; Private repositories start from 7$/month | Public repositories are free; Private repositories are free for up to 5 users; For more than 5 users the price starts from 10$/month | Free for 1 project with 1GB of storage For 2 projects the pricing plans start from 24$/month | Public repositories are free |
| **Bug and issue tracking** | Yes | Yes | Yes | Yes |
| **Code review tools** | Yes | Yes | Yes | Yes |
| **Third-party tools integration** | Collaboration tools, tools for planning and monitoring projects, tools for testing and deploying applications | Tools for tracking application errors, tools for building, testing and deploying applications | Tools for tracking application errors, tools for test management, collaboration tools, own messaging system | Collaboration tools like wikis |
| **Security** | SSL, HTTPS, SSH for data transmission; Two-factor authentication | HTTPS, SSH for data transmission; Encryption on data links between centers | HTTPS, SSL, IP restriction; VPN for server access | HTTPS |
| **Desktop/mobile applications** | GitHub for Windows and Mac; GitHub for Android (only keeps track of the project) | SourceTree for Windows and Mac | No | No |

Table 1 provides the results from the comparison. It is obvious that GitHub seems the preferred service for public (open-source) projects. Reasons for this are: the support of a mostly used [8] centralized (Subversion) and decentralized (Git) version control system, the presence of two factor authentication, the presence of desktop and mobile applications for connecting and working with the cloud repository. According to its site, GitHub is "the largest code host on the planet with over 17.9 million repositories". On the other hand Bitbucket, coming from one of the giants in this industry – Atlassian, offers free private repositories for small teams of up to five developers. As we can suppose, there is tight

integration with all the Atlassian products like JIRA (issue and project tracking software) and SourceTree (a desktop application). Assembla is free for one project with up to 1 GB of storage and has a long history of providing Subversion hosting. According to its site it is the "world's best host for Subversion". The last cloud service – Google code - is suitable exclusively for open source projects and provides a user-friendly representation of the source code and access to centralized and decentralized version control systems.

## 4.Conclusion

There are many issues to consider when it comes to choosing a cloud-based service provider which offers a version control system. First of all there is a need for a clear understanding of the advantages and disadvantages of a cloud-based solution and how much it is going to cost compared to an in-house solution. Security in the cloud is a topic that should be considered thoroughly. The current state of cloud-based version control shows that these services provide a great opportunity for an open-source project to grow faster and attract new contributors more easily. Although most of the hosted services offer enterprise/organization plans with private repositories and varying permissions, it is still not common for companies to migrate to cloud-based version control because they do not have high confidence in the security of cloud-based systems.

## References

1. Assembla. https://www.assembla.com/
2. Bitbucket. https://www.bitbucket.com/
3. GitHub. https://www.github.com/
4. Google code. https://code.google.com/
5. IDG Enterprise Cloud Computing Study 2014. (2014). [Online]. Available: http://www.idgenterprise.com/report/idg-enterprise-cloud-computing-study-2014
6. Kemper C., Oxley I. Foundation Version Control for Web Developers. APress, 2012.
7. Lanubile, F., Ebert C., Prikladnicki R., Vizcaino A. Collaboration Tools for Global Software Engineering. Software, IEEE, Vol. 27, Issue 2, 2010, pp. 52-55.
8. Skerrett I. Eclipse Community Survey 2014. (2014). [Online]. Available: http://www.slideshare.net/IanSkerrett/eclipse-community-survey-2014 [1.11.2014].

# Knowledge Extraction by Analyzing Student Data with Business Intelligence and Data Mining Tools

**Dorina Kabakchieva[1], Kamelia Stefanova[2]**

[1] UNWE, 1700 Sofia, Bulgaria
dkabakchieva@unwe.bg

[2] UNWE, 1700 Sofia, Bulgaria
kstefanova@unwe.bg

**Abstract.** Universities today are operating in a very competitive environment and making great efforts to attract students. The decreasing number of potential student candidates in Bulgaria is due to different factors, including demographic problems, the existence of many universities and the availability of a variety of new educational offers, the brain drain effect. The main purpose of this paper is to reveal the strengths and possibilities of using Business Intelligence and Data Mining tools for analyzing student data and extracting valuable knowledge to support decision making at the University of National and World Economy.

**Keywords.** Business Intelligence, Educational Data Mining

## 1.Introduction

Universities all over the world today are operating in a very complex and dynamic environment. Globalization processes and the rapid development of information and communication technologies have led to very strong competition not only between companies, but also between higher education institutions. Leading universities in the USA, Australia and Western Europe have already experienced the effects of those changes. In order to meet the challenges, they are making much more efforts to attract better students, organizing wide marketing campaigns, offering new attractive bachelor and master programmes, providing interesting PhD and post-doctorate research opportunities, flexible education possibilities, etc. However, the successful implementation of those activities requires innovative approaches. Therefore, universities now realize that to remain competitive and grab new opportunities, they need to be able to take quickly informed and competent managerial decisions. Advanced analytical technologies, including Business Intelligence (BI) systems and Data Mining tools, which are already widely used by companies in all industry sectors, have also attracted the attention of educational institutions' managers. Many universities have already introduced such systems; others are currently in the process of initiating or completing such projects. The main advantages of using BI systems include integrating data from various sources and achieving "a single version of the truth" for all organization members; receiving needed reports at the right moment, at all organizational levels, and in the best possible formats; possibilities for detailed analysis of the available data through OLAP (Online Analytical Processing) tools; ability to find patterns and predict outcomes based on historical performance through Data Mining tools.

Bulgarian universities are also facing the challenges of globalization and new technologies. During the last few years, the number of universities that find it more difficult to attract enough students for their bachelor and master programmes increases. The reasons

for the experienced difficulties are various and complex. On one hand, there is an increased competition between the universities in the country, due to the insufficient number of potential candidates, resulting from the demographic crisis and the brain drain. On the other hand, the Bulgarian universities have to compete with foreign universities as well, organizing massive marketing campaigns and providing attractive offers for high quality education and good perspectives for successful future realization. University managers in Bulgaria are now confronted with the necessity of quickly taking important decisions. They could be supported in their strategy definition and fulfillment by implementing advanced analytic technologies like Business Intelligence and Data Mining, following the experiences and good practices of universities worldwide.

## 2.Business Intelligence Systems at Universities Worldwide

There are a lot of examples about universities worldwide which are already using Business Intelligence systems for analyzing their available data and thus supporting decision making. For the purposes of this paper, nine cases of BI implementation in universities are considered and analyzed. Four of them are European universities – three in the UK and one in Sweden, four are American universities, and one is an Australian university. The analyzed information is taken mainly from the university websites or from case studies, describing the successful realization of the IT projects at the selected institutions.

Modern universities are collecting large volumes of data referring to their students, the organization and management of the educational process, research activities and other managerial issues. Data at universities is usually collected in various IT systems and comes from both, the traditional face-to-face form of classroom education and the new forms of web-based and distance learning, and the e-learning systems. One of the most important problems that is solved by implementing a Business Intelligence system at analyzed universities, is the integration of data from the various sources and achieving the so called "single source of truth" which allows for data consistency and thus eliminating doubt. The importance of working with high quality data and the great efforts needed for achieving that are highlighted in most of the considered cases. Combining data from different systems in a single Business Intelligence solution facilitates cross-functional analyses and leads to higher quality decision making and efficiency improvement.

Most of the university Business Intelligence initiatives are focused in three main areas – for supporting administrative and management activities, academic issues and university relations. In some of the considered cases, the BI solutions initially deal with only some of the available data (e.g. only financial or HR data is analyzed) and later they expand to other functional areas. BI platforms support analysis, monitoring, key performance indicators and dashboards. The user interface is usually suitable for different kinds of users, simplifying the processes of requesting and analyzing data, providing facilities for personalization of reports in accordance with individual needs, online access to intuitive self-service reports, ad-hoc analysis tools, and key business performance metrics. The access to the BI platform is often provided through a web portal, to support all types of user interfaces such as desktop computers, laptops, and tablets.

Business Intelligence systems always support administrative and management activities. During university admission campaigns, student applications and offers are processed, enrollment trends are analyzed. The BI system is also used as a planning and monitoring tool, e.g. for modelling student number plans and then monitoring during the plan period, for planning courses, for budgeting and financial planning and then monitoring the year outcomes. Monitoring is performed on Key Performance Indicators (KPIs), e.g. Progression and Achievement (the average score of student progression and achievement),

Student Satisfaction, Research Assessment, Graduate Level Destination, International Population (the number of full-time international students as a proportion of the total full-time student population), to enable management to monitor their current performance against agreed strategic targets. Performance monitoring and decision making is improved also through the abilities to monitor daily operations, to apply scenario modelling, to analyze student lifecycle trends, to make forecasts, to evaluate facts and figures for effective ROI. That ensures the university is kept informed of emerging trends and developments, and supports the reduction of costs related to the process of management and sharing of BI data. The BI platform is also used to transform structured data into functional information by providing capabilities for generating and distributing different kinds of reports, including universities' annual performance reports.

Academic performance is also an area that is usually widely supported by the BI solutions developed at universities. Some of the most aspects that need accurate information for better decision making include better understanding of departmental loads, disciplines and academic outcomes; establishing areas that performed poorly, in order to take early measures, or well, in order to identify best practice; finding the reasons for retention and monitoring progression; analyzing institutional research performance and standing; financial contributions, etc.

University relations, and particularly alumni management, are also very often supported by the university BI systems, providing accurate information and thorough analyses helping to ensure organizational sustainability and capability, university environment and engagement. Decision making is based on dashboards, scorecards and reports, allowing for graphically sharing important data with the stakeholders.

## 3.Implementing Data Mining Methods and Tools at Universities

The implementation of data mining methods and tools for analyzing data available at educational institutions is defined as "educational data mining" (EDM) (Romero et al. 2007) and is a new stream in the data mining research field. There are already a large number of research papers discussing various problems within the higher education sector and providing examples for their successful solutions reached by using data mining.

The application of data mining in educational systems can be addressed for contribution to different stakeholders – students, educators, administrators, managers, governmental institutions (Romero, 2007; Baker, 2009). Students could be supported by being recommended different learning resources, activities, tasks or even different learning paths. Educators can get more objective feedback and insight about the educational process that could help them to improve the content of the courses, to select adequate methods for the content provision, to differentiate students, based on their needs in guidance and monitoring, taking into consideration their learning abilities and peculiarities. Administrators could benefit from getting the right information at the moment when it is needed, thus supporting the decision-making process with adequate analysis. University managers could be assisted in their strategic tasks by receiving deep multi-dimensional analysis that could reveal trends and opportunities for improving the effectiveness and efficiency of the university management.

Extensive literature reviews of the educational data mining research field are provided by Romero and Ventura in 2007 (Romero et al. 2007), covering the research efforts in the area between 1995 and 2005, and by Baker and Yacef in 2009 (Baker, 2009), for the period after 2005. Four of the areas in which Data Mining tools could be successfully implemented for university data analysis are defined more than 10 years ago. According to Luan (2002, 2004), a high potential is seen in analyzing data concerning alumni management,

institutional efficiency, targeted marketing and admission campaign management. Many authors thereafter discussed the various possibilities for using Data Mining for analyzing university data. In 2008 Delavari et al. (2008) developed guidelines for selection and application of different data mining techniques to support decision making in educational institutions. The increasing number of Data Mining applications in the educational sector is discussed by Nandeshwar&Chandhari (2009), emphasizing the support for better organization and management of admission campaigns, analyzing academic performance and web-based learning, student retention, etc. A review about the implementation of data mining methods in universities is provided by Ranjan&Randjan (2010) in 2010, focusing the attention to important problems as improving student performance and dealing with student retention, supporting students in danger of retention, course selection, more efficient distribution of lecturers, better management of alumni financial contributions and research grants, decreasing expenses, etc. New potential Data Mining applications are considered in 2011 by Kumar&Chadha (2011), including better organization of new education programmes, prediction of student enrollment at different courses and student performance, revealing of cheating at exams in electronic format, finding errors in available data, etc.

## 4.Possibilities for Using BI and Data Mining Technologies at UNWE

Universities in Europe are facing new problems in their policy development today. On one hand, they build the curricula profile based on the needs of businesses and society, on the other, they should attract the right students that will be able to perform best and fulfill the university objectives. Taking into consideration the whole changing environment of educational processes within the enlarged Europe, universities are competing strongly to identify their own uniqueness and to select the most appropriate students. Bulgarian universities are also focusing on the future priorities and understand the importance of introducing innovative approaches in collecting and processing the required data for improving the decision making processes.

The University of National and World Economy (UNWE) is one of the leading universities in Bulgaria, and it is effectively using new information and communication technologies to improve its performance. It is providing electronic content for many university courses through its online education system, online services for students and lecturers, electronic test center, etc. Two database systems are developed, maintained and constantly improved at the university, containing data about the enrollment campaigns and the academic performance of students. However, this unique data is not extensively used and thoroughly analyzed. The standard reports that are generated periodically provide valuable information about decision making, but that information is not sufficient for effective planning and making important strategic decisions about the university future.

The development of Business Intelligence applications for analyzing the available university data could lead to substantial competitive advantages for the UNWE. Some of the most important benefits could be:

- Standardized reporting across university units;
- Quick answers to important management questions through reports on demand, containing clean, accurate, secure and reliable data;
- Availability of interactive self-service reports, constructed by the user itself, without the need of asking somebody else about the needed information;
- Integrated reporting and cross system analyses – possibilities for analyzing data from a variety of systems that are used throughout the university;

- Better performance monitoring and enabled execution of strategy - through flexible reports and dashboards, providing an overview of the Key Performance Indicators and possibilities for drilling down to increased levels of detail;
- Ability to predict outcomes based on historical performance;

The implementation of Data Mining tools and techniques on selected university data could provide a better insight on the students and the university performance. The problems that are most often attracting the attention of researchers and becoming the reasons for initiating data mining projects at higher education institutions are focused mainly on retention of students, improving institutional effectiveness, enrollment management, targeted marketing, and alumni management. Knowing the students better is expected to contribute to more efficient university enrollment campaigns, to attracting the most desirable students and to improving the organization and implementation of the educational and management processes at the university. Understanding student types, based on the characteristic features of university candidates (e.g. the country regions from which the students come, the type of school where they have received their secondary education, etc.), their learning styles and preferences, etc., and targeted marketing based on the data mining models derived from the data, could lead to better academic performance of university students and to decreasing the retention rate.

## 5.Conclusion

Universities are facing serious problems in their policy development today. Taking into consideration the whole changing environment of educational processes, they are competing strongly to identify their own uniqueness and to select the most appropriate students. In order to meet the challenges and the new available opportunities, the leading educational institutions worldwide have to using innovative approaches. The implementation of advanced analytical technologies, including Business Intelligence (BI) systems and Data Mining tools and techniques, provides possibilities for extensive use and thorough analyses of available university data, thus ensuring more effective and efficient performance, better management and informed decision making, based on clean, accurate, secure and reliable data.

## References

1. Baker, R., Yacef, K. (2009). The State of Educational Data mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, Vol.1, Issue 1, Oct. 2009, pp.3-17.
2. Delavari, N., Phon-Amnuaisuk, S., Beikzadeh, M. (2008). Data Mining Application in Higher Learning Institutions. Internationa Journal of Informatics in Education, published by the Institute of Mathematics and Informatics, Vilnius, 2008, Vol. 7, No. 1, pp.31–54.
3. Infotool Data AB & Dimensional Insight Netherlands (2014). A University BI-solution: Better insight and control over operations, Easy access for all users, Reliable and accessible information. Available at: http://www.infotool.se/140528_universitet_folder_lr.pdf?cms_fileid=368e75ee96405118205c58b756944989&disposition=inline
4. JISC (2011). Business Intelligence: Monitoring Performance and Planning Improvement. Available at: http://www.jisc.ac.uk/media/documents/publications/briefingpaper/2011/businessintelligence.pdf
5. Kumar, V., Chadha, A. (2011). An Empirical Study of the Applications of Data Mining Techniques in Higher Education. International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No.3, March 2011, pp.80-84.
6. Liberty University, USA. Available at: http://www.liberty.edu/informationtechnology/it-department-info/index.cfm?PID=21842
7. Luan, J. (2002). Data Mining and Its Applications in Higher Education. New Directions for Institutional Research, Special Issue titled Knowledge Management: Building a Competitive Advantage in Higher Education, Volume 2002, Issue 113, pp.17–36.
8. Luan, J. (2004). Data Mining Applications in Higher Education. SPSS Executive Report, SPSS Inc.
9. Nandeshwar, A., Chaudhari, S. (2009). Enrollment Prediction Models Using Data Mining. Available at: http://nandeshwar.info/wp-content/uploads/2008/11/DMWVU_Project.pdf
10. Northwestern University, USA. Available at: http://www.it.northwestern.edu/business-intelligence/system.html
11. Ranjan, J., Randjan, R. (2010). Application of Data Mining Techniques in Higher Education in India. Journal of Knowledge Management Practice, Vol. 11, Special Issue 1, January 2010. Available at: http://www.tlainc.com/articlsi10.htm
12. Romero, C., Ventura, S. (2007). Educational Data Mining: A Survey from 1995 to 2005. Expert Systems with Applications 33, 2007, pp.135-146.
13. Schaffhauser, D. (21010) Florida State U Transforms Reporting with Business Intelligence Campus Technology. Available at: http://www.tonybates.ca/2010/04/27/using-business-intelligence-tools-at-florida-state-university/
14. University of Central Lancashire (UCLan), UK. Available at: http://www.uclan.ac.uk/about_us/bird.php
15. University of Exeter, UK. Available at: http://as.exeter.ac.uk/it/systems/businessintelligence/
16. University of Gloucestershire – Business Intelligence Strategy. Available at: http://insight.glos.ac.uk/departments/fandp/documents/businessintelligencestrategy.pdf
17. University of Technology, Sydney, Australia. Available at: http://www.pqu.uts.edu.au/bi-program/
18. Wesleyan University , Connecticut, USA. Available at http://www.wesleyan.edu/its/services/adminsys/business_intel.html

# Interactive Virtual Avatars. Design & Application Challenges for Future Smart Homes

**ZlatogorMinchev[1,2], LubenBoyanov[1,3]**

[1]Institute of ICT / [2]Institute of Mathematics & Informatics, Bulgarian Academy of Sciences

[3]University of National and World Economy

Acad. Georgi Bonchev Str., Bl. 25A, 1113, Sofia, Bulgaria

zlatogor@bas.bg, lb@acad.bg

**Abstract.** Today, technological progress is opening vast perspectives for modern digital solutions and evolution of services. This, however, is closely related to some resulting cyber threats challenges, emerging from users' understandings and interactions with technologies. The paper is outlining some users' beliefs for future smart homes realities automation, noting the role of avatars and the resulting multiple cyber threats aspects. A description of a home automation system, providing interactive communication through virtual assisting agent – ALEX is presented. Users' preferences and engineering challenges, concerning the avatar ALEX are also noted. Results usage directions for security advancing in the new digital world are also discussed.

**Key words**. home automation, smart homes, cyber security, virtual assisting agents, avatar engineering

## 1. Introduction

Modern digital world is constantly progressing, facing a number of different technological solutions and services developments towards ubiquitous digital society. Key challenges in this process are the users' multiple services necessities in future smart homes [1].

Presently, the interactive social networks environment of communication, in addition to the 4G broadband Internet multimedia access, are producing a number of challenging virtual and augmented realities combinations with possible smart homes applications.

These however are generating a significant set of unrevealed cyber threats [2], [3], [4] for both users and smart homes (e.g. social engineering, privacy data control or even malware and targeted attacks towards new smart infrastructure).

An important point here is also the development of virtual assistants with different applications. It can be noted that this concept is also related to soft bot agents' embodiment, or "avataring" [5] regarding their appearance and character.

A special focus on the topic is the interactive digital assistance [6],used as a convenient (intelligent) human-machine interfacing, closely related to Building Automation Systems [7] and due to its mobility control capabilities – to Internet of Things [8], combining virtual and physical objects.

The paper presents two preliminary questionnaire based studies on:(i) modern users' beliefs for smart homes automation progress as new virtual/augmented digital realities, noting possible robots and avatars application;(ii) experts' multifaceted evaluation towards resulting cyber threats in future smart homes as element of the new digital world.

Further on, an experimental approach for engineering smart home, voice controlled virtual assisting agent– ALEX is considered. Different aspects of users' preferences on the avatar ALEX: appearance, character and voice for achieving better realism are also studied.

Results usage directions for security advancing in the new digital world are also discussed.

## 2. Trends of users' beliefs and cyber threats for future smart homes

An initial questionnaire based (q-based) survey amongst 180 students from University of National & World Economy – Sofia was performed to outline the boundaries of the technological developments and potential cyber threats. Their beliefs in five years' time horizon, concerning smart homes automation, new digital realities integration, robots and avatars usage were considered. Generalization of this survey results is given in Figure 1.

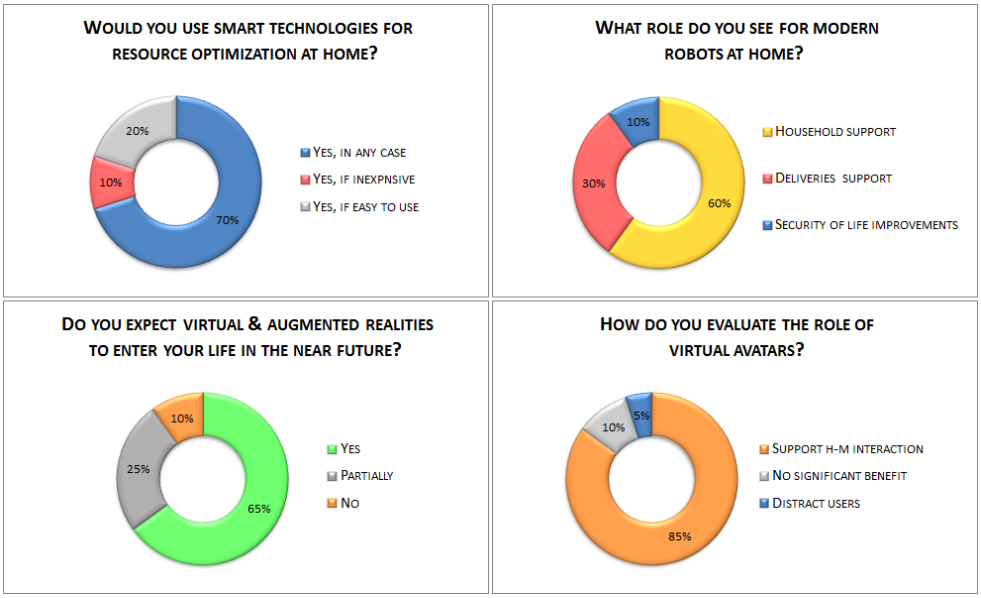



Figure 1. Q-based survey results about modern users' beliefs for smart homes automation, new digital realities integration, robots and avatars usage amongst 180 participants.

The participants' beliefs for the next five years are assuming some70% usage of smart technologies for resource optimization of future smart homes. The majority of innovations, concerning home robots applications are related to household support - 60 %. These are combined with the broader integration of digital realities (virtual & augmented) in the everyday life - 65 %, together with the expected Internet of Things technological boom [9]. As smart homes new technologies are related to new interfaces, the implementation of avatars in the human-machine interaction is expected to grow up too - 85 %.

This technological trends progress is related to a number of resulting cyber threats. Being emerging technologies as a whole, another recent study for their connectivity with web technologies is outlining their high significance for the near future and at the same time define uncertainties, influencing the next 30 years [10].

The negative influence towards way of life, incorporating stress and overweight, accentuating on communication via mobile smart phones in social networks is another trend noted recently [3].

These, provoked a multifaceted ("Human Factor", "Digital Society", "Governance", "Economy", "New Technologies", "Environment of Living") q-based survey on smart homes regarding different cyber threats (*Targeted Attacks*, *Compromised Devices*, *Malware*, *Technologies Influence*, *Privacy & Alienation*) in the digital world. The threats were harmonized with Red Book [2] recent discoveries, concerning cyber security for the next ten years.

The survey incorporated 75 national and international experts in the cyber security area (see Figure 2), interviewed in the framework of EU SysSec system security network of excellence [11] and DFNI T01/4 smart home [12] research projects.

| *Threat/Area* | Human Factor | Digital Society | Governance | Economy | New Technologies | Environment of Living |
|---|---|---|---|---|---|---|
| Targeted Attacks | | | | | | |
| Compromised Devices | | | | | | |
| Malware | | | | | | |
| Technologies Influences | | | | | | |
| Privacy & Alienation | | | | | | |

Figure 2. Q-based survey results about multifaceted evaluation towards resulting cyber threats in future smart homes as an element of the new digital world amongst 75 experts.

Three-level evaluation scale ("red" – severe, "yellow" – high and "blue" – uncertain) for cyber threats was adopted.

The results are nominating horizontally: *Privacy & Alienation* as highest cyber threat, together with *Targeted Attacks*. The rest of the threats are rather uncertain towards the evaluated facets. The most critical facet is the "Human Factor" and the most uncertain – "Digital Society".

This places the future smart homes as a critical infrastructure with an important multirole of their inhabitants. One can draw the conclusion that this is a rather important component of human-machine interaction and posts the significance of virtual assistants as a new technological trend [13].

These preliminary studies results are just outlining the technological progress of smart homes automation and the resulting cyber threats as users' beliefs. The challenge here appears to be finding a reliable validation framework and focus.

As outlined in [3], a scenario based approach and test-bed environment constructive validation could be used. A natural follow-up, concerning the interface design – namely - an avatar interface of a smart home automation system is proposed. Our virtual assisting agent– ALEX will be further noted and discussed.

## 3. Experimental engineering of smart home automation system with virtual assisting agent

The idea for smart home automation system engineering is rather intuitive and there are a number of prototypes [13] but its practical implementation could be quite complex and unique, in accordance with the environment and the level of automation.

Specific moments in such system design are the human-machine interface, agent intelligence organization and multiple communication channels arrangements.

The presented concept here (Figure 3) is the implementation of user interface trough virtual assisting agent with human based avatar, voice communication, together with multiple channel device (actuators) controlling and limited level of artificial intelligence decision logic, using "condition-action" agent-environment communication [14].

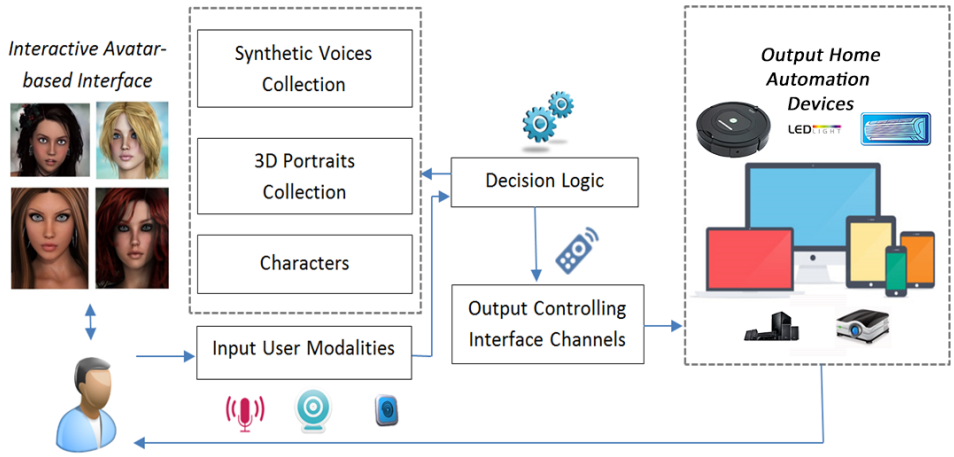



Figure 3. Engineering concept for smart home automation system design with interactive avatar based interface implemented in virtual assisting agent.

As seen in Figure 3, the system encompasses four basic modules: (i) input avatar-based interactive interface (a synthetic combination of voice, face, character and resulting behavior features); (ii) decision logic (a rule-based "condition-action" system that practically generates the agent intelligence); (iii) output controlling interface channels (like: Wi-Fi, Bluetooth, ZigBee, IrDA, ANT, etc.), (iv) output controlling devices (like: smart TVs, projectors, audio systems, home cinema\theatre solutions, home robots, lighting, conditioning, etc.). The input-output system connectivity uses human-in-the-loop feedback. Here is used voice control as input modality, but other users' activities (e.g. gestures, face emotions or even biometric body signals [15]) could also be added.

*Experimental Prototyping*

The real prototyping of the concept from Figure 3 was experimentally implemented by using a combination of commercial-off-the-shelf solutions, together with own created software and hardware modules. The working environment was MS Visual Studio 2010 Express©. Voice control, provided from a Plantronix© ML18 hands-free system and Bluetooth communication was used as input modality.

The avatars interface was designed by combining four DAZ Studio© models (blond, red, brown and black) with four IVONA© artificial voices (Salli, Emma, Amy, Kimberly). The working environment for avatars experimental study was CrazyTalk© Pro v 7.11 and the standard built-in flirting character. The SONY© KDL-32HX750 3D Smart TV was used as output multimedia home device.

The input digital assistant avatars were shown on a separate transparent screen with holographic effect via a SONY© Vaio Pro 13 ultrabook and Epson EH-TW5200 multimedia projector.

The control was performed via the SONY IR remote channel (using preliminary commands recordings, translated through the headphone or USB port and Arduino IR

dongle) and programming MS Windows© Media Player 12 functionalities (including: volume stepwise up/down level change, screen size – normal vs. full, track list control). The working environment was under MS Windows7/8.

Voice recognition and synthesizing of user input commands and avatars speech were organized with MS SAPI 5.3, defining own grammar rules set. The decision logic implements own rule-based system with fuzzy sets application for intuitive uncertainty copying and effective ambient noise suppression similar to [16].

As one of the key problems in creating such system is basically related to avatars' realistic development, a more detailed description of the process is given bellow.

*Avatar Users' Evaluation*

Generally, the creation of a realistic artificial avatar interface is quite comprehensive design task. The present study used a user-based evaluation.

The process organization covered two basic stages: (i) initial participants' current emotional state evaluation, (ii) rating via questionnaires selected input 3D models and synthetic voices collections.

Two focus groups of 57 participants were used: 42 young generation (Group I – high-school students –18 male and 24 female at an average age of 17.5 years) and 15 adult researchers (Group II – 8 female and 7 male at an average age of 45 years).

The initial emotional state was evaluated [17] via an instruction for choosing a photography (Object A – garden, Object B – fireworks, Object C – desert, Object D – seal specimens) and melody sound (Sound 1 – carousel, Sound 2 – buzzing, Sound 3 – robin, Sound 4 – country night specimens) from the databases of the Center for the Study of Emotion and Attention, University of Florida IAPS/IADS [18].

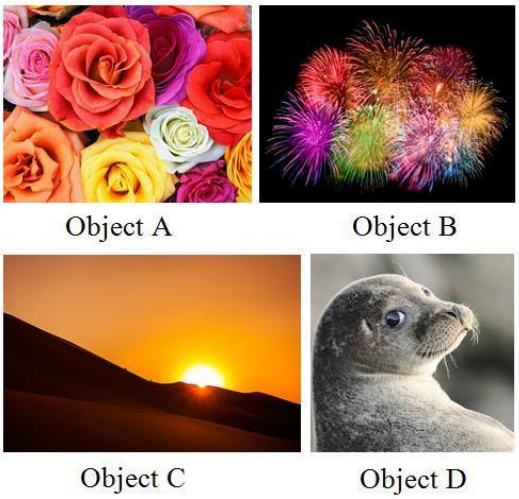



Figure 4. Emotional image stimuli collection replicas to IAPS.

The studied subjects produce the overall positive emotional state, choosing mostly Object D (95 %, VAR = 0.3) and melody Sound 3 elements (98 %, VAR = 0.2).

The second task of the groups was to put together four DAZ Studio©3D models' faces (Object1, Object 2, Object 3 and Object 4, see Figure 5) and IVONA voices (Voice 1 – Amy, Voice 2 – Emma, Voice 3 – Kimberly, Voice 4 – Salli) [17]. They were pronouncing one and the same word sequence with normal speech speed, equal duration, volume level, typical flirting character and mouth-to-text synchronization.

We should note here some other similar studies on the problem, referring users' predisposition towards: trustworthiness, competence, dominance, extroversion, regarding human face and voting [19], which are partially addressing the avatar design problem.

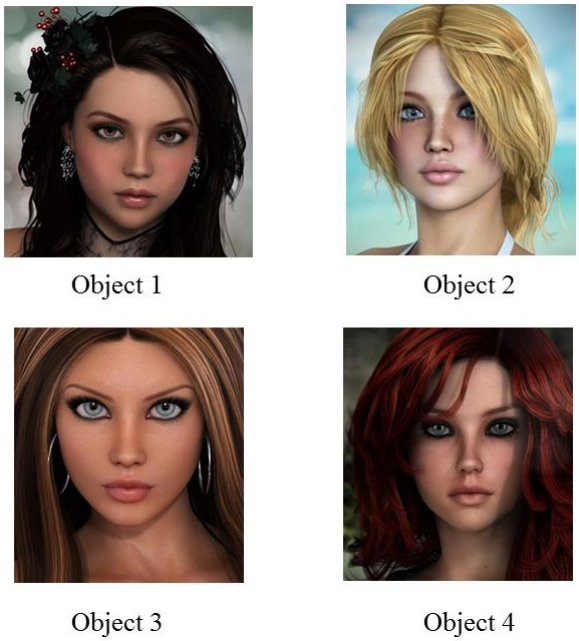



Figure5. Implemented avatars DAZ Studio©3D face models.

The averaged results for the two groups have shown a dominated preference for Voice 4 vs. Object1 for Group I (Figure 6a) and Voice2 vs. Object 1 for Group II (Figure 6b).

The most common avatar combination selection from this survey was: Object 1 and Voice 4 combination and it was named "ALEX"– Avatar Live EXample.

A live video demo of ALEX virtual assisting agent is available on-line at [20].

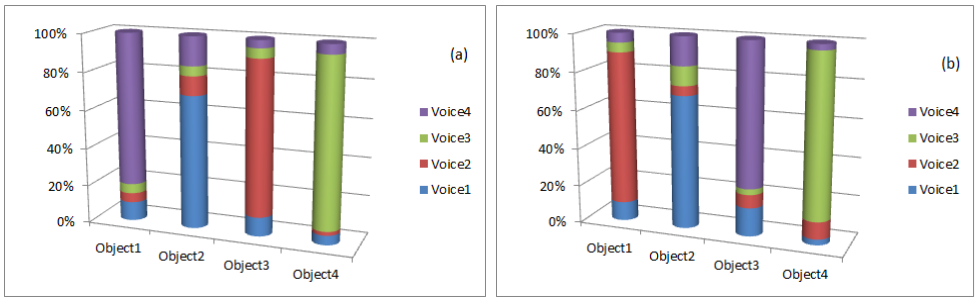



Figure 6 .Averaged results for avatars faces and voices combinations for Group I (a),VAR=0.35 and Group II (b),VAR = 0.45.

## Discussion

Our current research efforts are showing some promising results, concerning smart home automation and integration of interactive virtual assisting avatars.

The practical development however requires design, interface communications and intelligence balancing.

Additional sensors information has to be considered for further integration, following the Internet of Things fast progressing trends.

The resulting future smart homes potential cyber threats challenges are also progressing with the new avatar interface.

Special accent has to be given to social engineering, as avatars interface could produce overtrust in certain situations from regular daily life in the future smart homes.

Another focus for cyber security advancing is related to new critical infrastructure looming and the natural communication channels raised security demands, facing future environment of living technical evolution.

These new directions, related to both – social engineering and critical infrastructure protection will produce cyber security advancing, concerning future smart homes.

## Acknowledgement

The authors express gratitude for the technological support to: "A Feasibility Study on Cyber Threats Identification and their Relationship with Users' Behavioural Dynamics in Future Smart Homes", Research Grant, funded by Fundamental & Applied Scientific Research in Priority Fields, Bulgarian Science Fund, Ministry of Education Youth and Science, 2012-2015, DFNI-T01/4,www.smarthomesbg.com.

A special appreciation for the context definition and q-based survey maintenance is given to: Joint Training Simulation & Analysis Center team and EU Network of Excellence in Managing Threats and Vulnerabilities for the Future Internet – SysSec, FP7 Grant Agreement No. 257007, 2010 - 2014, www.syssec-project.eu.

## References


1. Boyanov, L. Modern Digital Society, Lik Publishing House, Sofia, 2014 (in Bulgarian)
2. Markatos, E., Balzarotti, D., Minchev, Z., et al, The Red Book. Roadmap for Systems Security Research, SysSec Consortium, September, 2013, http://www.red-book.eu/ [Online]
3. Minchev, Z. and Boyanov, L. Smart Homes Cyberthreats Identification Based on Interactive Training, ICAICTSEE – 2013, July, 72 – 82, 2014, http://icaictsee-2013.unwe.bg/proceedings/ICAICTSEE-2013.pdf [Online]
4. Boyanov, L., Minchev, Z. Cyber security Challenges in Smart Homes, In Proceedings of NATO ARW "Best Practices and Innovative Approaches to Develop Cyber Security and Resiliency Policy Framework", Ohrid, Macedonia, June 10-12, 2013, Published by IOS Press, NATO Science for Peace and Security Series - D: Information and Communication Security, 38, 99-114, 2014
5. Faloon, G. Using Avatars and Virtual Environments in Learning: What do They have to Offer?, British Journal of Education Technology, Vol. 41, No.1, 108 – 122, 2010
6. Reeves, B., The Benefits of Interactive Online Characters, Stanford University, http://www.sitepal.com/pdf/casestudy/Stanford_University_avatar_case_study.pdf [Online]
7. Brambley, M., et al. Advanced Sensors and Controls for Building Applications: Market Assessment and Potential R&D Pathways, U.S. Department of Energy, Contract DE-AC05-76RL01830, 2005, http://apps1.eere.energy.gov/buildings/publications/pdfs/corporate/pnnl-15149_market_assessment.pdf [Online]
8. Höller, J., Tsiatsis, V., Mulligan, C., Karnouskos, S., Avesand, S., Boyle, D. From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence, Elsevier, 2014
9. Balzarotti, D. (Editor) Final Report on Threats on the Future Internet: A Research Outlook, SysSec Consortium, September, 2014, http://www.syssec-project.eu/m/page-media/3/syssec-d4.4.pdf [Online]
10. Minchev, Z. Security of Digital Society. Technological Perspectives and Challenges, In Proceedings of Jubilee International Scientific Conference "Ten Years Security Education in New Bulgarian University: Position and Perspectives for the Education in a Dynamic and Hardly Predicted Environment" Sofia, Planeta – 3 Publishing House, 438-444, 2013 (in Bulgarian)
11. SysSec Project Web Page, http://www.syssec-project.eu [Online]
12. DFNI T01/4 Project Web Page, http://smarthomesbg.com [Online]
13. Automated Avatar Is Home's Personal Electronic Greeter, Smart Home, April, 2013, http://www.electronichouse.com/daily/smart-home/an-avatar-is-born/ [Online]
14. Russel, S., Norvig, P. Artificial Intelligence: A Modern Approach, 3rd Edition, PEARSON, 2009
15. Ioanidis, S., Stamatogiannakis, M., Petsas, T. (Eds.) Advanced Report on Cyberattacks on Lightweight Devices, SysSec Consortium, September, 2013, http://www.syssec-project.eu/m/page-media/3/syssec-d7.3- CyberattacksLightweightDevices.pdf [Online]
16. Minchev, Z. An Intuitionistic Fuzzy Sets Application in Infrared Object-Reflecting Sensors of a Mobile Robot, Notes on Intuitionistic Fuzzy Sets, Vol. 10, No.4, 82 – 85, 2004
17. ALEX Avatar Audio Media Collection, http://snfactor.com/snfactor/sites/files/ALEX%20Avatar%20Audio%20Media%20Collection/ALEX%20Avatar%20Audio%20Media%20Collection.zip [Online]
18. Center for the Study of Emotion and Attention, University of Florida IAPS/IADS Data Bases, http://csea.phhp.ufl.edu/media.html [Online]
19. Olivola, Ch., Funk, F., Todorov, A. Social Attributions from Faces Bias Human Choices, Trends in Cognitive Sciences, Vol. 18, Issue 11, 566–570, November 2014
20. ALEX Video Demo, http://snfactor.com/snfactor/sites/files/clips/Alex_intro.wmv [Online]

# Algorithm for Quick Numbering of Large Volumes of Data

**Krassimira Ivanova**[1]

[1] UNWE, 1700 Sofia, Bulgaria
krasy78@mail.bg

**Abstract.** An original algorithm for numbering large datasets by means of Natural Language Addressing (NLA) is presented in the paper. We use a counter to number different instances and store its current value in the container NL-addressed by the instance. If the instance is repeated, from this NL-address we receive its already assigned number. The algorithm is implemented in an experimental program RDFArM for storing large RDF-datasets. The provided experiments have shown that NL-access time for one instance (triple or quadruple) does not depend on number of already stored instances from the dataset. This is very important for storing Big Data.

**Keywords.** Natural Language Addressing, Big Data, Numbering Large Datasets.

## 1.Introduction

There are three main problems of so called "Big Data" connected to its volume, variety, and velocity [42]:

- The sheer volume of data being stored today is exploding. It is no longer unheard of for individual enterprises to have storage clusters holding petabytes of data;
- Enterprises must be able to analyze variety of all types of data, both relational and non-relational: text, sensor data, audio, video, transactional, etc.;
- The velocity at which data is generated has changed. In traditional processing, one can think of running queries against relatively static data. With streams computing, one can execute a process similar to a continuous query to get continuously updated results in real time.

Large unstructured or semi-structured datasets require a high level of computational sophistication because operations that are easy at a small scale - such as moving data between machines or in and out of storage, visualizing the data, or displaying results - can all require substantial algorithmic ingenuity [39].

Consider a real time stream of incoming strings of symbols with possible many repetitions of the same strings. Strings may have different size (number of symbols) which may be very long, for instance, several hundred symbols. To work with such strings, it is convenient to create a unique numeration of strings, i.e. the equal strings have to have equal numbers. It is important for the next generation of information systems. The problem we have to solve is to propose an algorithm which may create unique numeration of a stream of strings in real time, i.e. without spending time for rebuilding the indexes and recompilation of information base. This paper presents such algorithm with constant complexity for quick numbering Big Data sets.

In the next point we will outline the main approach for solving the problem. Further we will present the algorithm as well as an example of its implementation. Analysis of the results and possible future research conclude the paper.

## 2.Natural Language Addressing (NLA)

"Natural Language Addressing" (NLA) is an approach for storing large semi-structured or unstructured datasets. The idea of NLA [20; 21; 22; 23; 24; 25; 26; 27; 28; 29] consists in using the computer encoding of symbols as logical address of connected to it information stored in a multi-dimensional numbered information spaces [31; 32; 33]. This way no indexes are needed and high speed direct access is available. It is similar to the natural order addressing in a dictionary where no explicit index is used but the concept by itself locates the definition.

To illustrate NL-addressing in the multi-dimensional information spaces let see an example (Figure 1).

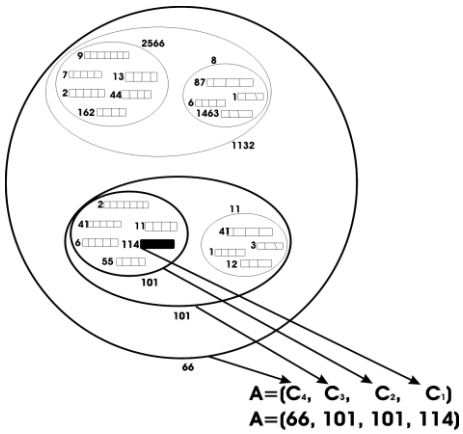



**Figure 1.** Example of a space address

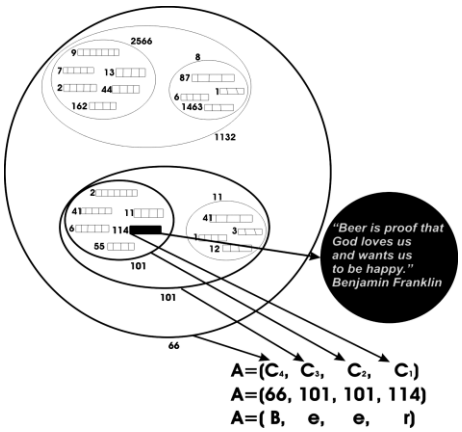



**Figure 2.** Example of Natural Language Space Address and content pointed by it

Our storing model is a hierarchy of hash tables. The leaves of the hierarchy are containers which may store basic information elements (BIE) which may be a letter, a word, a phrase of words, or, simply, string of symbols. Such container is colored in black on Figure 1 and its number is 114.

Let a set of BIE is numbered and stored in any archive (file). If we have several sets of BIE, we may number them again and store in a common archive, and so on. This way we receive a specific hierarchy of numbered sets. If one will write the sequence of numbers of the sets starting from the one which contain all others, he will create a *space address*. The space address on Figure 1 means that the space with number 66 contains the space with number 101. The numbering is unique for every set. Because of this, there is no problem to have the same numbers in the included sets what is illustrated at the Figure 1 – set 101 contains element with number 101 which is a set. Finally, the last set contains element with number 114 which is not a set but container with a string of symbols.

In other words, the space address of this string is **A = (66, 101, 101, 114)** and its content may be written or read directly using this address.

Now we may illustrate the idea of NL-addressing.

Consider the space address we just have seen – (66, 101, 101, 114).

If we assume these numbers as ASCII codes, i.e. 66 = B, 101 = e, 114 = r, we may "understand" the space address as the word "Beer" (Figure 2).

At the end, we have to illustrate the BIE (content) which may be stored at such NL-space address. It may be arbitrary long string of words. In our example we choose the BIE to

be the remarkable aphorism of Benjamin Franklin: "Beer is proof that God loves us and wants us to be happy" (Figure 2).

In this case (Figure 2) the couple {(space address A), (BIE)} is:

{(B, e, e, r), ("Beer is proof that God loves us and wants us to be happy." Benjamin Franklin)}

To access the text, we have to convert index (B, e, e, r) to index (66, 101, 101, 114) and to use corresponded access operations, i.e. we have the consequence:

Beer =>
=> (B, e, e, r) =>
=> (66, 101, 101, 114) =>
=> ("Beer is proof that God loves us and wants us to be happy." Benjamin Franklin).

## 3.Algorithm for Quick Numbering by means of NLA

Consider the problem of numerating a sequence of arbitrary words including multiple repeating of words but the same words has to be numbered with the same numbers independently of quantity of its repeating. If the length of the sequence is not so great, we may use a binary tree to store every word and its number. To find if it is already numbered we need to provide binary search and if the word is already indexed than to assign the same number.

The challenge of "Big Data" is that the sequence may be with unlimited length (several trillions) and in the same time words may come permanently during the time. In such case the static search trees could not be used because after every new word we have to reconfigure the tree. We need a new approach to solve this problem. For this purpose we may use Natural Language Addressing (NLA).

The algorithm for numbering by NLA is very simple – we need to have a counter which will count every new different word and to store its current value in the container NL-addressed by the new word. When the same word is repeated from this NL-address we will receive its already assigned number.

The ***algorithm for assigning unique numbers*** is as follow:

```
BEGIN
  counter := 1; // a counter is used, it starts from 1
  input (string); // string to be numbered
  STRNumber := NLA-read (string); // from NL-archive, using the incoming string as
path, obtain its number
  IF STRNumber = 0 // if no number exists in the container
      THAN begin
          STRNumber := counter; // assign counter value as number of string
          NLA-write (string) := STRNumber; // store number in the container located
 by the string as path
          INC (counter); // increment the counter by 1
      end
END.
```

The algorithm is illustrated on Figure 3.

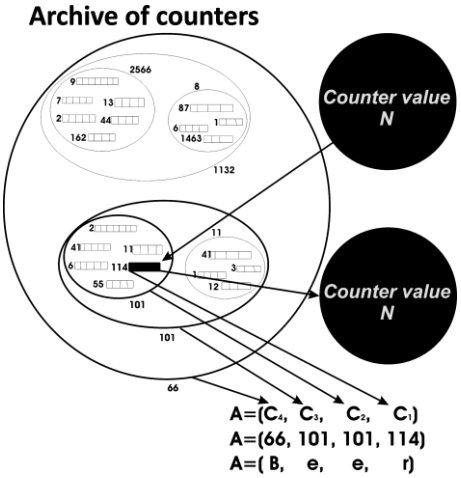



**Figure 3.** Illustration of the algorithm for assigning unique numbers

## 4.Experimental implementation of algorithm

Now, our next step is to implement our algorithm for solving real practical problem. Such very actual problem is storing Resource Description Framework (RDF) datasets. RDF is approach for representing Big Data. The primary goal of RDF is to handle large non regular or semi-structured data [35]. RDF provides a general method to decompose any information into pieces called triples [4]:

- Each triple is of the form <"Subject", "Predicate", "Object">;
- Subject and Object are the names for two things in the world. Predicate is the relationship between them;
- Subject, Predicate, Object may be given as URI's (stand-ins for things in the real world);
- Object can additionally be raw text.

The power of RDF relies on the flexibility in representing arbitrary structure without a priori schemas. Each edge in the graph is a single fact, a single statement, similar to the relationship between a single cell in a relational table and its row's primary key. RDF offers the ability to specify concepts and link them together into a graph of data [17].

The experimental implementation of proposed algorithm is illustrated on Figure 4. The main idea is to use NL-addressing for quick unique numbering of elements of triples/quadruples and after that to use these numbers as co-ordinates for storing information in the archives. In this case we have two kinds of archives: (1) archives of counters and (2) archives of values. The storing model we used is multi-layer [23]. Subjects, Predicates, Objects, and Contexts are numbered separately and these numbers are used to construct storing co-ordinates. We assume that the triple datasets contain empty context which has to be omitted.

In Figure 4 we illustrate storing of RDF-triple (beer, is, proof that...).

First we assign a number to subject: "beer".

The same we do for the relation: "is".

And after that we used these numbers as coordinates of the container where the object "proof that ..." has to be stored.

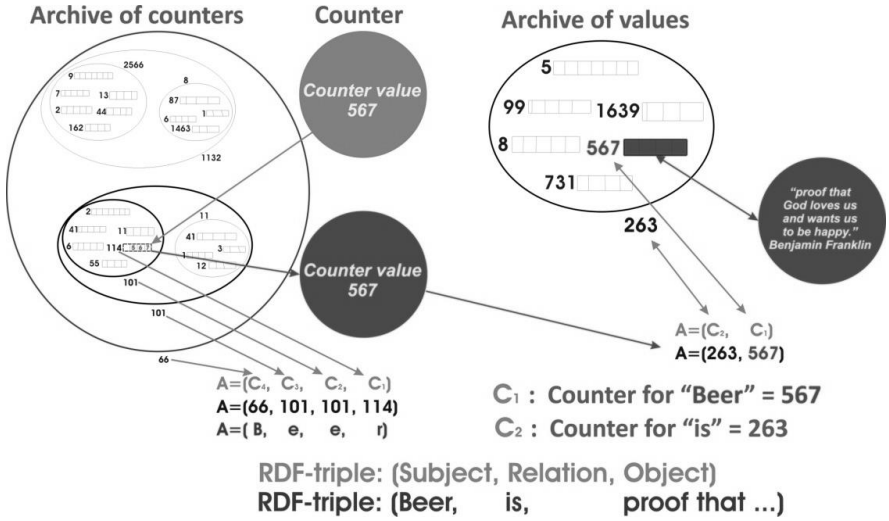



**Figure 4.** Illustration of the experimental algorithm for storing RDF triples

## 5.Experimental program RDFArM

We have realized experimental program RDFArM for storing middle-size and large RDF-datasets. Experiments were provided with both real and artificial datasets. Experimental results were systematized in [28]. We have compared RDFArM with well-known RDF-stores: OpenLink Virtuoso Open-Source Edition 5.0.2 [41], Jena SDB Beta 1 on Postgre (SQL 8.2.5 and MySQL 5.0.45) [30], and Sesame 2.0 [40], all tested by Berlin SPARQL Bench Mark (BSBM) team and connected to it research groups [1; 6; 8; 9].

We have provided experiments with:

- Middle-size RDF-datasets based on selected real datasets from DBpedia [14; 15] and artificial datasets created by BSBM Data Generator [2; 3; 5].

- Large RDF-datasets, based on selected real datasets from DBpedia's homepages [16], geocoordinates datasets [1], and Billion Triple Challenge (BTC) 2012 [12; 13]. The artificial large RDF-datasets were generated by BSBM Data Generator [5] and published in Turtle format [7; 8; 9; 10; 11], i.e. in N-quads [38]. We converted it to N-triple format using "rdf2rdf" program developed by Enrico Minack [34].

We have used the Friedman and Nemenyi tests to detect statistically significant differences between the systems [19]. The Friedman test is a non-parametric test, based on the ranking of the systems on each dataset. It is equivalent of the repeated-measures ANOVA [18]. We have used Average Ranks ranking method, which is a simple ranking method, inspired by Friedman's statistic [36]. The null-hypothesis states that if all the systems are equivalent than their ranks $R_j$ should be equal. In our case, the null-hypothesis was rejected; we could proceed with the Nemenyi test [37] which is used when all systems are compared to each other. The performance of two systems is significantly different if the corresponding average ranks differ by at least the critical difference. The Nemenyi test results for tested systems (Figure 5) had shown that RDFArM is at critical distances to Jena and Sesame. RDFArM is nearer to Jena than to Sesame. RDFArM, Jena, and Sesame are significantly different from Virtuoso [28].

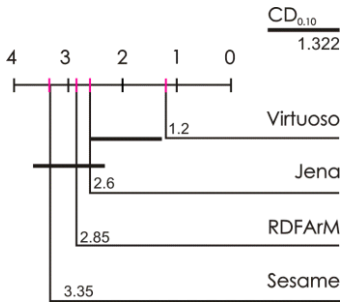



**Figure 5.** Visualisation of Nemenyi test results

All experiments had shown constant time for storing of one triple independently of the number of already stored ones. It is illustrated on Figure 6 by comparison of graphics of log(n) (black line) and average time in ms for storing one triple from BSBM 100M (gray line).

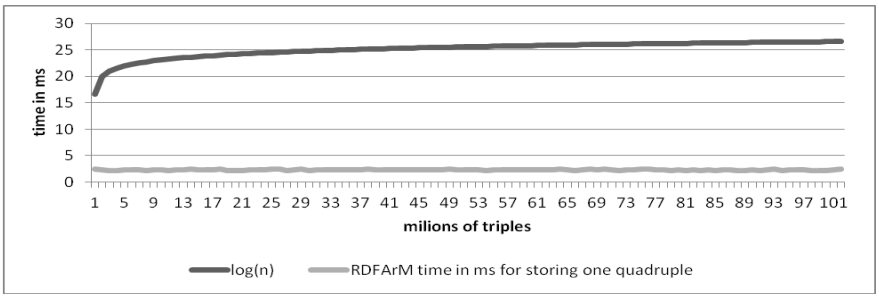



**Figure 6.** Comparison of log(n) and average time for storing one triple from BSBM 100M

## 6.Conclusion

An original algorithm for numbering by means of Natural Language Addressing has been presented in the paper. It is very simple: we have a counter, which count every new different word, and store its current value in the container NL-addressed by the new word. When the same word is repeated, from this NL-address we receive its already assigned number if such exists. In other case we assign the current value of counter to the new word.

We have realized experimental program RDFArM for storing middle-size and large RDF-datasets. We provided series of experiments to estimate the storing time. Experiments were provided with both real and artificial datasets. The main conclusion is optimistic because RDFArM is at critical distances to Jena and Sesame.

NL-access time for one instance (triple or quadruple) does not depend on number of already stored instances from the dataset. This is very important for storing Big Data.

## References

1. Christian Becker, "RDF Store Benchmarks with Dbpedia", Freie Universität Berlin, 2008, http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/ (accessed: 05.04.2013)
2. Christian Bizer, Andreas Schultz: Benchmarking the Performance of Storage Systems that expose SPARQL Endpoints; In: Proc. of the 4th International Workshop on Scalable Semantic Web knowledge Base Systems (SSWS2008), http://www4.wiwiss.fu-berlin.de/bizer/pub/BizerSchulz-BerlinSPARQLBenchmark.pdf (accessed: 31.07.2013)
3. Christian Bizer, Andreas Schultz, "The Berlin SPARQL Benchmark", In: International Journal on Semantic Web & Information Systems, Vol. 5, Issue 2, Pages 1-24, 2009, http://wifo5-

03.informatik.uni-mannheim.de/bizer/pub/Bizer-Schultz-Berlin-SPARQL-Benchmark-IJSWIS.pdf; see also http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/ (accessed: 31.07.2013)

4. Mario Briggs "DB2 NoSQL Graph Store", What, Why & Overview, A presentation, Information Management software IBM, 2012, https://www.ibm.com/developerworks/mydeveloperworks/blogs/nlp/resource/DB2_NoSQLGraphStore.pdf?lang=en (accessed: 01.12.2012)
5. Data Generator and Test Driver, In: Berlin SPARQL Benchmark (BSBM) - Benchmark Rules, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/spec/BenchmarkRules/index.html#datagenerator (accessed: 31.07.2013)
6. Berlin SPARQL Benchmark, http://www4.wiwiss.fu-berlin.de/bizer/BerlinSPARQLBenchmark/ (accessed 09.04.13).
7. Berlin SPARQL Benchmark Results, V1, 2008, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/V1/results/index.html (accessed: 31.07.2013)
8. Berlin SPARQL Benchmark Results, V2 2008, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V2/index.html (accessed: 31.07.2013)
9. Berlin SPARQL Benchmark Results, V3, 2009, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V3/index.html (accessed: 31.07.2013)
10. BSBM Results (V5) for Virtuoso, Jena TDB, BigOWLIM, 2009, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V5/index.html (accessed: 31.07.2013)
11. Berlin SPARQL Benchmark Results, V6, 2011, http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V6/index.html (accessed: 31.07.2013)
12. Billion Triple Challenge 2012 Dataset http://km.aifb.kit.edu/projects/btc-2012/ (accessed: 16.03.2013)
13. BTC data set from Datahub, http://km.aifb.kit.edu/projects/btc-2012/datahub/data-0.nq.gz (accessed: 16.03.2013).
14. DBpedia dataset "homepages.nt" dated 2007-08-30, http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/homepages-fixed.nt.gz (accessed: 31.07.2013)
15. DBpedia dataset "geocoordinates.nt" dated 2007-08-30, http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/geocoordinates-fixed.nt.gz (accessed: 31.07.2013)
16. DBpedia dataset "infoboxes.nt" dated 2007-08-30, http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/infoboxes-fixed.nt.gz (accessed: 31.07.2013)
17. David C. Faye, Olivier Cure, Guillaume Blin. A survey of RDF storage approaches. Received, December 12, 2011, Accepted, February 7, 2012, ARIMA Journal, vol. 15 (2012), pp. 11-35.
18. R. A. Fisher, "Statistical methods and scientific inference (3rd edition)", Hafner Press, New York, 1973, ISBN 978-002-844740-7
19. Friedman, M.: "A comparison of alternative tests of significance for the problem of m rankings", Annals of Mathematical Statistics, Vol. 11, 1940, pp.86-92.
20. Krassimira Ivanova, Vitalii Velychko, Krassimir Markov. "About NL-addressing" (К вопросу о естествено-языковой адрессации) In: V. Velychko et al (ed.), Problems of Computer in Intellectualization. ITHEA® 2012, Kiev, Ukraine - Sofia, Bulgaria, ISBN: 978-954-16-0061 0 (printed), ISBN: 978-954-16-0062-7 (online), pp. 77-83 (in Russian).
21. Krassimira Ivanova, Vitalii Velychko, Krassimir Markov. "Storing RDF Graphs using NL-addressing", In: G. Setlak, M. Alexandrov, K. Markov (ed.), Artificial Intelligence Methods and Techniques for Business and Engineering Applications. ITHEA® 2012, Rzeszow, Poland; Sofia, Bulgaria, ISBN: 978-954-16-0057-3 (printed), ISBN: 978-954-16-0058 0 (online), pp. 84 – 98.
22. Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Introduction to the Natural Language Addressing", International Journal "Information Technologies & Knowledge" Vol.7, Number 2, 2013, ISSN 1313-0455 (printed), 1313-048X (online), pp. 139–146.
23. Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Introduction to Storing Graphs by NL-Addressing", International Journal "Information Theories and Applications", Vol. 20, Number 3, 2013, ISSN 1310-0513 (printed), 1313-0463 (online), pp. 263 – 284.

24. Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Storing Dictionaries and Thesauruses Using NL-Addressing", International Journal "Information Models and Analyses" Vol.2, Number 3, 2013, ISSN 1314-6416 (printed), 1314-6432(online), pp. 239 - 251.
25. Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "The Natural Language Addressing Approach", International Scientific Conference "Modern Informatics: Problems, Achievements, and Prospects of Development", devoted to the 90th anniversary of academician V. M. Glushkov. Kiev, Ukraine, 2013, ISBN 978-966-02-6928-6, pp. 214 - 215.
26. Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Storing Ontologies by NL-Addressing", IVth All–Russian Conference "Knowledge-Ontology-Theory" (KONT-13), Novosibirsk, Russia, 2013, ISSN 0568 661X, pp. 175 - 184.
27. Krassimira Ivanova, "Informational and Information models", In Proceedings of 3rd International conference "Knowledge Management and Competitive Intelligence" in the frame of 17th International Forum of Young Scientists "Radio Electronics and Youth in the XXI Century", Kharkov National University of Radio Electronics (KNURE), Kharkov, Ukraine, Vol.9,2013, pp6-7.
28. Krassimira Ivanova, "RDFArM - A System for Storing Large Sets of RDF Triples and Quadruples by means of Natural Language Addressing", International Journal "Information Models and Analyses" Vol.3, Number 4, 2014, ISSN 1314-6416 (printed), 1314-6432(online), (in print).
29. Krasimira Ivanova, "Storing Data using Natural Language Addressing", PhD Thesis, Hasselt University, Belgium, 2014
30. Apache Jena, http://jena.apache.org/about_jena/about.html (accessed: 23.03.2013)
31. Krassimir Markov, "A Multi-domain Access Method", Proceedings of the International Conference on Computer Based Scientific Research, PLOVDIV, 1984, pp. 558 - 563.
32. Krassimir Markov, "Multi-domain information model", Int. J. Information Theories and Applications, 11/4, 2004, pp. 303 - 308
33. Krassimir Markov, "Co-ordinate based physical organization for computer representation of information spaces", (Координатно базирана физическа организация за компютърно представяне на информационни пространства) Proceedings of the Second International Conference "Information Research, Applications and Education" i.TECH 2004, Varna, Bulgaria, Sofia, FOI-COMMERCE – 2004, стр. 163 - 172 (in Bulgarian).
34. Enrico Minack, "RDF2RDF converter", http://www.l3s.de/~minack/rdf2rdf/ 2010, (accessed: 31.07.2013).
35. Andrae Muys, "Building an Enterprise Scale Database for RDF Data", Seminar, Netymon, 2007.
36. Neave, H., Worthington, P., "Distribution Free Tests", Routledge, 1992.
37. Peter Nemenyi, "Distribution-free multiple comparisons Unpublished", PhD thesis; Princeton University Princeton, NJ, 1963
38. N-Quads: Extending N-Triples with Context http://sw.deri.org/2008/07/n-quads/ (accessed: 16.03.2013).
39. National Research Council, "The Mathematical Sciences in 2025", Washington, DC: The National Academies Press, USA, 2013. ISBN-13: 978-0-309-28457-8. http://www.nap.edu/catalog.php?record_id=15269.
40. Sesame, OpenRDF, http://www.openrdf.org/index.jsp http://www.openrdf.org/doc/sesame2/2.3.2/users/userguide.html#chapter-sesame2-whats-new (accessed: 01.12.2012)
41. OpenLink Virtuoso Universal Server: Documentation http://docs.openlinksw.com/pdf/virtdocs.pdf, http://virtuoso.openlinksw.com/ (accessed: 23.03.2013)
42. P.C. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, G. Lapis, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGraw-Hill Companies, USA, 2012, ISBN 978-0-07-179053-6. 166 p.

# Preferences and Prescriptive Utility Presentation: A Stochastic Approximation Approach

**Yuri P. Pavlov**

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences,
yupavlov15@isdip.bas.bg, yupavlov14@hotmail.com

**Abstract.** This paper presents an approach to evaluation of human preferences. Here are presented different aspects of stochastic utility function evaluation and its applications. The stochastic assessment is based on mathematically formulated axiomatic principles and stochastic procedures. The evaluation is a preferences-oriented machine learning procedure which reduces the "certainty effect and probability distortion". The uncertainty of the human preferences is eliminated in ways typical for stochastic programming. The stochastic utility evaluation approach is used for the development of preferences-based decision support in machine learning environments and iterative complex control design.

**Keywords**: Preferences, Utility, Stochastic Approximation, Adaptive Learning, Optimal Portfolio

## 1.Introduction

The objective of the paper is to present a strict mathematical approach for modeling and evaluation of human preferences. The productive merger of the mathematical exactness with the empirical uncertainty in the human notions is the main challenge for modeling process in complex systems. People's preferences contain uncertainty due to the cardinal type of the empirical (expert) information. This uncertainty is of subjective and probability nature. Stochastic programming and expected utility theory address decision making and machine-learning under these conditions (Fishburn 1970, Keeney 1978, 1993). The uncertainty of the subjective decision maker's (DM's) preferences could be taken as an additive noise that could be eliminated, as is typical in the stochastic approximation and machine-learning (Aizerman, 1970, Pavlov, 2005, 2013). This article deals with stochastic algorithms for assessment of expert utilities and values on the basis of the expressed preference (Pavlov 2005, 2013). They can be considered as a modeling approach within the framework of the prescriptive decision-making based on the normative mathematical tradition.

## 2.Mathematical formulations, preference relations and utility function

The evaluation method presented here rest upon the achievements of the theory of measurement (scaling), utility theory and, statistic programming. The difficulties that come from the mathematical approach are due to the probability, and subjective uncertainty of the DM's expressions and the cardinal character of the expressed human preferences. The mathematical description follows. Let $\boldsymbol{X}$ be the set of alternatives and $\boldsymbol{P}$ be a set of probability distributions over $\boldsymbol{X}$. A utility function $u(.)$ will be any function for which the following is fulfilled:

$( p \succ q , (p,q) \in \boldsymbol{P}^2 ) \Leftrightarrow (\int u(.)dp > \int u(.)dq)$.

In keeping with Von Neumann and Morgenstern (Fishburn, 1970) the interpretation of the above formula is that the integral of the utility function $u(.)$ is a measure for comparison of the probability distributions $p$ and $q$ defined over $\boldsymbol{X}$. The notation ($\succ$) expresses the preferences over $\boldsymbol{P}$ including those over $\boldsymbol{X}$ ($\boldsymbol{X} \subseteq \boldsymbol{P}$). There are different systems of mathematical axioms that give satisfactory conditions for the existence of an utility function. The most famous of them is the system of *von Neumann and Morgenstern*:

(A.1) The *preference* relation ($\succ$) is transitive, i.e. if ($p \succ q$) and ($q \succ r$) then ($p \succ r$) for all $p,q,r \in \boldsymbol{P}$**;**

(A.2) *Archimedean Axiom*: for all $p,q,r \in \boldsymbol{P}$ such that ($p \succ q \succ r$), there is an $\alpha,\beta \in (0,1)$ such that (($\alpha\, p + (1-\alpha) r) \succ q$) and ($q \succ (\beta p + (1-\beta) r)$);

(A.3) *Independence*: for all $p,q,r \in \boldsymbol{P}$ and any $\alpha \in [0, 1]$, then ($p \succ q$) if and only if
(($\alpha\, p + (1- \alpha)r$) $\succ$ ($\alpha\, q + (1- \alpha)r$)).

Axioms (A1) and (A3) are insufficient for the solution of the utility function existence. Axioms (A1), (A2) and (A3) give a solution in the interval scale (precision up to an affine transformation): (($p \succ q$) $\Leftrightarrow$ ($\int v(x)dp \succ \int v(x)dq$) $\Leftrightarrow$ ($v(x) = a u(x) + b$, $a,b \in \boldsymbol{R}$, $a>0$)). The *"indifference"* relation ($\approx$) based on ($\succ$) is defined in (Fishburn, 1970):

($(x \approx y) \Leftrightarrow \neg( (x \succ y) \vee (x \prec y) )$).

The presumption of existence of a utility function $u(.)$ leads to the "*negatively transitive*" and "*asymmetric*" relation ($\succ$). The "*negative transitivity*"

($(\neg(p \succ t) \wedge \neg(t \succ q)) \Rightarrow \neg(p \succ q)$) is equivalent to the expression (Fishburn 1970):
($(p \succ q) \Rightarrow ((p \succ r)$ or $(r \succ q)$ for $\forall\, (p, q, r) \in \boldsymbol{P}^3$)). The following proposition is almost obvious.

**Proposition 1.** The DM "*preference*" relation ($\succ$) defined by the utility function $u(.)$ is "*negatively transitive*" ( $\neg(p \succ t) \wedge \neg\, (t \succ q)$ ) $\Rightarrow \neg(p \succ q)$ ).

The following proposition outlines the transitivity of the equivalence relation ($\approx$). This property is violated in human decision making most often.

**Proposition 2.** In the case of a "*negative transitivity*" of ($\succ$) the "*indifference*" relation ($\approx$) is transitive ( $((x \approx y) \wedge (y \approx t)) \Leftrightarrow (x \approx t)$ )).

**Consequence:** Let the preference relation ($\succ$) is "*irreflexive*" ($\neg(p \succ p)$) and "*negatively transitive*". Then the "*indifference*" relation ($\approx$) is an "*equivalence*" (reflexive, symmetric and transitive).

The irreflexivity and the negative transitivity of the preference relation split the set $\boldsymbol{X}$ into non overlapping equivalence classes. The factorized set of these classes is denoted by $\boldsymbol{X}/\approx$. We need the next two definitions. A "*weak order*" is an asymmetric and "negatively transitive" relation. The transitivity of the "*weak order*" ($\succ$) follows from the "asymmetry" and the "negative transitivity". A "*strong order*" is a "*weak order*" for which $(\neg(x \approx y) \Rightarrow ((x \succ y) \vee (x \prec y))$ is fulfilled. It is proven in (Fishburn 1970) that the existence of a "*weak order*" ($\succ$) over $\boldsymbol{X}$ leads to the existence of a "*strong order*" preference relation ($\succ$) over $\boldsymbol{X}/\approx$.

*Consequently the presumption of existence of a utility function u(.) leads to the existence of: asymmetry* ($(x \prec y) \Rightarrow (\neg(x \succ y))$), *transitivity* ($(x \succ y) \wedge (y \succ z) \Rightarrow (x \succ z)$) *and transitivity of the "indifference" relation* ($\approx$).

The transitivity of the relations ($\succ$) and ($\approx$) is violated most often in practice. The violation of the transitivity of the relation ($\succ$) could be interpreted as a lack of information, or

as a DM's subjective mistake. The violation of the transitivity of the relation ($\approx$) is due to the natural "uncertainty" of the human's preference and the qualitative nature of expressions of the subjective notions and evaluations (Cohen, 1988, Kahneman, 1979, Raiffa, 1968).

There is different utility evaluation methods, all of them based on the "lottery" approach (gambling approach). A "lottery" is called every discrete probability distribution over X. We mark as $<x,y,\alpha>$ the lottery: here $\alpha$ is the probability of the appearance of the alternative $x$ and $(1-\alpha)$ - the probability of the alternative $y$. The most used evaluation approach is the assessment: $z \approx <x,y,\alpha>$ , where $(x,y,z)\in \boldsymbol{X}^3$, $(x \succ z \succ y)$ and $\alpha\in[0,1]$. Weak points of this approach are the violations of the transitivity of the relations and the so called "certainty effect" and "probability distortion", identified by Kahneman and Tversky (Kahneman, 1979) and discussed in Choen and Jaffray (Cohen, 1988). Additionally, it is vary difficult to determine the alternatives $x$ (*the best*) and $y$ (*the worst*) on condition that $(x \succ z \succ y)$, where z is the analyzed alternative. *Therefore, the problem of utility function evaluation on the grounds of expert preferences is an important one*. The violations of transitivity of the relation ($\approx$) also leads to inconsistencies in the utility assessment. All these difficulties could explain the DM behavior observed in the famous Allais Paradox that arises from the "*independence*" axiom (A.3).

The determination of a measurement scale of the utility function $u(.)$ originates from the previous mathematical formulation of the relations ($\succ$) and ($\approx$). It is accepted that ($\boldsymbol{X}\subseteq\boldsymbol{P}$) and that $\boldsymbol{P}$ is a convex set ($(q,p)\in\boldsymbol{P}^2\Rightarrow(\alpha q+(1-\alpha)p)\in\boldsymbol{P}$, for $\forall\alpha\in[0,1]$). Then the utility function $u(.)$ over $\boldsymbol{X}$ is determined with the accuracy of an affine transformation (i.e. interval scale) (Fishburn 1970):

**Proposition 3**. If ($(x\in X \wedge p(x)=1)\Rightarrow p\in\boldsymbol{P}$) and ($((q,\ p)\in\boldsymbol{P}^2) \Rightarrow ((\alpha p+(1-\alpha)q)\in\boldsymbol{P}$, $\alpha\in[0,1]$)) then the utility function $u(.)$ is defined with precision up to an affine transformation ($u_1(.) \approx u_2(.)\Leftrightarrow$ ($u_1(.)=au_2(.)+b$ , $a>0$) (in the case of utility function existence).

This property is essential for the application of the utility theory, since it allows a decomposition of the multiattribute utility functions into simpler functions (Keeney, 1993). This proposition reveals that the utility measurement scale of the utility function is equivalent to the temperature scale (interval scale).

Several non-expected utility theories have been developed in response to the transitivity violations (Kahneman, 1979, Cohen, 1988). Among these theories the rank dependent utility model and its derivative cumulative prospect theory are currently the most popular. In the rank dependent utility (RDU) the decision weight of an outcome is not just probability associated with this outcome. It is a function of both probability and the rank (the alternative) $x$. For example the RDU of the lottery ($p_1$, $x_1$; $p_2$, $x_2$; ..; $p_n$, $x_n$) is:

$$RDU = \sum_{i=1}^{n} W(p_i)u(x_i)$$

Based on empirical research several authors have argued that the probability weighting function W(.) has the inverse S-shaped form, which starts concave and then becomes convex. It is supposed that $W(p_i)= p_i+\Delta W(p_i)$. Such a function is shown in Figure 1.

## 3.Utility evaluation and stochastic approximation

Starting from the properties of the preference relation ($\succ$) and indifference relation ($\approx$) and from the weak points of the "lottery approach" we propose the following stochastic approximation procedure for evaluation of the utility function. In correspondence with **Proposition 3** it is assumed that ($\boldsymbol{X}\subseteq\boldsymbol{P}$), $((q,p)\in\boldsymbol{P}^2\Rightarrow(\alpha q+(1-\alpha)p)\in\boldsymbol{P}$, for $\forall\alpha\in[0,1]$) and that the utility function $u(.)$ exists. We define two sets:

$\boldsymbol{A}_{u^*}=\{(\alpha,x,y,z)/(\alpha u^*(x)+(1-\alpha)u^*(y))>u^*(z)\}$,

$\boldsymbol{B}_{u^*}=\{(\alpha,x,y,z)/(\alpha u^*(x)+(1-\alpha)u^*(y))>u^*(z)\}$.

The notation $u^*(.)$ is the DM's empirical utility assessment. The following proposition is fundamental for the stochastic approximation approach:

**Proposition 4.** We denote $\boldsymbol{A}_u=\{(\alpha,x,y,z)/(\alpha u(x)+(1-\alpha)u(y))>u(z)\}$. If $\boldsymbol{A}_{u1}=\boldsymbol{A}_{u2}$ and $u_1(.)$ and $u_2(.)$ are continuous functions, then is true ($u_1(.)=au_2(.)+b$, $a>0$) (Y. Pavlov 2005)**.**

The approximation of the utility function is constructed by pattern recognition of the set $\boldsymbol{A}_u$ (Aizerman, 1970, Pavlov, 2013). The proposed assessment process is a machine-learning approach based on the DM's preferences. The machine learning is a probabilistic pattern recognition procedure because ($\boldsymbol{A}_{u^*}\cap\boldsymbol{B}_{u^*}\neq\varnothing$) and the utility evaluation is a stochastic approximation with noise (uncertainty) elimination**.** Key element in this solution is Proposition 4. *The following presents the evaluation procedure***:**

*DM compares the "lottery" $\langle x,y,\alpha\rangle$ with the simple alternative z, $z\in\boldsymbol{Z}$ ("better-$\succ$", $f(x,y,z,\alpha)=1$", "worse-$\prec$", $f(x,y,z,\alpha)=(-1)$" or "can't answer or equivalent- $\sim$", $f(x,y,z,\alpha)=0$", $f(.)$ denotes the qualitative DM's answer ). This determine a learning point ($(x,y,z,\alpha)$, $f(x,y,z,\alpha)$). The following recurrent stochastic algorithm constructs the utility polynomial approximation* $u(x)=\sum_i c_i\Phi_i(x)$:

$$c_i^{n+1}=c_i^{n}+\gamma_n\left[f(t^{n+1})-\overline{(c^n,\Psi(t^{n+1}))}\right]\Psi_i(t^{n+1}), \tag{1}$$

$$f(t^{n+1})=\left[\mathrm{D}'(t^{n+1})+\xi^{n+1}\right]$$

$$\sum_n\gamma_n=+\infty,\ \sum_n\gamma_n^2<+\infty,\ \forall n,\gamma_n>0.$$

In Equation (1) are used the following notations (based on $\boldsymbol{A}_u$): $\boldsymbol{t}=(x,y,z,\alpha)$, $\boldsymbol{\psi}_i(t)=\boldsymbol{\psi}_i(x,y,z,\alpha)=\alpha\Phi_i(x)+(1-\alpha)\Phi_i(y)-\Phi_i(z)$, where $(\boldsymbol{\Phi}_i(x))$ is a family of polynomials. The line above the scalar product $\bar{v}=\overline{(c^n,\Psi(t))}$ means: ($\bar{v}=1$), if ($v>1$), ($\bar{v}=-1$) if ($v<-1$), and ($\bar{v}=v$) if ($-1<v<1$). The coefficients $c_i^n$ take part in the polynomial representation $g^n(x)=\sum_{i=1}^{n}c_i^n\Phi_i(x)$ and $(c^n,\Psi(t))=\alpha g^n(x)+(1-\alpha)g^n(y)-g^n(z)=G^n(x,y,z,\alpha)$ is a scalar product. The learning points are set with a pseudo random sequence.

*The mathematical procedure describes the following assessment process:*

The expert relates intuitively the "learning point" $(x,y,z,\alpha)$ to the set $\boldsymbol{A}_{u^*}$ with probability $D_1(x,y,z,\alpha)$ or to the set $\boldsymbol{B}_{u^*}$ with probability $D_2(x,y,z,\alpha)$. The probabilities $D_1(x,y,z,\alpha)$ and $D_2(x,y,z,\alpha)$ are mathematical expectations of $f(.)$ over $\boldsymbol{A}_{u^*}$ and $\boldsymbol{B}_{u^*}$ respectively, $(D_1(x,y,z,\alpha)=M(f/x,y,z,\alpha))$ if $(M(f/x,y,z,\alpha)>0)$, $(D_2(x,y,z,\alpha)=(-)M(f/\ x,y,z,\alpha))$ if $(M(f/x,y,z,\alpha)<0)$. Let $D'(x,y,z,\alpha)$ is the random value: $D'(x,y,z,\alpha)=D_1(x,y,z,\alpha)$ if $(M(f/x,y,z,\alpha)>0)$; $D'(x,y,z,\alpha)=(-D_2(x,y,z,\alpha))$ if $(M(f/x,y,z,\alpha)<0)$; $D'(x,y,z,\alpha)=0$ if $(M(f/x,y,z,\alpha)=0)$. We approximate $D'(x,y,z,\alpha)$ by a function of the type :

$$G(x,y,z,\alpha)=(\alpha g(x)+(1-\alpha)g(y)-g(z)), \qquad (2)$$

where $g(x)=\sum_i c_i\Phi_i(x)$.

The coefficients $c_i^n$ take part in the polynomial approximation of $G(x,y,z,\alpha)$:

$$G^n(x,y,z,\alpha)=(c^n,\Psi(t))=\alpha g^n(x)+(1-\alpha)g^n(y)-g^n(z),\ g^n(x)=\sum_{i=1}^{N}c_i^n\Phi_i(x).$$

The function $G^n(x,y,z,\alpha)$ is positive over $\boldsymbol{A}_{u^*}$ and negative over $\boldsymbol{B}_{u^*}$ depending on the degree of approximation of $D'(x,y,z,\alpha)$. The function $g^n(x)$ is the approximation of the utility function u(.). The stochastic convergence of the evaluation given by formula (1) is analyzed in (Aizerman, 1970, Pavlov, 2013).

## 4.Computational modeling

*The dialogue between the expert and the computer was simulated*, where the expert was replaced by a model function in the numeric presentation (the solid line in figure (2)). In the modeling, an additive noise determines the uncertainty in DM's answers. The expert was modeled by the seesaw lines in figure (2). The learning points $(x,s,z,\alpha)$ are set with a pseudo-random sequence. The utility approximation is shown in figure (2) with star line (*).

*Our approach permits direct assessment of the dependence of the utility function on probability.* An example of such an evaluated utility function $f_2(y,\alpha)$ is shown in figure 3. The explicit formula of the cumulative utility function $f_2(y)$ has the foolowing mode:

$$f_2(y)=\int_0^1 f_2(y,\alpha)d\alpha.$$

This function is shown in Figure 4. If the cumulative function W(.) has a symmetric form (figure (1)) then is true:

$$\int_0^1 \Delta W(p)f_2(y)dp=0.$$

In this case the expected von Neuman and Morgenstern's utility function $f_2(.)$ is exactly the integral because $p$ is evenly distributed. In this way the "certainty effect" and "probability distortion" identified by Kahneman and Tversky could be reduced.

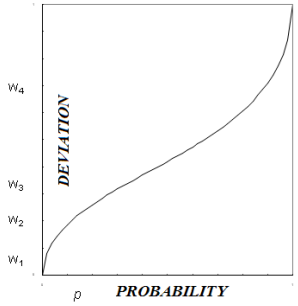



Fig.1 RDU Weigthing function W(.)

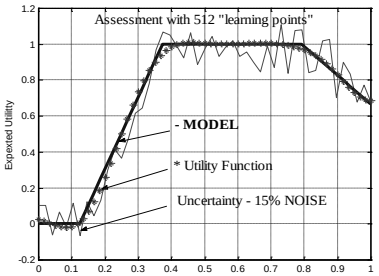



Fig.2 Modeling, 512 "learning points"

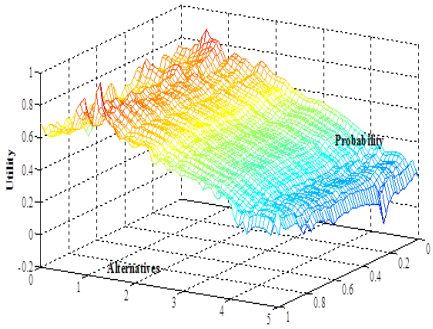



Fig. 3 Utility function $f_2(y,\alpha)$

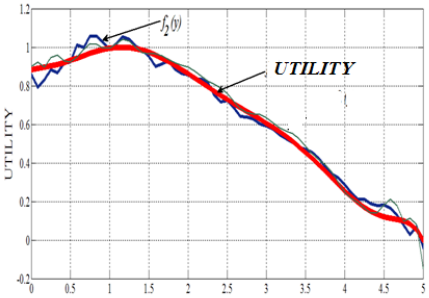



Fig. 4 Utility function $f_2(y)$

## 5.Aplications, Optimal Portfolio Utility Allocation

This problem is discussed in different scientific sources and has practical significance (Touzi, 2012). But the problem of the choice (or of the construction) of the utility function is out of discussion. In our exposition we will propose a more complex utility function for description of DM's portfolio allocation. This utility (objective) function is constructed (approximated) by the stochastic procedure described in the previous subsections and is in agreement with the DM's preferences.

We use a classical dynamic model for description of a financial market. Consider a non-risky asset $S^0$ and risky one S. Following the presentation in (Touzi, 2012) the Black-Scholes stochastic differential equation is given by:

$$dS_t^0 = S_t^0 r dt \quad and \quad dS_t = S_t \mu dt + \sigma dW_t .$$

Here $r$, $\mu$ and $\sigma$ are constants ($r$=0.03, $\mu$=0.05 and $\sigma$=0.3) and $W$ is a one dimensional Brownian motion (Touzi, 2012). By $X_t$ we denote the state space vector of the controlled dynamic process. The investment policy is defined by a progressively adapted process $\pi=\{\pi_t, t\in[0,T]\}$ where $\pi_t$ represents (defines) the amount ($X_t \pi_t$, $\pi_t \in[0,1]$) invested in the risky asset at time $t$. The remaining wealth ($X_t - \pi_t X_t$) at the same moment $t$ is invested in the risky asset. The time period $T$ is 50 weeks. The liquidation value of a self-financing strategy satisfies:

$$dX_t^{\pi} = \pi_t X_t^{\pi} \frac{dS_t}{S_t} dt + (X_t^{\pi} - \pi_t X_t^{\pi}) \frac{dS_t^0}{S_t^0} = (r X_t^{\pi} + (\mu - r)\pi_t X_t^{\pi}) + \sigma \pi_t X_t^{\pi} dW_t .$$

It is obvious that in the chosen parameters for the investment policy is true $E\int_0^T(\pi_t X_t^{\pi})^2 dt < \infty$. Here E denote mathematical expectation defined in the initial filtered probability space ($\Omega, F, F, P$) with canonical filtration F=$\{F_t, t\geq 0\}$ of the Brownian motion defined over the probability space ($\Omega, F, P$). The objective of the investor (DM) is to choose the control (the amount $\pi_t$ invested in the risky process) so as to maximize the expected utility of his terminal wealth at moment T, i.e:

$$V(t,x) := \sup_{\pi\in[0,1]} E[U(X_T^{t,x,\pi})]$$

In this formula $X^{t,x,\pi}$ is the solution of the controlled stochastic differential equation with initial condition (initial wealth) $x$ at time $t$. For the liquidation value is supposed that if the state space vector is zero in a moment $t$ then it remains zero to the end T ( $X_T^{t,x,\pi} = 0$ ). The optimal control could be determined step by step from the Hamilton-Jacobi-Bellman partial differential equation following the dynamical programming principle (Touzi, 2012):

$$\frac{\partial w}{\partial t}(t,x) + \sup_{\pi\in[0,1]}[(rx+(\mu-r)\pi x)\frac{\partial w}{\partial t}(t,x) + \frac{1}{2}\sigma^2\pi^2x^2\frac{\partial^2 w}{\partial^2 t}(t,x)] = 0.$$

Following the presentation in (Touzi, 2012) and passing through generalized solution of the Black-Scholes stochastic differential equation we suppose that the solution of the Hamilton-Jacobi-Bellman partial differential equation has the form $w(t, x)=U(x)h(t)$ where $U(x)$ is the DM's utility function.

We suppose that the total initial DM's wealth is 40000 BGN. But the DM does not wish to invest the whole initial sum in the risky process. We suppose that the DM invest in the risky process at moment $t\in[0,T]$ in accordance with his utility function in figure 5 and 6. Over this sum (BGN) we choose the utility function of the form $U(x)=(U(40000)/40000^{\gamma})(40000+(x-40000))^{\gamma}$ where $x\in[40000, \infty)$ and ($\gamma$=0.3).

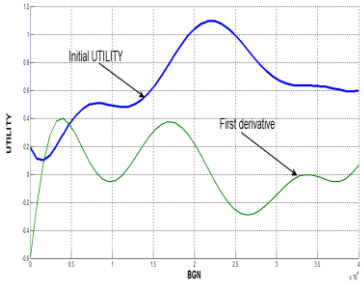



Fig.5 Utility function between 0 and 40000 BGN

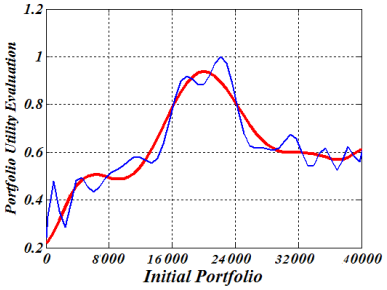



Fig.6 Utility evaluation

The determination of optimal control is easy because we know the first and the second utility function derivative analytically and the extremal point is looked for in the closed interval [0, 1] in a finite set of points. In the following two figures (respectively 7 and 8) we show the solution with optimal control or without control. In figure 7 is seen the effect of optimal control strategy. It is necessary to note that the optimal strategy is in accordance with DM's preferences presented by the utility function.

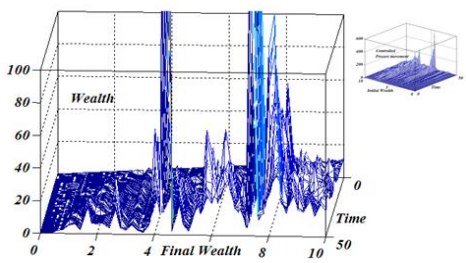



Fig. 7 Liquidation value with control

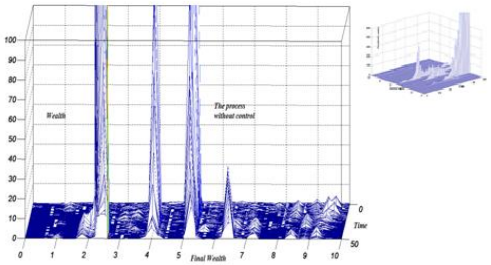



Fig. 8 Liquidation value without control

## 6. Conclusion

Our experiments confirm the applicability of this approach. The proposed method has the following advantages:

- Having understood the term "lottery", DM is relatively fast in learning to operate in the software environment.
- The proposed assessment is a machine-learning approach based on DM's preference.

The utility function is a mathematical abstraction presented here within the framework of the normative approach, i.e. the axiomatic systems of von Neumann and Morgenstern. Proposition 4 reveals the existence of a mathematical expectation, measured in the interval scale utilizing the DM's preferences within the gambling approach. This mathematical expectation, evaluated with the stochastic procedures, could be interpreted as an approximation of the expected utility function.

## References

1. Aizerman, M. A./ E. Braverman/ L. Rozonoer 1970: Potential Function Method in the Theory of Machine Learning, Moscow, Nauka.
2. Cohen, M./ J-Y. ,Jaffray 1988: Certainty Effects versus Probability Distortion: An Experimental Analysis of Decision Making Under Risk, in Journal of Experimental Psychology: Human Perceptions and Performance, vol. 14 (4). pp. 554-560.
3. Farquhar, P. 1984: Utility Assessment Methods, in: Management Science. N30, pp. 1283-1300.
4. Fishburn, P. 1970: Utility Theory for Decision-Making, New York, Wiley.
5. Kahneman, D./ A.Tversky 1979: Prospect theory: An analysis of decision under risk, in: Econometrica, vol. 47, pp. 263-291.
6. Keeney, R. & H. Raiffa 1993: Decisions with multiple objectives: preferences and value tradeoffs, Cambridge Univ. Press.
7. Pavlov Y. and R. Andreev, *Decision control, management, and support in adaptive and complex systems: Quantitative models*, Hershey, PA: IGI Global, 2013.
8. Pavlov, Y. 2005: Subjective preferences, values and decisions. Stochastic Approximation Approach, Proceedings of Bulgarian Academy of Sciences, Vol. 58 (4). 367-372.
9. Raiffa, H. 1968: Decision Analysis. Introductory Lectures on Choices under Uncertainty, New York,Addison-Wesley.
10. Touzi N. & Tourin Ag., (2012), *Optimal Stochastic Control, Stochastic Target Problems, and Backward SDEs*, Fields Institute Monographs, Vol. 29, Springer- Business & Economics.

# Teaching Undergraduate Students to Manage Service Production Processes by Scenarios

**Vanya Slantcheva-Baneva**

MT&M College, 1612 Sofia, Bulgaria
v.slantcheva@mtmcollege.bg

**Abstract.** The paper reviews the outcomes of conducting operation management simulation in a class of Undergraduate students within "Services Marketing Management" course in 2014. A hypothetical call of "service operations" was made delineating services as scenario-driven production processes, drawing attention on the best strategy design and the service process performance itself. A rehearsal of best scenario-driven operation strategy was performed on the grounds of achieving the highest profit result, which contrasted some operations insights on a quantitative and qualitative base of argumentation.

**Keywords.** servuction, scenarios, simulation, operations management, service industry.

## 1.Introduction

Teaching Undergraduates to manage service production processes by scenarios become an interesting educational task in the course of Services Marketing Management. Usually students claim that doing marketing of a service is to manage, more or less, its communication mix *for the sake of* the targets, and in reverse. Moreover they think that doing management of a service means to design marketing strategies for running that service in its best performance. But services are naturally determined by their inseparability and simultaneity of their inner processes of production and consumption. Cosequently, managing a service turns into overseeing service synchronization *(by time and mode)* between the consumer and the service provider, and steering service feasibility (*by encounter, or interaction),* considering its conceptual 7P framework. Than an issue in teaching students to manage services becomes "to understand it", rather than "know about it". Likewise product offer, service offer is to be operationalized one, but contrariwise, service itself is a service-, product-dominated, or hybridized band of components that determine its servuction-and-consumption offer.

The purpose of this report is to review the challenge for the students to create the best service offer by generating servuction-consumption pitches and to make prescriptions of how the challenges to be overcome. Hence, using a simulation tool to manage production-consumption processes of a service is a good reasoning scenario-thinking paradigm to be demanded.

## 2.Defining Services as Scenario-Driven Servuction Processes

The paper leans on the definition of services based on Christian Grönroos's writings [1]. Generally he defines services as processes and activities that are intangible in nature, and more precisely, "service" as a process consisting of series of intangible activities that take place in interactions between the customers and service employees and/or physical goods

and systems of the service provider, which are provided as solutions to the customer problems. That rationale helps him among whole the range of characteristics of services suggested and discussed in the literature, to identify three basic ones: 1) services are processes consisting of activities or a series of activities rather than things; 2) services are produces and consumed simultaneously, and 3) the customer participates in the service production process.

Understanding service management and marketing of services it is critical to realize that the consumption of a service is process consumption rather outcome one, or the customer perceives the service production process as part of the service consumption, but not the outcome of that process, as in traditional marketing of physical goods. Because of that inseparability of the process and the consumption of the service, the production process is delineated as an *open process* and the consumption – as process *consumption*. Then service encounter or customers' interactions with the production process become part of their consumption process, and the consumption of the physical goods included becomes partly process consumption. Hence, the service provider offers a factory-related service element in the solution to the customer problems. Here, considered as a service operations management issue, that interpretation is reposed in the service production process in a restaurant service.

Coined by Pierre Eiglier and Eric Langeard, the term "servuction", combining the terms service and production, describes the service operations system as that part of the service provider's physical environment and goods, which is visible and *open* for processing to customers, contact personnel, and the customer in person. Christopher Levelock expanded that conceptual framework and embraced three overlapping elements: *service operations* (batching, ordering, dining time in a restaurant), where inputs are processed and the elements of the service product are created; *service delivery,* which embraced process consumption and where final "assembly" of those elements take place and the product is delivered to the customer; and *other contact points*, i.e. all points of contact with customers, including advertising, billing, and market research [2].

Defining here services as scenario-driven servuction processes, an assumption of intersecting two disciplines is made. Scenario method and Decision analysis rationales are reffered as management tools for building-up comparable futures that unfold in better service operations. George Wright and George Cairns combine the scenario thinking paradigm and the decision analysis algorithms in a six-stage formal framework to evaluate strategies against decision maker's objectives. Considering that top-down-oriented organizational framework, an articulation of its approach from strategic to operational level management is made here so as to have service production processes strategized, to design alternative strategies and consequently the best performing strategy to be selected [4]. Formulating scenarios on operations level also take place as a context for objective formulation and strategy design, but the very contribution is in the perception of the servuction process as a performance of scenario-driven service operations strategy.

That is why when assuming services as scenario-driven servuction processes we have been taking into account service strategy performance and the service process performance itself – as a simultaneous production and consumption.

## 3.Simulating Service Performance by Scenario-Driven Operation Strategy

"Operation Management Simulation: Benihana©" is a single-player online simulation tool that help students explore the principles of operations and service management while working through a series of challenges set during a single evening at a busy Benihana restaurant.

Customers start in "the bar area" for drinks and then move into "the dining room" where chefs prepare the food right at the table. Each simulation challenge examines a particular aspect of the restaurant operation.

The teaching plan is to help students systematically unearth the elements of Benihana's profitability, determining how each aspect of the operation contributes to superior financial performance. In the process, students learn how to apply important principles of operations management. Students come to realize terms as throughout capacity, demand variability, capacity utilization and service time apply not only manufacturing environment but also to a service production process. The most important message the simulation conveys to the learners is that effective Operations Management practices can have a powerful impact on a firm's profitability, or the profitability could be the KPI for designing and evaluating scenario-driven servuction strategies. Other KPIs to be analyzed are: capacity, demand rates, cycle time, and throughput in a service operation.

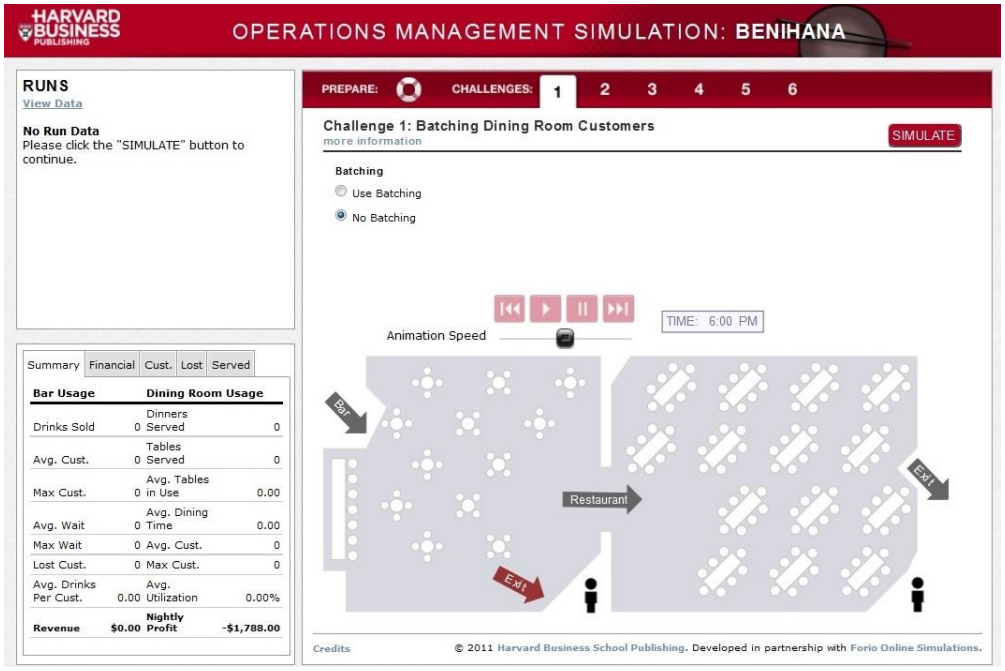



**Fig. 1.** Interface of "Operation Management Simulation: BenihanaV2.

The simulation is a tool that helps students unlock in a quantitative manner the leverages of Benihana's profitability. It is organized as a series of challenges, each with a different set of options for managing the servuction and consumption. In all of these challenges, the goal is to manage the bar and dining area in order to maximize utilization, throughput, and total profit for the evening.

The specific challenges are [3]: 1) batching dining room customers, which strategy affects throughput; 2) optimally designing the size of the bar for maximum profitability; 3) reducing dining time across the evening; 4) boosting demand with advertising and special programs, or redouble service marketing efforts to affect service production and profitability; 5) scheduling different, more complex forms of batching at different times to have an impact on operations and profitability; and 6) designing best strategy as a combination to maximize profitability.

Throughout the simulation students understand how batching strategies improve throughput and how increasing capacity improves bottlenecks. Students digest the importance of optimizing capacity in an operation in order to minimize or eliminate demand variability (cyclical, stochastic, batch size, and service time). They learn to optimize multiple variables in the service production process and ensure consistency in the overall strategy.

## 4.Rehearsing Scenario-Driven Service Production Strategies

Teaching how to develop service operation strategies, "Operation Management Simulation: Benihana©" was conducted in a class of undergraduate students within a course of "Services Marketing Management" during the academic 2013/14 year. The outcomes received confirmed that the simulation helped students rehearse their best scenario-driven operation strategies. It was the final Challenge 6 that required students to consider the lessons learned in the previous challenge-stages – form 1 to 5 – and to design a strategy that maximizes utilization, throughput, and total profit for an evening at the restaurant. When designing their best strategy, they made bulk of combinations of the options given in the sub-stages (see Fig. 2) that ought to bring service production process at Benihana to its efficiency and superior performance.

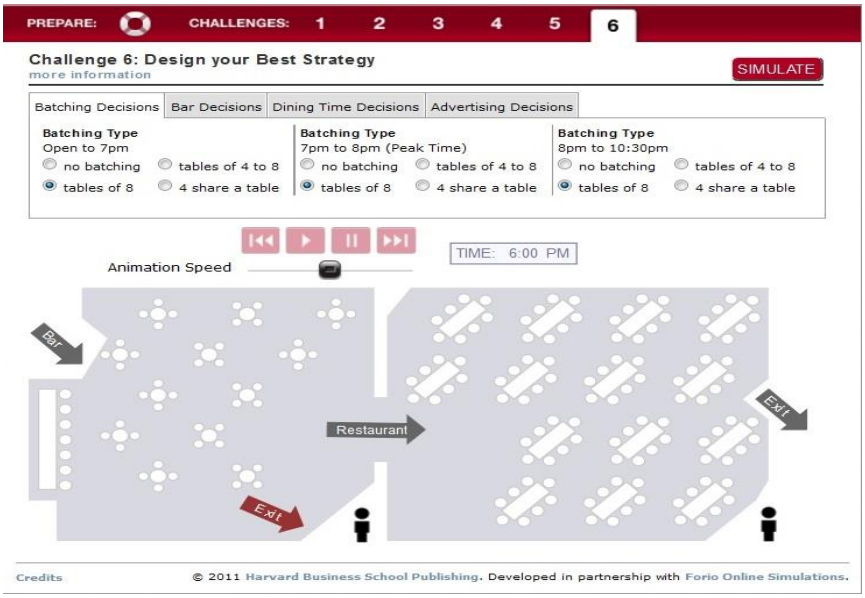



**Fig. 2.** Simulation Interface of Challenge 6: Design your Best Strategy.

Students systematically unearthed the elements of Benihana's profitability, determining how each aspect of the service production contributes to best financial performance. The highest ranked profitability element students used as litmus for best-

performed servuction process is "profit per night". As an evidence of how "profit" navigates any options' blend for simulating "Best strategy design", here is provided the example of student outcomes after 70 runs along the challenges. Her scenario-driven strategy traced a flow of repeated footprints that she accepts as best-practice service process with superior profit (in her case $622.03 per night). She claimed that the best strategy design for the restaurant is based on (see Table 1): 1) non-batching – neither in the dining room, nor in the bar; 2) inclination to increase numbers of the bar seats and decrease restaurant tables; 3) dining time should be kept short up to 45 minutes, especially in pre-peak and the very peaks; according to the student customers could be allowed to relax during the post-peaks; 4) boosting demand with advertising and special marketing programs should be a privilege but with a normal ad-budget, and on expenses of earlier open time of the restaurant (at 5 pm if the peas is between 7 and 8 pm); and finally 5) if any batching is possible it should be for tables of 8 customers.

**Table 1.** "The highest profit" scenario-driven service production process.

| CHALLENGE | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | **2** | | **3** | | | **4** | | | **5** | | |
| **Batching dining room customers** | **Design the bar** | | **Change dining time (minutes)** | | | **Boost demand with advertising and special programs** | | | **Use different types of batching at different times** | | |
| | Batching | Bar Seats/ Restaurant tables | Pre-peak | Peak | Post-peak | Ad Budget | Ad Campaign | Opening Time | Pre-peak | Peak | Post-peal |
| no | no | 15/19 | 45 | 45 | 45 | None | Awareness | 5 pm | no | | |
| yes | yes | 55/14 | 60 | 60 | 60 | 1-2x | Discounts | 6 pm | tables of 4 to8 | | |
| | | 87/10 | 75 | 75 | 75 | 2-3x | Happy hour | 7 pm | tables of 8 | | |
| | | | | | | 3-4x | | | 4 share a table | | |

The simulation performed as a tool that helps students understand key operational insights under the paradigm of scenario thinking and decision analyses. The comparison of three executed strategies envisions an important operational insight (see Table 2).

**Table 2.** Comparison of three challenged highest-profit strategies.

| SCENARIO DETAILS (AVERAGE RUN) | "The highest profit" scenario | "The average profit" scenario | "The lowest profit" scenario |
|---|---|---|---|
| User | Zadgorska, Ina | | |
| Challenge | 6 | 6 | 6 |
| Scenario | 9 | 11 | 13 |
| **KEY RESULTS** | | | |
| Dinners Served | 377.65 | 297.6 | 338.05 |
| Drinks Served | 295.2 | 121.4 | 125.9 |
| Bar Revenue | $442.78 | $242.78 | $251.75 |
| Dinner Revenue | $3,776.50 | $2,976.00 | $2,873.42 |
| Total Revenue | $4,219.28 | $3,218.78 | $3,125.17 |
| Total Costs | $3,597.25 | $2,999.54 | $3,096.00 |
| Lost Customers | 5.75 | 1.25 | 1.2 |
| Profit | $622.03 | $219.23 | $29.17 |
| **DECISIONS** | | | |
| Bar Seats | 79 | 55 | 55 |
| Batching (Pre-Peak) | Tables of 8 | Tables of 8 | Tables of 8 |
| Batching (Peak) | Tables of 8 | Tables of 8 | Tables of 8 |
| Batching (Post-Peak) | Tables of 8 | Tables of 4 to 8 | Tables of 4 to 8 |

| | | | |
|---|---|---|---|
| Dining Time (Pre-Peak) | 45 minutes | 45 minutes | 45 minutes |
| Dining Time (Peak) | 45 minutes | 45 minutes | 45 minutes |
| Dining Time (Post-Peak) | 50 minutes | 45 minutes | 45 minutes |
| Advertising Budget | 1.6x | 1x | 1x |
| Advertising Campaign | Happy Hour | Awareness | Discount Promotions |
| Opening Time | 5:00 pm | 6:00 pm | 5:00 pm |

The student agrees that the decisions made in any of the three scenarios are built on achieving highest-profit outcomes and that it is generated when boosting the volume of consumption by means of dinners and drinks served within an elongated working day, rather than in keeping low the number of the customer lost. The student turned that observation into a hypothesis that undermines customer-centered boosting demand.

## 5.Conclusion

The paper reasonably shows that designing best operation strategies for service providers originates in understanding services as scenario-driven servuction processes. Teaching students to design service production processed appears to be an ambiguous task because, from one side, service operation management should have powerful impact on provider's profitability, and, from the other, servuction itself does not exist separately from service consumption where we talk about productivity drivers, i.e. about customer perceived value and service quality. Building up service operation scenarios should not be reduced to quantitative combinations of options on site, but they should be considered as a context for scenario thinking and decision analysis for long term and strategic perspective.

## References

1. Grönroos, C., *Service Management and Marketing: Customer Management in Service Competition.* 2nd ed., John Wiley & Sons, 2000.
2. Lovelock, C. and J. Wirtz, *Services Marketing: People, Technology Strategy.* 6th Ed., PEARSON Prentice Hall, 2007.
3. W. Earl Sasser , Jr.and R. Ernst *Operations Management Simulation: Benihana,* V2, Harvard Business Publishing, 7004, Rev. July, 15 2014.
4. Wright, G. and G. Cairns, *Scenario Thinking: Practical Approaches to the Future.* Palgrave Macmillan, 2011.

# Problems with the Reliability of Information on the Internet Media Environment

**Anelia Popova**

Department of "Media and Public Communications", UNWE, 1700 Sofia, Bulgaria
aneliapopova@abv.bg

**Abstract.** The lack of a guarantee of reliability of information on the Internet media environment has emerged as a serious problem. With the increasingly rapid pace of development of information technology is increasing and the need to control the flow of information on the web. This report analyzes the classical methods for verifying the accuracy of the information disseminated on the web, ways in which online information are manipulated by the media, as well as possible alternatives to eliminate this problem.

**Keywords:** Reliability of Information, Internet, Media, Manipulation, Virtual Space

## 1. Introduction

In the age of Internet information is a paramount. The lack of credibility of information on the Internet media environment, however, has emerged as a serious problem. With the increasingly rapid pace of development of information technology is increasing and the need to control the flow of information on the web.

In this report we will try to trace the ways in which information in the online environment are altered in meaning, hence manipulated by the media, as well as possible alternatives to eliminate this problem.

## 2. Focus problem

One of the main characteristics of the online environment is that it is open for free supplement. In other words, anyone can submit a supplement to the total amount of information without going through the necessary controls to verify the authenticity of the information. This in turn leads, in many cases, to disinformation, as the veracity of the information disseminated within the network shall be subjected to doubts. When information is altered in effect turn into manipulation of the media. A systematic attempts to manipulate with incorrect information, outlines the increasingly serious trend not only in the direction called. virtual community, but also to his rest, which in one case or another, referring to information circulated on the Internet media environment. Daily, and probably every minute network is increasingly becoming the main channel and a way to communicate with the world - when we are face to face with him. And since the online resource center is a matter of personal choice - no matter whether we talk about social networks, web services, blogs, forums and discussion groups, the need for systematic control is imposed by itself.

The information published on the Internet is not always reliable. In most cases, and unedited, and wide open to all. Online media often give us false information, unedited and manipulated information in order to attract attention. In many cases even resorted to so-called "synthetic data" for the purpose of modeling reality and its transformation into "enhanced reality" / the term was created in 1990 by Thomas Caudill /. As a result of such

practices, finds a decline in confidence in the flow of information in cyberspace (http://izdatelstvo.shu.bg/userfiles/editor/file/bio2008.pdf).

Should not be overlooked however and the fact that professional creators and distributors of information on the Internet through the media, including social networks and bloggers are journalists. Of course, the information spread by the organizations themselves, through their authorized officials. Survey of 2013 shows that the accuracy of the information **as journalists** is displayed in the following percentage indicators (Alexandrova, E., Popski, D. Stoitsova, T. Journalists and bloggers for PR in Bulgaria. Yearbook Department "Mass Communications", NBU, 2009):

- Executive Director - 67.8 %
- Independent expert or analyst - 59.6 %
- Representative of the senior management of the organization - 59.3 %
- Technical Expert in the company - 56.5 %
- User - 47.0 %
- Representatives of the middle management level - 38.1 %
- An employee of the company - 36.0 %
- PR manager / spokesperson of the organization - 28.7 %

**According to bloggers percentage breakdown is as follows:**

- Independent expert or analizator- 60 %
- User – 55 %
- Technical expert in company - 54.1 %
- Executive director- 41.3 %
- Representatives of NGOs - 35.6 %
- A representative of a consumer organization – 35 %
- A representative of senior management - 32.8 %
- An employee of the company - 32.2 %
- PR manager / spokesperson of the organization - 19.7 %

**According to the accuracy of the information consumer is:**

- User - 61.6 %
- Independent expert or analyst of the industry - 59.5 %
- Technical Expert in the company - 47.6 %
- A representative of a consumer organization - 47.1 %
- An employee of the company - 44.9 %
- Executive Director - 41.2 %
- Representatives of senior management - 38.3 %
- Representatives of NGOs - 31.7 %
- PR Manager / speaker - 24.1 %

## 3. Ways of manipulating information on the Internet

Internet there are two competing objectives - information and communication, as examples of online information with untrustworthy character, as we have said, are countless. For example, the majority of bloggers do not check exported in the public domain facts, even if they are not inclined to fool the audience. Passivity and lack of principle in them to individual elements of the process writing and transmission of written, actually defining the problem. And one against another in its meaning stand allegations that individually are serious enough argument for it to raise the level of public control over the information circulated. On the one hand, the public need for traffic news, regardless of their nature, and

the other - the quality of individual information reaching the audience in the information chain. Thus gradually but surely build a model of a new world in which we become more dependent on free information submitted - where missing links are checking, assurance. The threat similar picture to become incontestable reality outside cyberspace comes from practice that a different and subjective point of view, given the web in the form of consumer reaction, is capable of rejecting even deny a number of complex research , rational studies even adopted hitherto truth on certain issues. Worrying that begin to conform more and more entrepreneurs in the online industry (Singer. Larry. magazine. Knowledge, Issue 2 - February 2010). Significantly, and Larry Singer (Larry Singer is an American philosopher and co-founder of Wikipedia in 2001, recently created a new free encyclopedia, Citizendium.) said the following: "Do not believe what you soak in the Internet, because I know how to do digital dish. Internet companies build and participate in Internet communities since the beginning of 1994 and it taught me that most societies are based, not to ensure perfect accuracy of the information and to create a field for fair skirmish in a complex games."

Manipulation of the information published on the internet, mainly happens through the layout. Shaping information in turn can be viewed in two ways:

- Manipulation of viewpoint images;
- Manipulation in terms of the text

The images are often processed with various programs such as Photoshop, while the text is happening so-called layout, resulting in the size of letters, words figurative. Known practice of processing audience is when the purpose of its attraction in capital letters to write something that is outrageous sense or seemingly impossible sounds and letters in italics below explains the real meaning of what was said. Another type of manipulation is semantic targeting. This is a technique that aims to show consumers a selective basis, only what is interesting to them, but actually does not show all the results of the requested information. There are many other ways to manipulate the information on the Internet. "Violence" is the information and behavioral targeting - prevalent mostly among online publishers to reach potential customers. Also, so-called "spamdeksing" that indexes the information stored for a faster search without reviewing individual documents. Often resorted to manipulation of e-mails, fraudulent, often for money. At the risk of dissemination of false information could be considered and rapid exchange of information on the Web - such as posting links, exchange links, copying and citation, which multiplies short time and most treacherous and unpleasing information. A number of examples in this direction. For example, in 2008, when a bank not only suffered from the inability to curb snowballing speculation created online.

## 4. Classical methods for verifying the accuracy of information

Classical methods for verifying the accuracy of information are compelling. We are expressed in effective control of the accuracy of the information resources based on analytical work, including:

- Verification of the facts of the actual intentions with similar material from other sources;
- Clarifying the status of the document;
- Obtaining information on the competence of the author of the material;

And if journalism rule authentication is considered verification of information from at least two independent sources, the web is tens to form credibility. A well-known thesis is that a lie repeated a hundred times, assumed the status of truth.

The modern consumer information is oversaturated, lazy and forgetful. Almost no interest in firm-mediator in contact with the requested content. He is looking for content that is easy, economical, convenient to use and attention grabbing. Its main requirement for

information to objects is to spare no effort and possibly be a click away. It was found that the user hardly interested in behalf of the firm-mediator, while looking for specific content, since it owns commercial awareness, pragmatism and high dose of selfishness. This requires the content to be mainly oriented towards personalized messages and information packages to be not only relevant, but also flexible and most importantly - free - as to a web information can be accessed through various channels (http://www.oreillynet.com/lpt/a/5666, 21.10.2013 г. Lessig, Lawrence. Remixing Culture. An Interview with Lawrence Lessig by Richard Koman O'Reilly Media, Inc., Published 02/24/2005).

Today, the main indicator of quality is speed. This is expected already in all daily activities associated not only with information but also with communication. The modern consumer prefers everything that happens around him, minimized and maximized efforts from such benefits, the maximum proportion to the pace of life. But when the pace gets too fast, this entails considerable risks. On the one hand, sharply reduces people's ability to think in depth, and the other - the level of reliability of dynamic supply facts drastically decline. The implication in such a chain of actions that transience replaces durability and duration in the hierarchy of values. Information revolution requires changes that are not always positive character for overall development level of the audience. A global society is increasingly controlled Internet, where the possibilities seem limitless. All these possibilities, however, does not always encourage people to activity. There are cases where the recipient of the information only register it as such, without this leading to counter participation on his part. And exactly this type users can most easily be fooled by misleading or false. In these cases, manipulation of the facts is particularly dangerous because it can lead to distorted perceptions of reality, provoking in turn a number of problematic situations. Too much information it can be just as dangerous as too meager. And every opportunity to complete and utter position in the web, we often offer contradictory, conflicting or completely false information.

Internet freedom leads to uncontrollable in terms of content information. Already difficult to talk about the Internet as a reliable and credible source of information, because much of the information there are apocryphal nature. Even seemingly serious information expressed in terms of personal opinion or interpretation of a set of facts by the author, it is possible to create an erroneous impression. Even the fact is that today much of the print media publish information generated by users online. And thus incorrect information is supplied once, then quickly disseminate to the public, even sold. Google search engine has been adopted by many as a major source of information that does not always lead to the reliability of the information provided. But despite the controversy of the web environment, the Internet still exists a huge amount of reliable academic and correct information - generated by competent people, scholars, professional analysts and journalists. The problem occurs in parallelism, which is quoted at the same time and distorted facts, outright propaganda, unreliable data, etc. Thereby misleading the audience in a sea of facts and statements online. The lack of balance in the Internet is also part of a general problem with accuracy of information. Furthermore, studies show that the anonymity of the web makes people say things they would not otherwise say even write things that violate intellectual property law (http://e-vestnik.bg, Magazine. "Newsweek", February 2008).

## 5. Possible alternatives to eliminate the problems and preventions for the veracity of the information

Possible alternatives to eliminate the problems are several:

- Obligatory verification of the information provided in the Internet media environment, and outside cyberspace - in books, encyclopedias, scientific articles, etc.;

- Expanding the horizons towards the concepts of "truth," "fact," "objective," "subjective". This is the specificity of the Internet, which significantly alter the psyche of the people, and their idea of the real environment. In this case, it becomes deeply subjective;

- The ability for good judgment in screening information expressed in minimum double check the facts;

- Encryption of information when it is for the transfer of personal data to protect against unauthorized access;

- Platform capable calculated accuracy of rumors in various media, languages and social networks;

As prevention of the authenticity of the information on the Internet is considered the creation of software in the European project Pheme, which automatically detects whether a rumor is true or not. The project is co-funded under the Seventh Framework Programme of the European Union for research in the field of information and communication technologies. The software will enable a better understanding of the influence of rumors on social media - from where originate, how they spread, what and how, and by whom - in multiple copies in different areas. This in turn would enable companies and public organizations to track the spread and veracity of information, ideas and other online data. Daily we see to what extent information and news affecting all levels of society - from shares of companies to government debt, political ratings or even ordinary consumer perspectives on important socially significant issues. This mostly applies to the information age in which rumors "swarm" like viruses in the global network (http://pcworld.bg/22091_evropejski_proekt_shte_izsledva_sluhovete_v_internet, Pheme e new research project. The project is co-funded under the Seventh Framework Programme of the European Union for research in the field of information and communication technologies.).

The analysis of the authenticity of the information spread in social networks, for example, poses serious challenges arising from its laconic, hard to define and highly contextualized nature. Topics that spread like viruses from users of social media are also known as "memes." This may be different ideas, rumors, patterns, content types and more. The emphasis in this software appearing is a necessary prevention for the veracity of the information falls primarily on defining a model of reliability, supported by a platform for analyzing large data sets that identify, track and check the "Family" or Internet "memes "which is added value" true "or" false "while being disseminated through various media and social networks.

Using information extracted from huge banks with existing public data, in combination with content gathered from social networks and media, this software will also develop technology that will automatically assess the reliability of manufactured consumer content.

## 6. Conclusion

At this stage, we are in between WEB 2.0 and WEB 3.0, WEB 4.0 is the only idea on to form the so-called web of symbiosis or arranged no information to be accessed and edited

from both people and automated applications. The idea is associated even with the concepts of artificial intelligence, WEB OS and aggregate intelligence.

The problems with the reliability of information in the Internet media environment is yet to be deployed, thanks to the rapid pace of development of information technologies. And yet dynamic environment, it appears that according to studies, the Internet appears to be only the third most influential and reliable channel for receiving information according to Bulgaria / study of Synovate/. Objectives ¾ of people trust the words spoken by a friend more than writing web. 29 % percent of Bulgarians do, accept for reliable information on corporate websites sponsored links in search engines, e-mail and banner ads. For information - Internet advertising outperforms the percentage TV / study "Nielsen"/ (Global information and media company "Nielsen"., http://prnew.info/tag/medii/page/309/ и http://www.newmedia21.eu/content/2011/11/digital_politics.pdf).

Conclusions of everything written here is that the confidence of consumers, bloggers and journalists in general to information on the Internet is low. There was a need search engines to rank results by rating, for users to have an idea of the accuracy of the information they use. Also available is the need to control information resources that outweighs refused to verify the facts. And last but not least, the fact is the need to prevent the information disseminated on the Internet.

## References:

1. Marinov, Russia. Modern communications and strategic communications. NBU, the Almanac of the department "Mass Communications", 2001, p. 59-60
2. Dermendzhieva, Greta. Online journalism. The media in the digital world. S .: University Press "St. Kliment Ohridski ", 2012, p. 452-453
3. Pavlov Plamen. Onlain Journalism. S .: IC "Vanya Nedkov", 2010, p. 79-80
4. http://kultura.bg/web/ work-at-any-network-is-creator, 21.10.2013, Terziev Krassimir. Creativity Network: everyone is an artist. Portal Kultura.bg published on 08.23.2013
5. http://rhetoric.bg/медии-и-нови-медии, 25.09.2013 on Kassabova Primrose. Media and New Media - a change of rhetorical paradigm. Defining and redefining the concepts. Electronic scientific journal "Rhetoric and Communication" / number 5 September 2012 /
6. http://media-journal.info/. Scientific online magazine for media, PR, journalism, business communication and advertising
7. http://tuj.asenevtsi.com/InfResurs/InfResurs29.htm
8. http://www.vesti.bg/bulgaria/obshtestvo/internet-treti-po-dostovernost-za-bylgarite-2539351
9. http://pcworld.bg/22091_evropejski_proekt_shte_izsledva_sluhovete_v_internet
10. http://computerworld.bg/45464_ontotekst_shte_proveryava_dostovernostta_na_internet_sluh ove
11. http://www.capital.bg/politika_i_ikonomika/bulgaria/2007/03/23/321761_tehnologiite_sresht u_pravoto /, 21.10.2013, the Ognyanova Nelly. Technology against the right - they are eligible for direct exchange technology files and filtering of Internet traffic legally. Online version of the newspaper "Capital", published on 23 Mar 2007

# Comparative Study of Network-Centric Approach for Management of Complex Systems

**Boyan Lazarov[1] and Georgi Kirov[1]**

[1]ISER - BAS, Acad. G. Bonchev Str., Bl. 2, 1113 Sofia, Bulgaria
lazarov.b@gmail.com

**Abstract.** This paper discusses the application of the Network-centric Operations (NCO) approach in the context of crisis management and disaster recovery. The study identifies some shortcomings of the traditional approach to crisis resolution and identifies areas in which NCO could bring advantage and enable faster and more accurate response to potential threats. The authors also propose a technical implementation based on the Data Distribution Service (DDS) standard, which allows for fast development of network-centric applications which can provide the required real-time information exchange needed for the implementation of the NCO approach in the crisis management framework.

**Keywords.** network-centric approach, crisis management, DDS.

## 1.Introduction

In the recent years we have witnessed a number of critical incidents and crises both on local as well as on international level. Due to the ever-increasing reliance of society on technology and infrastructure, a corresponding increase of the impact of these critical incidents can be observed. A crisis can quickly cross functional, temporal and even political borders and thus have an effect over multiple domains. In order to mitigate those increasing effects, an increasing number of resources are required, as well as the participation of multiple organizations which must operate in close coordination in order to minimize both human and economic losses. Key to this cooperation is the effective, real-time information exchange between the participating organizations and the ability to quickly make adequate decisions and organize a coordinated response [1]. The focus of this paper will be to analyse how the network-centric approach can be applied in order to provide the mechanisms for such information exchange and what benefits it will bring to the crisis management process.

## 2.Conventional Approach to Crisis Management

The response to a typical crisis usually involves multiple different organizations, each with its own capabilities and responsibilities during the aftermath of an incident. These organizations are further characterized by their specific internal structure, communication capabilities and protocols. Examples for such organizations are emergency medical services (EMS), police, fire brigade, search and rescue, civil defense etc. At the scene of a large incident, these organizations are usually represented by one or more teams of their respective first responders, who are the ones who actively interact with the environment and are the prime source of information regarding the critical situation. Teams belonging to the same organization usually report to a coordinator, who works as the connecting link between the field assets and the operational level, where the decision making takes place (Fig 1). Information is passed hierarchically from the first responders, through the coordinator, to the

operational level, typically by means of status reports. In the opposite direction, once the situation has been evaluated at the operational level, orders are forwarded down to the field assets for execution.

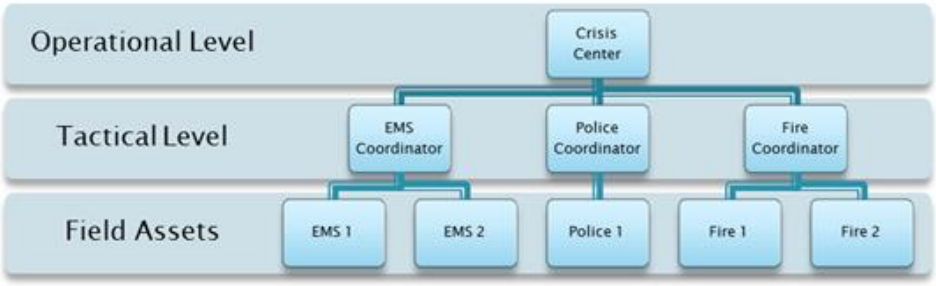



Fig. 1. Hierarchical operational structure during a response to a crisis

One characteristic of this approach is that the higher the level in the crisis management hierarchy the more non-real time the information becomes. This is due to the fact the information undergoes some form of processing (summarization, preparation of status reports) before being forwarded to a higher level. During a critical event, the situation could change dramatically in a very short time frame, therefore creating the risk that until a decision is taken and respective reaction is generated, the response is no longer adequate to the progression of the crisis. In addition, if a new organization has to join the crisis management effort, it needs to be brought "up to speed" with the situation. This requires additional time and there is often the problem that not all parties involved have access to all the required information. All these factors add up to additional time required to include another organization effectively in the crisis management mechanism.

At the Field Assets level, information exchange is mostly limited to sharing between different teams belonging to the same organization. On one hand, this is due to the different communication channels specific to each organization – for example, EMS and police voice communications usually operate on different frequencies, and in some cases communications equipment used by one organization is incompatible with the one used by the other. On the other hand, even if equipment constraints are resolved (e.g. one organization lending communication equipment such as radio stations to the other), the different communication protocols, habits, as well as specific professional terminology make the efficient exchange of information very challenging. Therefore, most of the information sharing between different participating organizations happens at the tactical level, between the respective team coordinators, which then forward any information of interest to the first responders. Naturally, this limits the amount and type of the exchanged information as well as introduces delays and the possibility of introducing errors.

In summary, the hierarchical organization of the information flow during a critical event does not facilitate data exchange between first respondents from different organizations, and also introduces delays in forwarding information up the command chain to the operational level. At the Operational Level (and, to a smaller extent, at the Tactical Level as well) this results in limited situational awareness, creating the risk that decisions become outdated even before they reach the execution level. In order to overcome this, improvements in information collection and sharing are required. The following sections of this paper discuss how Network-centric Operations can be applied to crisis management in order to create efficient information exchange mechanisms.

## 3.Network-Centric Approach and Operations

The term network-centric (NC) is introduced by the U.S. Department of Defense (DoD). In military connotation the NC is frequently associated with a term Net-centric Operations (NCO). The Concept of the NCO is to establish the architectural model for integrating all DoD information systems into Global Information Grid (GIG). The purpose of the NCO is to achieve shared awareness, increased speed of command, higher tempo of operations, greater, increased survivability, and a degree of self-synchronization and standardization [2].

The basic concept is based on the idea of representing management interactions with the help of the network model. In the context of crisis management the network points can be illustrated with the various ministries, agencies and organizations, including specialized action units, analysis centres, etc. The nodes' basic functions are to elaborate assessment of the crisis and to propose particular measures for tackling a concrete situation or accident. To understand what is the new in network centric systems (NCS) and to be able to understand the increased power that is associated with it, one has to concentrate on four levels (Fig. 2) and on their interactions - information level, cognitive level, physical level, and the social level.

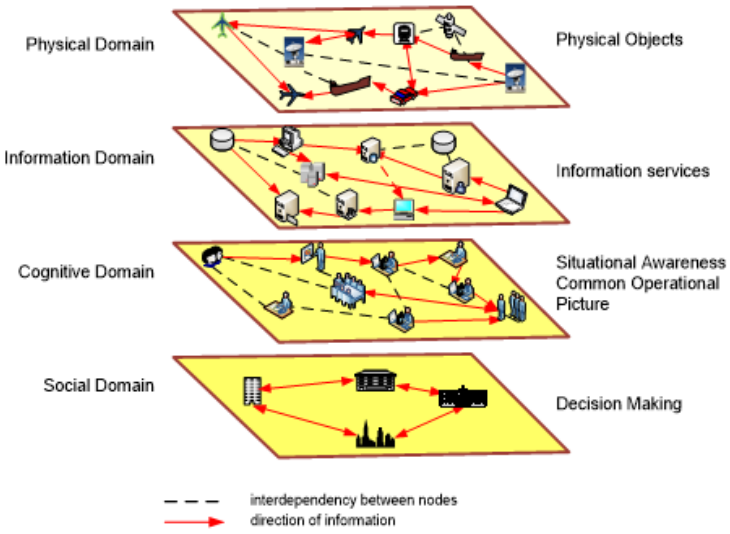



Fig. 2. Conceptual network-centric model.

***Physical domain*** – it is the place where the NCS seek to influence. In the context of crisis management, it includes the physical domain where the critical even has occurred, including all infrastructure and field assets. According to [3], all organizational units must be networked and should support connectivity and interoperability between each other.

***Information domain*** – it is the place where information is created, gathered, transmitted, and shared. It includes information exchange between territorially separated participants. The collaboration aspect aims at taking into account the opinions of all participants in the crisis management process before the most appropriate decision is made.

***Cognitive domain*** - it regards the awareness and understanding of the critical situation. Individuals need not only to develop their own situational awareness, but they will need to share this awareness with a wide range of participants. This domain takes into account a level of training and experience.

***Social domain*** – it regards the ability of organizations to conduct collaborative decision-making and formulate adequate responses to the unfolding situation. This includes considering the impact which decision-making may have on tasks performed by other organizations.

Different organizations during the management of a crisis usually pursue their individual goals – EMS must treat wounded, police is in charge of maintaining the peace and order at the scene, conduct search and rescue etc. In the traditional approach, each of these organizations would follow their own action plan. In a complex environment, however, this approach is not optimal, because it does not consider the effects that one team's actions might have on the other team. For example, police deciding to block access on certain roads might influence or disrupt transportation of victims by the EMS, if the decision is not coordinated or communicated in timely fashion between the teams. Therefore, it is better for organizations to agree on a common action plan rather than follow individual agendas. The Social Domain of NCO envisions the ability to conduct cooperative decision-making, which in turn requires that the goals on the lower levels of the model have been reached. Ultimately, it comes down to the ability to share information and share agreement on its meaning.

## 4.Enhanced Crisis Management Through NCO

This section focuses on a short example of how the network-centric approach can be applied in practice and what quantitative and qualitative improvements it can bring to the crisis management effort. A similar study has been documented in [4], and some of the observations are discussed here as well.

In the example, we assume a combined search and rescue attempt by teams from different organizations. We compare the two scenarios – a traditional one, where teams from different organizations rely only on voice communications with their own controller at the tactical level, and a scenario where the voice communication is supplemented by a network-centric application. The application provides a GIS interface with real-time position of all participating teams, as well as the ability for each user to enter points of interests (POIs) with short descriptions - for example location and state of a victim. Data entered by one user is automatically synchronized to all participants.

We assume a situation where a first response team from one organization wants to share information with the rest of the rescues – for example, a police search and rescue team has come across an injured person. Fig. 3 depicts the information flow in a scenario where only voice communications are available.

In order for the information to reach the EMS personnel on field, it has to be transmitted by the police team to their coordinator at the tactical level. The information is then forwarded to the EMS controller, who informs the EMS responders. For simplification, we assume that the time needed to forward data about one victim from over one link is $t=\delta$ and remains constant, regardless of the specific participants in the link. Therefore, relying on voice only, it would take $t=3*\delta$ before the EMS team is informed about the victim. On Fig.4, the same situation can be observed, this time taking advantage of the available NCO application.

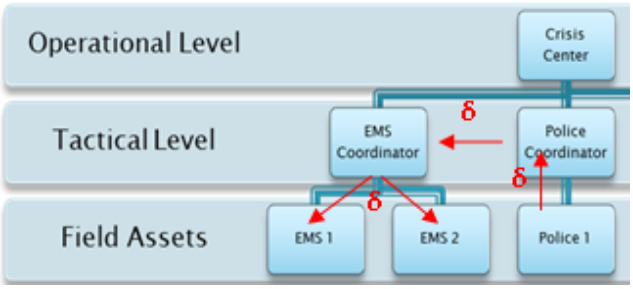



Fig. 3. Information flow with voice-only communication.

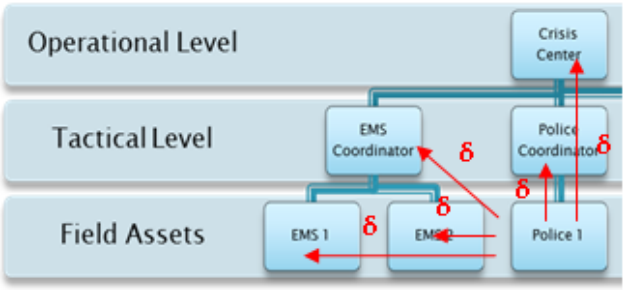



Fig. 4 Information flow with NCO applications.

Upon discovery, the police team enters the data as POI in the NCO application. The data is transmitted in real-time simultaneously to all participants using the application. For simplification, it is assumed that the amount of time required to interact with the application and enter the data is the same as the time needed to transmit all necessary details over a voice channel. The time required for the actual transmission can be neglected. The result is reduction of the reaction time compared to the traditional approach by 2/3rds. Furthermore, simultaneous to informing the EMS team, the information is also transmitted up the functional hierarchy to the Operational Level. In the traditional scenario, Operational Level is usually updated about the situation by means of status reports, resulting in a static, non-real-time snapshot of the event. The proposed NCO enhancement allows real-time updates to be provided to the decision-making authority, thus improving the quality and accuracy of the response actions.

In addition to the quantitative improvement, the application of NCO to the crisis management procedure also improves the data exchange qualitatively. Going back to the discussed example, in the voice-only scenario, the police team will transmit the information about the victim as a description of where the victim is found, in the best case perhaps with exact coordinates. Upon receiving the information, the EMS coordinator at the tactical level will have to process the update, locate the victim on a map of the area, select the closest team and direct them to the location or forward the coordinates and let them find the best route. Besides introducing further delays, using this approach it is much harder to create a comprehensive overview of the situation. For the first responders, who might be visiting the location for the first time, it can be hard to orientate themselves based only on a description of the location or even based on coordinates. Use of the proposed NCO application and having a readily available GIS representation of the information can improve both the processing at the Tactical and Operational Levels as well as response of the field personnel.

## 5.Technical Implementation of NCO Based on The DDS Standard

In this section, we propose a technical solution for development of network-centric applications, which can be used to develop software that could assist in implementing NCO for the purposes of crisis management.

Data-Distribution Service (DDS) is emerging specifications created by the Object Management Group (OMG) in response to the need to standardize a data-centric publish-subscribe programming model for distributed systems. The purpose of the specification is to provide a common application-level interface that clearly defines the data -distribution service [5].

DDS defines a set of services for efficiently distributing application data between participants in a distributed application [5, 6]. The specification provides a platform independent model (PIM) that can be mapped into a variety of concrete platforms and programming languages. DDS provides formal definitions for the Quality of Service (QoS) settings that can be used to configure the services.

The overall Data-Centric Publish-Subscribe (DCPS) model consists of the following entities [4, 7]:

Domain – It is a basic DCPS building unit. All DDS entities that belong to the same domain can interact with each other.

***Topic*** – A fundamental means of interaction between publishing and subscribing applications. Each topic has a unique name and a specific data type that it publishes. When publishing data, the publishing process always specifies the topic. Subscriber requests data via topic. The association between publications and subscriptions is accomplished by means of Topic objects.

***Publisher*** - It is responsible for the dissemination of publications to all relevant subscribers in the domain. The Publisher decides what information is to be published at what time.

***DataWriter*** - It is a type specific interface for the Publisher. It allows an application to offer samples (data) for a specific topic to the Publisher, which will then perform the actual transmission of these samples. The DataWriter is responsible for marshalling the data and passing it to publisher for transmission.

***Subscriber*** - It is responsible for collecting information coming from various publications. The Subscriber decides what information is to be retrieved at what time.

***DataReader*** – It is a type specific interface for the Subscriber. It allows an application to access samples of a specific topic from the Subscriber, which actually collects all incoming samples. The DataReader takes the data from subscriber and de-marshals it into the appropriate type for a topic.

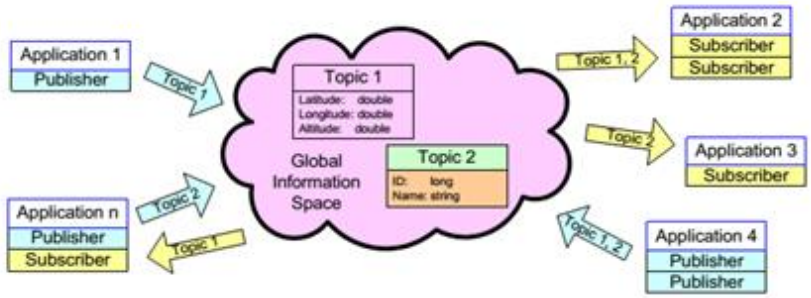



Fig. 5. DDS Publish-subscribe software architecture

## 6.Object-Oriented Layer for DDS Applications

The main goal of OOL is to provide functions and services for working with the DCPS Data Model [6, 7]. Once this is done, we can write the business logic on-top of these abstractions. For the purpose of this investigation a SIMD library is used, which is an open source C++ API intended to simplify the use of DDS. Our aim is to develop an object-oriented architecture based on SIMD that reduces the lines of code that need to be written for a DDS application. Fig. 5 shows the overall architecture. DDS OOL has a publish-subscribe communication architecture that supports object modelling and the notion of disseminating updates to DDS object instances. It allows exchanged data to be described not only in terms of attributes, but also in terms of methods and relationships with other objects. The DDS OOL provides more direct access to the exchanged data through OO data-accessing constructs. In OOL, a Topic corresponds to an object class.

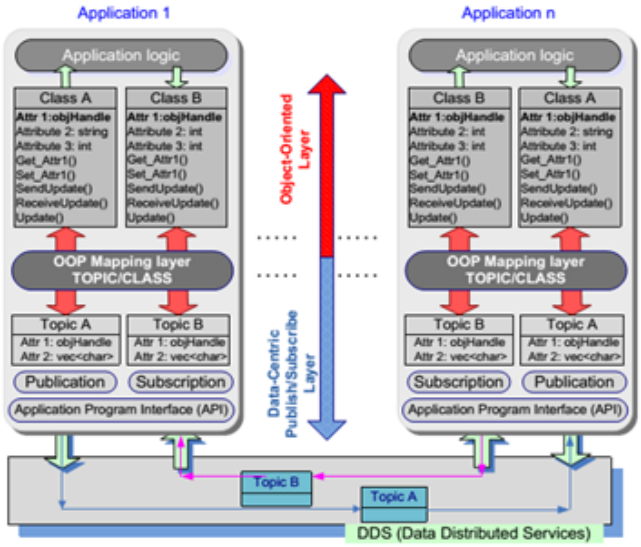



Fig. 6. Object-oriented layer for DDS systems

## 7.Conclusion

Modern-day crises are characterized by their large-scale impact on multiple domains. The traditional hierarchical information exchange during the response to a crisis cannot efficiently cope with the dynamics with which the situation can change. In order for the best counter-actions to be taken, it is necessary to be able to make fast, well-informed decisions. Network-centric operations provide a model for efficient information generation and exchange, which results in better situational awareness and coordinated decision-making. The proposed technical solution based the DDS standard enhanced by an object-oriented layer can be used to implement the network-centric approach in applications that can provide the necessary communication for an efficient network-centric based crisis management.

## References

1. Dimitrov D., System for monitoring of the university computer network performance, Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education – ICAICTSEE-2012, Sofia, 400-405.
2. Benssam, A., Berger, J., Boukhtouta, A, Debbab, M., Ray, S., and Sahi, A. (2007, *What middleware for network centric operations?*, Knowledge-Based Systems, vol. 20, pp. 255–265.
3. Alberts, David S., and Hayes Richard E. (2007). Planning: complex endeavors. DoD Command and Control Research Program, Washington, DC, available at: www.dodccrp.org.
4. Van de Ven et al., Network Centric Operations in Crisis Management, Proceedings of the 5thInternational ISCRAM Conference – Washington, DC, USA, May 2008 F. Fiedrich and B. Van de Walle, eds.
5. Joshi, R., G. Castellote, A Comparison and Mapping of Data Distribution Service and High-Level Architecture, Real-Time Innovations, 2004. Research Program, Washington, DC, available at: www.dodccrp.org.
6. Corsaro, A., L. Querzoni, S. Scipioni, S. Tucci-Piergiovanni, A. Virgillito, Global Data Management, IOS Press, 2006.
7. Wang, N., D. Schmidt, H. Hag, A. Corsaro, Toward an adaptive data distribution service for dynamic large-scale network-centric operation and warfare (NCOW) systems, Proceedings IEEE Military Communications Conference MILCOM, 2008, pp. 1-7.
8. Ryll, M., S. Ratchev, Precision Assembly Technologies and Systems, Four International Precision Assembly Seminar IPAS 2008, Chamonix, France, 2008, 359-369.

# Simulation Approach to Analysing Critical Infrastructure Interdependencies

**Georgi Kirov**

[1]ISER - BAS, Acad. G. Bonchev Str., Bl. 2, 1113 Sofia, Bulgaria
kirov@iser.bas.bg

**Abstract.** The paper is dedicated to the simulation approaches for analysis of the Critical Infrastructure Interdependencies (CII). As a result of an analysis of emergency management domain infrastructure interdependencies and their relevance to computer simulation is shown. The process of development and execution of High Level Architecture (HLA) simulation is explained. It provides a generic platform that allows communication between simulation models based on HLA/RTI standard for information exchange. This approach can address the multiple independent aspects of critical infrastructure interdependencies and simulates the overall effect on the vital infrastructure objects.

**Keywords.** infrastructure interdependencies, HLA/RTI, computer simulation.

## 1.Introduction

Critical infrastructure elements comprise a lot of basic facilities, services, information systems, and communication networks needed for the functioning of a community or society [2]. The people in developed countries are becoming more and more dependent from these resources and assets due to the critical operations and infrastructures they support. However, the infrastructure objects are vulnerable to a variety of disruptions, ranging from mild (short-term power outage, disk drive failure) to severe (natural disaster, equipment destruction, fire). More over, the infrastructure elements have complex relationships and interdependencies that cross infrastructure boundaries. This fact increases the spectrum of the threads. It raises the question for effective emergency management that considers infrastructure interdependencies and vulnerability so that the appropriate control can be put into place to either prevent incidents from happenings or to limit the effect of an incident.

The above-mentioned considerations show that the Critical infrastructure protection has to be a purpose not only of common strategies but also this problem has to be examined on a higher level concerning the national and international security [1]. There is a growing need to analyze and better understand the interdependencies that cross multiple infrastructure sectors and can induce potentially cascading effects [3].

The modeling and analysis of interdependencies between critical infrastructure elements is a relatively new and very important field of study. A number of simulation models have been developed and more are being developed for studying individual aspect of infrastructure elements [4]. The value of these models decreases because they don't consider all aspects of a disaster. The simulation models addressing different aspects of an emergency situation need to be integrated in common framework to provide the whole picture of a situation to planners, trainers, and responders [5].

The purpose of the paper is to present an analysis of the interdependencies between critical infrastructure elements and to explain the process of development and execution of a simulation based on HLA/RTI (High Level Architecture/Run Time Infrastructure) standard.

This approach will provide a possibility of doing an assessment of the elements of the Critical Infrastructure and their interdependencies affected by an emergency situation.

## 2.Interdependencies and cascading effects

In this chapter, the interdependencies and cascading effects between the infrastructure objects are analyzed. According to one of the most widespread definition, an interdependency is a bidirectional relationship between infrastructure objects through which the state of each infrastructure object is influenced by the state of the other. The interrelationship among infrastructure objects is a precondition for cascading effects. The cascading effects can occur when an infrastructure object disruption spreads beyond itself to cause appreciable impact on other infrastructure objects, which in turn cause more effects on still other infrastructure objects. The consequences of the cascading effects due to an infrastructure failure can range from the mild to the catastrophic.

For example, interdependencies exist between different sectors: national energy, water supply system, transport, telecommunication, and emergency services. A power outage can cause an effect on water supply system. The impact of the disruption may not stop at this level. It may go on to adversely affect other critical infrastructure objects (Fig. 1).

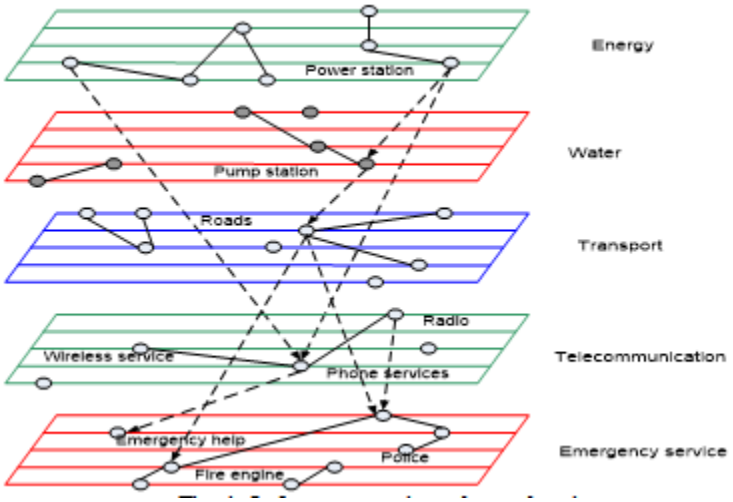



Fig. 1. Infrastructure interdependencies

Given the current state of shrinking budgets and growing number of threats, it is obvious to almost any observer in the field of the emergency management that there is a critical need for a system analysis of the critical infrastructure interdependencies. According to the system analysis there are four classes of interdependencies: physical, cyber, geographic, and logical [8].

*Physical interdependency*: two infrastructure objects are physically interdependent if the state of each depends upon the material output of the other. A requirement for this interdependency is a physical reliance on material flow from one infrastructure to another.

*Cyber interdependency*: two infrastructure objects are cyber interdependent if the state of the one depends on information transmitted through the information system of the other. This is relatively new type of interdependency which basic element is information transfer between infrastructure objects. For example, a lot of infrastructure objects use the

SCADA systems for control and analysis. Therefore, the infrastructure objects have an information dependency from the SCADA systems.

*Geographic interdependency*: two infrastructure objects are geographically interdependent if a local environmental event can change the state in the two objects. The main requirement for this type of interdependency is the existence of a physical proximity. For example, a fire may affect and disrupt all the infrastructure objects located in the area. It is important to notice, that geographic interdependency exists not due to physical connections between infrastructure objects; rather, it arises from the influence the event exerts on all infrastructure objects simultaneously.

*Logical interdependency*: two infrastructures are logically interdependent if the state of each depends upon the state of the other via some mechanism that is not a physical, cyber, or geographic connection. This type of interdependency shows that infrastructure components may affect societal factors such as public opinion and cultural issues. For example, various regulatory mechanisms can give rise to logical linkage among two or more infrastructure objects.

Modeling and simulation of the interdependencies between the critical infrastructure objects is a complex, multidisciplinary process, which implies the work of the different specialists – mathematicians, programmers, physicists, and so on. In the last decade the simulation became a powerful approach for analyzing and design of complex interdependent systems. Simulations allow the engineer to investigate and understand systems that either do not exist or cannot be used for experimentation. According to a simple definition, the simulation is the imitation of a real process or system over the time. The behavior of infrastructure objects can be studied by developing a *simulation model*, which is based on an *interdependency formalization* - a physical, mathematical, or logical representation of network infrastructure processes and systems.

One of the most used ways to present the infrastructure interferences is an interdependency matrix (Table 1). It presents by coefficients ($r_{i,j}$) the influence or impact, that one infrastructure object can have, either directly or indirectly, upon another. The coefficients are defined by experts, which are responsible for an emergency management.

Table 1. Interdependency matrix

| **Interdependency matrix** | | | | |
|---|---|---|---|---|
| **Infrastructure objects** | *Object 1* | *Object 2* | *...* | *Object n* |
| *Object 1* | | $r_{1,2}$ | $r_{1,..}$ | $r_{1,n}$ |
| *Object 2* | $r_{2,1}$ | | $r_{2,..}$ | $r_{2,n}$ |
| | $r_{.,1}$ | $r_{.,2}$ | | $r_{..,n}$ |
| *Object n* | $r_{n,1}$ | $R_{n,2}$ | $r_{n,..}$ | |

When the experts fill in the interdependency matrix they have to take into account the next two definitions [6]:

An infrastructure network *I* is a set of nodes, which are interconnected each other by a relation (connection) presented by function. The relation can be directional or bi-directional. The internal dependencies (connections) in an infrastructure *I* are presented by edges *(a, b)*, with a, $b \in I$.

If *Ii* and *Ij are infrastructure networks,* in which $i \neq j$, $a \in Ii$, and $b \in Ij$, then an interdependency is external one. It is defined as relation between infrastructures and presented by edge *(a,b)*. It means that node *b* is dependent upon node *a*: $(a,b) \rightarrow (b,a)$.

Concerning the results from the multidisciplinary analysis the interdependencies are characterized by *parameters* given in Table 2. These parameters can be used as criteria for a choice of a simulation technology. The parameters show the modeling and analysis of the infrastructure interdependencies is a complex process that makes the use of traditional simulation approaches inefficiency. For example, traditional simulators do not attempt to address the all aspect of an incident response. Therefore, they cannot simulate the infrastructure interdependencies and cascading effects. It results in critical need for a new and innovative simulation approach that allows an information exchange and time synchronization between infrastructure objects. The major difference between traditional simulation systems and proposed one is certainly that contemporary simulation applications not only require the exchange of data during their runtime but also require the exchange of synchronization information regarding their advancement of simulation time.

The interdependency formalization shows the applicability of the use of the computer simulation approach for analysis of the infrastructure interdependencies [5].

Table 2. Interdependency parameters

| **Parameters** | **Parameter analysis – *relevance to computer simulation*** |
|---|---|
| *Interoperability and reusability* | It means that simulation models can be reused in different simulation scenarios and applications. Closely related to reusability is the property of interoperability, which means that the reusable component simulations can be combined with other components without the need for re-coding. Interoperability implies an ability to combine component simulations on distributed computing platforms of different types, often with real-time operation. |
| *Tame scale* | The critical infrastructure has a very important characteristic – dynamics. The dynamics investigation is hard task because it varies from milliseconds to decades. The interdependency analysis requires the using of the simulation technology, which can control the time. |
| *Geographic Scale* | Simulation technology has to support different geographic scale. It allows an investigation of different scenarios in respect to geographic situation of the infrastructure objects. Varying from cities to national level. The ability for scaling increases the quality of interdependency analysis – resolution, information exchange level, and so on. |
| *Cascading effects* | The simulation technology has to simulate the cascading effect, i.e. it has to integrate different simulation models in a common framework to provide the whole picture of a disaster. |

## 3.Simulation execution and development process

The infrastructure interdependency simulator can be presented as a complicated system that consists of great number of simulation models (objects). The processes in the simulation system perform in real time and have random properties. The main problem in the realization of such simulators is associated with an operational environment that is fast

moving, uncertain, and flooded with information. Therefore, just maintaining situational awareness in the simulator is a significant challenge for the architecture specialists. The most critical for the simulation of interdependencies are information fusion and management at different levels, communication, planning and execution monitoring. Add to this, the new requirements for rapid deployments, joint operations of the models, and an information infrastructure that must be fast realized in simulation environments, the current technologies are no very effective.

The proposed new technology allows communication between simulation models based on High Level Architecture/Run Time Infrastructure (HLA/RTI) standard for information exchange. This software technology can address the infrastructure interdependencies between multiple independent aspects of emergency situation and simulates the overall effect on the vital infrastructure objects.

With HLA simulation infrastructure interdependencies can be modeled with almost any level of detail desired and the design space can be explored more finely than is possible with analytical-based approaches or measurements. HLA simulation can combine different simulation models easily, and incorporate measured characteristics of infrastructure objects and their interdependencies.

The HLA Simulation Development and Execution Process (SDEP) is a high level process that describes a generalized lifecycle for building HLA simulation (Figure 2). The main purpose of the SDEP is to guide simulation developers through the series of activities necessary to design and build HLA simulation. For the purpose of emergency management SDEP provides a framework for HLA simulation construction into which lower-level development practices native to each individual model in the area can be easily integrated. Moreover, SDEP defines a systems engineering methodology for HLA simulations, which is a multidisciplinary approach that comprises the whole lifecycle of a complex simulation system – from the early stages of system specification to the implementation and exploitation of the system. SDEP covers five basic steps: definition of simulation requirements, conceptual simulation model development, design and models development, integration of the models and testing, and simulation execution.

At the first stage of the SDEP, a set of needs to be addressed through the development and execution of an HLA simulation is discussed concerning simulation purposes, available resources, and sponsor requirements. This step includes a description of the tasks and activities, operational elements, and information flows required to accomplish or support an infrastructure object operation. It contains descriptions (often graphical) of the operational elements, assigned tasks and activities, and information flows required to support the object functioning. It defines the types of information exchanged, the frequency of exchange, which tasks and activities are supported by the information exchanges, and the nature of information exchanges in detail sufficient to ascertain specific interoperability requirements.

The second step comprises an interdisciplinary process that demonstrates the applicability of interoperable simulation to the emergency management research field by building a working HLA-based simulation. It implies a translation the portion of the real infrastructure objects to be modeled into a suitably abstract and computationally tractable representation [7].

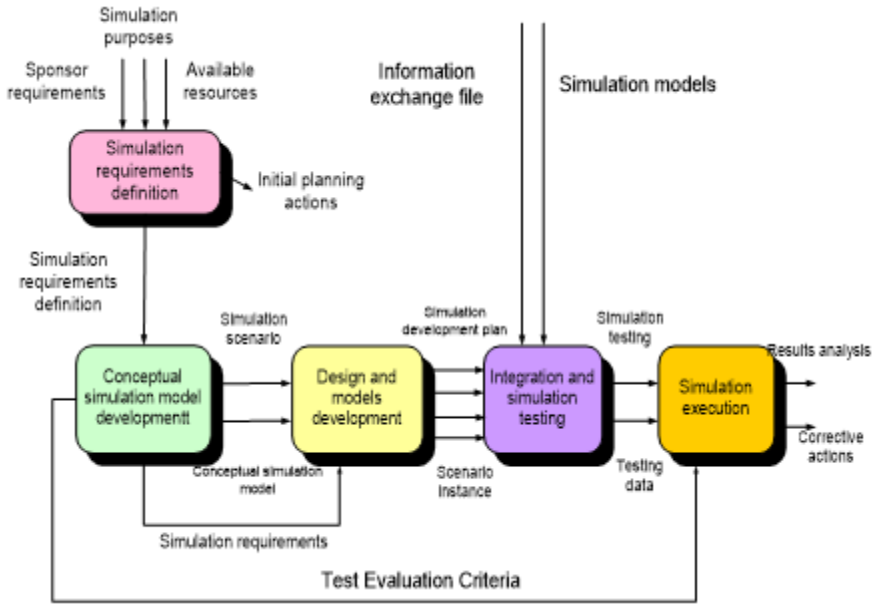



Fig. 2. Steps of the simulation development and execution process

The third step of the SDEP requires the selection of all models that will compose the simulation and the assignment of needed functionalities to each of these models. For example the simulation can be composed of the following elements: power station model that simulate the functioning of power station, water supply model, roads model, plus model simulating emergency services.

At the next step, the software experts integrate the models into a common simulation environment. The impediments to model-to-model interoperability are identified and resolved, and all required modifications and enhancements to the simulation are implemented. The integration of the models allows addressing different aspects of an emergency situation and providing the whole picture of an emergency situation.

At the last step, all required linkages and interconnectivity between infrastructure models are established and testing is performed to ensure that all interoperability requirements in respect to time synchronization are satisfied. The results from the simulation of an emergency situation can be used to analyze infrastructure interdependencies and simulate the overall effect on the vital infrastructure objects.

## 4.Conclusion

The paper is dedicated to the simulation approaches for analysis of the critical infrastructure interdependencies. The analysis of emergency management domain shows the properties of the infrastructure interdependencies and their relevance to computer simulation. The publication presents a HLA development and execution process that defines a basic activities which constitute the end-to-end process of developing and fielding an HLA simulation. This type of simulation provides communication infrastructure for the models' coordination and time synchronization. The benefit of this approach is a possibility of doing an assessment of the elements of the critical infrastructure affected by an emergency situation.

The HLA technology can be used for implementation and support of the Computer Assisted Exercises (CAX) in the field of the civil protection. CAX provide an efficient and cost effective environment to train and analyze the work of the personnel responsible for emergency management.

## References

1. Analysis and Assessment for Critical Infrastructures Protection (ACIP), http://www.iabg.de/acip/index.html.
2. Dimitrov D., System for monitoring of the university computer network performance, Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education – ICAICTSEE-2012, Sofia, 400-405.
3. Jain S. Jain, C. McLean, A Framework for Modelling and Simulation of Emergency Response (paper presented at the Winter Simulation Conference, Dec. 7-10, New Orleans, Louisiana, 2003), 1068-1076.
4. Jain S., C. McLean, Modelling and Simulation of Emergency Response,( paper presented at workshop Relevant Standards and Tools, National Institute of Standards and Technology Internal Report), NISTIR-7071.
5. Kirov G., Integration of simulation models of the elements of the Critical Infrastructure in common simulation environment, (paper presented at "Scientific support of the transformation in security sector conference, September 2006, Center for National Security and Defense Research – Bulgarian Academy of Sciences, ISBN-10: 954-322-175-8, ISBN-13: 978-954-322-175-2), 143-159.
6. Murphy S., A. Feldman, J. Coolahan, R. Lutz, R Saunders, and R. Mukkamala, Distributed Biomedical Simulations: The Application of the HLA for Cardiac and Cardiovascular Simulation Integration, (paper presented at the 2002 Summer Computer Simulation Conference, 2002).
7. Pederson P., D. Dudenhoeffer, S. Hartley, and M. Permann, Critical Infrastructure Interdependency Modeling: A Survey of U.S. and International Research, (Paper prepared for the Technical Support Working Group under Work for Others Agreement 05734, INL/EXT-06-11464, August 2006).
8. Rinaldi S., Modeling and Simulating Critical Infrastructures and Their Interdependencies, (paper presented at the 37th Hawaii International Conference on System Sciences, 2004).

# System Recovery Management Basics

**Nedko Tagarev**

UNWE, 1700 Sofia, Bulgaria,department "National and Regional Security"
ntagarev@gmail.com

**Abstract.**System recovery management(SRM) also known as "disaster recovery" is a process of recovering lost data, result of act of unsuspected problem or problem in the system. The main problem of this paper addresses issues in the information security system. The main burden of the recovery workflow lies on experts to recover the assets of the information. Methodology "try-error" is highly used during the process and the system managers prepare protocols and standards for the elimination of defects in the system. System recovery management is not tied to "backup" the information.

**Key words:** System recovery management, disaster recovery, back-up, archive

## 1.Introduction

"Backup" (backup) is a function in information technology (also popular, as computer technology) that makes copies of the information (data), allowing the recovery of information (data) if they are lost. Backup is the backbone of disaster recovery. Disaster recovery is the backbone of business continuity.

Backing up has two main objectives, which can be connected to the values of information (cyber) security (availability, integrity (reliability) and confidentiality). The first objective is the restoration of data (information) after their loss due to deleting or damage (data corruption). The loss of information can be subsequently of action of each of the users of the information system of neglecting or their common action. In a study done in 2008 found that two thirds of users of information systems based on the information (digital) technologies have information loss.[1] The second goal of restoration of data is to find information from an earlier date. This process is connected with the policy of continuous data (information) backup as defined by the standards of information security. Also, this process is associated with the company's policy length (duration) to keep the information in storages. The most popular understanding of archiving is associated with popular shape – "disaster recovery". The archives should not be considered as a policy for disaster recovery. The disaster recovery is a complex systems which contains archives or backup systems or backup applications we are able to use, such as:

- computer systems;
- complex configurations;
- computing clusters;
- active directory server;
- database servers;
- clouds;
- others.

This is because the fact that recovery is due only to backup data (information). Availability of backup (archive) suggests that at least one of the entries has a value that deserves to be archived. Cost of maintenance of such records can be significant. Fund

governance processes for the storage and organization of the space required for them are complex and need initiatives. Models for information storage building support structures for storage. That exist many devices to store data that can serve as backup. This backup can be organized according to different methodology and can provide transmission address encryption to ensure security of information. Before the data (information) to be transferred to their storage space it is subject to selection, retrieval and processing. There are many techniques to improve the backup process. These processes include:

- Optimization for dealing with open files;
- Dealing with active (live) data sources;
- "Compression";
- "Encryption";
- Optimization of repeats (de-duplication).

Each backup system should have a verification process (Dry run (testing)), which ensures the reliability of the information that is archived. The verification process in computer systems is an activity carried out by a computer programmer where it checks the behavior of the source code, at a certain step, when it is started. In theoretical computer science Dry run (testing) is a series of mental algorithms, sometimes made of pseudo tests where computer analysts study processes is step by step. This process was visually assisted by tables which may be electronic or conventional, which are seen in parallel with the output code.

The role of consumers in those backup systems is very limited. This largely eliminates the possibility of human error or information security threats that come from man.

## 2. Data storage

The first step is associated with the models which acquire to the accuracy of the data (information). Archived information must be stored and organized to a predetermined level. This level is defined in IT policy of the company. Organization can be extremely simple to paper with a list of archived digital information with the location and content of the baseline. Much more complex system may include digitized indexes, catalogs or related databases. Different approaches have different advantages and disadvantages. Part of this system is the rotation of backups (backup rotation scheme). This system is a backup data to a computer or other storage media which, through reuse, minimizes the amount of media used. This process is defined in step where the cost-benefit analyses acquire. Scheme determines how and when each element of the memory is used to archive business information and for how long remains after archived information contain itself. Accumulation of experience can develop various techniques for balance in data storage and recovery, over the time. This is accomplished through additional media storage. Such a scheme can be quite complex, if required by the security policy of the organization - incremental backups, multiple periods of storage, and off-site storage. It should be recognized that "the exclusion of the repository" should be done in constant controlled condition, requiring additional resources. These models or schemes are using limited resources or search economy.

One of the basic models for this type of backup is the First in, First out (FIFO). It is a process of saving space, making changes on the old digital media in the system, as this media containing the oldest and thus the least beneficial previously archived data and put in their place new record. If you perform a daily backup of 20 medium depth of the archive will be up 20 days. In daily backup oldest media is used in the creation of the archive or reserve copy. This is the simplest model for the rotation. Usually the first archive is used in case of problems with a shortage of digital space. This scheme has the advantage in terms of the

longest daily backup of unimportant information or information (data) in the short term. These data should be relevant after the period of rotation. This model suffers from the possibility of data loss. The problem is that there may be an error in the different generations of archived data. Thus, all archived files may contain an error in itself. Therefore, it is better to have a presence of at least one version of archived information, which is not packs error.

Another model for rotational archiving is Grandfather-father-son. In this model there are more backup cycles. These cycles may be divided into days, weeks, months, or other scheme. Basically using the same scheme as in FIFO but rotation is done on a daily basis, weekly basis, monthly basis, etc. This model can be improved by building additional records quarterly, half year or year. These records can be saved separately. Very often, these archives of major cycles are used for disaster recovery.

Method the Tower of Hanoi is a mathematical method which is based on a mathematical puzzle. "Rotation" in this model allows organization to keep a few:

- Current copies of the data;
- Several copies weekly;
- Several copies monthly;
- Copies several years.

Each medium (media system) uses a different number of times. When the added carrier (carrier system) is reused with any rotation older carriers are used in every fourth rotation every eighth rotation, etc. Rotation of carriers can be done daily or weekly. For example, if there are five sets of carriers weekly with sample designations A, B, C, D and E, the system of rotation: ABACABADABACABAE.

Another model which is not systematic, is related to “weight” (importance) of the information (Weighted-Random Distribution), which is based on pre-distribution rule for deleting the information in different generations of the archives, excluding the oldest and latest entries. This is done to company policy with respect the random value (weight) of the information. Importance asset to the each record of information define the probability that it be deleted. Using large variables allows systematic distribution of records between generations. The smaller dimensions result in the allocation of a number of new and less old generations. This techniques disables ensure of even distribution of information between generations in time and space for storage. This method has no advantages over other systemic methods.

One popular method, of storage and/or archiving of information, is unstructured approach. It consists in keeping the information carrier without being structured and described. Description of the information usually consists of the components content and time by answering questions: What is written?, Where and recorded?. This is the most affordable method, but the possibilities for recover of information are quite limited.

Known method is image archiving system (System imaging). Backup of this method represents full storage "image" of the entire system which is made at a predetermined or certain moments of time. This technique is mainly used for recording the best configurations. In principle, this technique is not very good for keeping the current configuration, or copies of the information.

Data Protection is an optional data backup and recovery feature for VMware vCloud Air that enables self-service, policy-based protection of business-critical data by backing up vApps and their associated virtual machines within Dedicated or Virtual Private Cloud service types. Compared to traditional file-based backup and recovery solutions, image-level backups are used in Data Protection to ensure all operating system, file system and application data encapsulated within a virtual machine are captured as a snapshot image before being committed to backup media.[5]

Another method is the point system. This model of storage aims to build storage capability backups at different points of time by organizing data in steps in case of data over time. Another goal is elimination of storing data unchanged (duplication of information), even prior archived data change as their new amendment. Usually full backup copy includes all files, which is done once or at different intervals of time and are used as the basis for the backup spot. Thereafter incremental backups are built at regular intervals. Restoring the system starts from the last record of information that is made before loss of information. Then on the system are "imposed" the older archives. Some backup systems allow the creation of full backups.

Differential backups don't stored data that have changed since the last full backup. This makes it possible to use only two data sets for system recovery. One drawback compared to partial backup method is that over time the last full backup (and thus accumulated changes in the data) increases, so does the time to perform a differential backup. Restoration of the entire system, starting from the latest full backup and then apply only the latest differential backup since the last full backup.

Reverse delta method consists building a "mirror" of the output data, and a comparison between current "mirror" of information and previous states. Reverse delta backup starts with a normal full backup. After a full system backup is performed periodically, which is synchronized with the full backup of the active copy, while the data is stored that are needed for recover of older versions. This can be done by means of a rigid connection, or using a binary differential. This system works well for large, slowly changing sets of information. Examples of programs that use this method are rdiff-backup and Time Machine.

Assessing opportunities to provide backup to the products on the market [2]:

Table.1.[2]

| Name | Linux | Mac OS | Windows | Free Software | Comments |
|---|---|---|---|---|---|
| Acrosync | No | Yes | Yes | No | Alternative client with built-in file monitor to support automatic upload. Website |
| Back In Time | Yes | No | No | Yes | |
| BackupAssist | No | No | Yes | No | Direct mirror or with history, VSS. Proprietary |
| Carbon Copy Cloner | No | Yes | No | No | Proprietary tool for cloning, backing up and synchronising volumes/folders. |
| Cwrsync | No | No | Yes | No | Proprietary. Free Edition available. Based on Cygwin |
| DeltaCopy | No | No | Yes | Yes | Open Source, Free, Based on Cygwin - WebSite – Download |
| Dirvish | Yes | Partial | No | Yes | Backup software for taking incremental snapshots. Free software (Open Software License v2.0). |
| Fpart | Yes | Yes | No | Yes | Split a file tree into sub-trees and launch external command (such as rsync) over generated parts (C, BSD-licensed) |
| gadmin-rsync | Yes | No | No | Yes | Part of Gadmintools |
| Grsync | Yes | Yes | Yes Grsync for Windows | Yes | Graphical Interface for rsync on Linux Systems |
| Handy Backup | Yes | No | Yes | No | Proprietary software. Uses rsync for delta-copying and for differential backup. |

| | | | | | |
|---|---|---|---|---|---|
| LuckyBackup | Yes | Yes | Yes | Yes | |
| QtdSync | Yes | No | Yes | Yes | |
| rdiff-backup | Yes | Yes | Yes | Yes | Incremental backups. *archfs* (nowadays called rdiff-backup-fs which is more accurate) allows the backup to be mounted as a drive, making all versions accessible as snapshots. |
| RipCord Backup | No | Yes | No | | |
| rsnapshot | Yes | Yes | Yes | Yes | Snapshot-generating backup-tool using Rsync and hard links |
| RsyncX | No | Yes | No | Yes | |
| Syncrify | Yes | Yes | Yes | No | Free for personal use, uses rsync protocol over HTTP(S), AES encryption, GUI, 2-way synchronization, written in Java |
| Tym | Yes | No | No | Yes | time machine - Time rsYnc Machine (tym) - bash script - free |
| Unison | Yes | Yes | Yes | Yes | Two-way file synchronizer using Rsync algorithm |
| Space Machine | Yes | Yes | No | Yes | Simplifies internet sync with job files, desktop/email notification, compressed archive, open source bash script |

In continuous data protection not make copies protective and constantly make changes to the host system. This is done by saving files or differences in "blocks" and not in the various levels of files. This method differs from the ordinary mirror, because it allows a return through the log and allows old recovery of data.

## 3. Places to store information

Regardless of the model that uses the information (data) are stored in specific locations, which are called repositories. Repositories are directly related to the availability, integrity (availability) and confidentiality. Although some of the technologies are obsolete and dated but more than 40, all of them are still in use.

Magnetic tape has long been the most commonly used media for recording, storage, backup, archiving and exchange. The tape is usually possible for a better balance between capacity and cost compared with hard drives. In recent years, the correlations between the band-hard drive are much closer. When it comes to archiving, the magnetic stripe is second in use. There are many sizes and types of magnetic tapes, which are private for a specific user or a specific market. They may be a mainframe or a particular brand compatible with certain computer configurations. Specifically in that the strip has a sequential access storage media. Although "finding" the data is slow, the recording and reading of information is relatively fast. Some strip adaptations are faster than some of today's hard drives in terms of recording, retrieving, storing and archiving information. Speed of search, retrieval, recording, archiving are indicators which can assess information security.

HDD ratio power/price improved rapidly and significantly over recent years. This makes it more competitive against the magnetic tape as the recording media in the systematic condition. The main advantages of the hard drive as a means of storage and backup times are lower requirements for access to media, access to information (data records), capacity (volume) and easier to use. External drives can be connected via local interfaces like SCSI, USB, FireWire or ESATA, or remote locations by technologies such as Ethernet, ISCSI, or

Fiber Channel. Based on hard drive backup systems, such as virtual libraries, optimization of duplication of information requirements are significantly reduced capacity of space used daily or weekly backup of the information. The main disadvantages of the records or copies stored on a hard drive are that they are easily damaged - in transportation (backing out of it), cleaning or common activities which contains human intrusion. They are vulnerable to physical attacks or physical threats and risks. On the other hand, the stability of the system is unknown period of several years.

Optical storage media such as CD-minute, DVD-minute and Blu-ray discs are most commonly used in personal computers, because the whole unit costs are low to media. The capacity and speed of these and other optical discs are generally lower than those of the hard disk or magnetic tape. Optical disc formats of memory is a type of ROM (Read only memory), which makes them useful for archival purposes on side of completeness of the information, as the data cannot be changed. Using auto-changer or jukebox can make optical discs possible use for large-scale backup systems. Some optical storage systems make it possible to build a catalog of archived information without existing human contact with the discs. This allows for a longer time to maintain data integrity. Also as in magnetic information carriers has unknown period of saving up to few years.

Solid storage facilities that are known as flash memory, thumb drives, USB flash drives, CompactFlash, SmartMedia, Memory Stick, Secure Digital Cards, etc. These devices are relatively expensive for their small capacity. They are suitable for archiving relatively low amounts of data. Solid storage does not contain any moving parts, allowing an enormous capacity of about 500Mbit / and 6Gbit / sec. SSD discs are already available in the range of 500GB and speed of TBS. There is much more uncertainty regarding the possible periods of storage of information.

There are remote backup is associated with the spread of broadband access. Backup via the internet to a remote location can protect against some archives worst-case scenarios such as fires, floods, earthquakes, which are the most common technological threats and destroyed all the records that are in close proximity to the attack, along with all other material system assets. There are a number of serious shortcomings in reusing services for remote backup. Firstly, the speed of transport of the information that Capacity provides Internet connection is usually slower than the local device data storage. Mass broadband creates problems with system backup upstream is very slowly. Connection is used only occasionally to download a file from a backup. A trend to this type of backup is to be used as limited to relatively small amounts data of high value or critical information. On the other hand, consumers must have confidence in the provider, which is a third party provide services to maintain the confidentiality and integrity of archived (stored) data. Confidentiality can be ensured by encrypting the information before transmission to the firm performing the service backup with encryption key that is known only to the user and stored in other location. Service backup is using one of the methods for archiving. It makes use of that storage more complex way of doing backups.

## 4. Managing data store

Regardless of the model of data archive and the type of media in the repository, the IT management is always looking for a balance between the availability, integrity, confidentiality and costs.[3] These methods and resources are not mutually exclusive. They are often combined, depending on the needs and abilities of the user. Typical is measured using online media to manipulate the data before the information is archived in the main library.

Online storage (online), backup is one of the most accessible methods. It can start recovery or access to records within milliseconds. Examples include external drives or disk arrays. This process is done through Storage area network (SAN). Method is expensive, but pretty fast. This storage types are highly vulnerable to attacks by margin increases vulnerability. They can range from deleting or overwriting the information in consequence of an accident, disaster, intentional malicious act as a power attack or malware.

Near-line is not widely available and is more expensive than online storage. However, this method is used for storing information. Measured on a library whose entries are stored within a few seconds to several minutes. For the replacement, media is using mechanical devices. This media has a high degree of safety.

Off-line require direct human intervention in any of the operations for the storage of information. This is related to the changing information media changing the cables and more operations related to hardware. Information is not available to all computers except in limited periods of time. Data is written and read in reverse order. Thus data is protected classes in lack of short cycles archiving. Backup time is determined by the state and place the information carrier.

Off-site data protection is a method used to obtain protection against disaster or other specific vulnerabilities associated with the data carrier. Many organizations prefer to keep the information in the archives in the "vault" to another location. These locations range from personal computers, system administrators, alternative office, where the requirements can reach the level disaster or temperature controlled room to cyber bunkers with a high degree of security of the backup media.[4] The important thing is that the copy is located off-site storage of basic archive, and this process can be done via the Internet. These copies have very limited capacity and differ essentially from the offline copies.

Backup site or also called Disaster recovery center (DR center), is a method that applies when the available backup data media are not sufficient to restore the information system. This process is necessary and needs specified configuration of computer systems and networks. Many companies have their own systems to recover data and configurations tailored against specific scenarios. Another variant was to use contractors or third parties. Construction of such a center is very expensive, and very rarely is preferred over more traditional backup methods. Classic way of doing it is maximally build a "mirror image", which transfer basic archive center to disaster recovery.

## 5. The state-of-art software as a conclusion

“VMware vCloud® Air™ Data Protection offers secure, policy-based backup and recovery in the cloud for virtual machines hosted exclusively on vCloud Air. Available across both Dedicated Cloud and Virtual Private Cloud infrastructure-as-a-service types, Data Protection includes the following feature highlights:

- Backup policy affinity controls per Virtual Data Center (VDC) or per vApp
- Daily (24-hour) Recovery Point Objective (RPO) guarantee

- Virtual machine (image-level) Restore Granularity Objective (RGO)
- Custom backup window scheduling
- Configurable data retention
- On-demand backups
- Synthetic-full backup images
- Standard AES-256 encryption of data at-rest
- Intelligent consumption tracking and activity reports

Table.2.

| | Dedicated cloud | Virtual private cloud |
|---|---|---|
| Data Protection Capacity Units | 1TB Front End Protected Storage | 1TB Front End Protected Storage |
| Unlimited Backups | Yes | Yes |
| Unlimited Restores | Yes | Yes |
| Backup Operations (Concurrent) | 16 | 16 |
| Restore Operations (Concurrent) | 16 | 16 |
| Production Support | 24 Hours/Day<br>7 Days/Week<br>365 Days/Year | 24 Hours/Day<br>7 Days/Week<br>365 Days/Year |
| Subscription Terms | 1 month<br>12 months<br>24 months<br>36 months | 1 month<br>3 months<br>12 months<br>24 months<br>36 months |

Key benefits

- Business continuity in the cloud – Data Protection ensures business continuity compliance for enterprise workloads hosted in the cloud, by enabling self-service backup and recovery operations as a foundational layer in risk mitigation and data loss prevention.
- Agentless backup architecture – Data Protection leverages both vStorage APIs and VMware vSphere®-based snapshot technology to protect any virtual machine hosted on vCloud Air, regardless of the OS and applications installed and without any dependencies on resource- intensive guest agents.

Simplified manageability – Eliminate the complexities of routine backup policy administration by taking advantage of the intuitive, self-service controls provided by Data Protection. Scheduling backups, configuring data retention and specifying affinity settings per Virtual Data Center or per vApp provide maximum flexibility in how data is protected in the cloud.[5]

## References:

1. http://www.kabooza.com/globalsurvey.html (14.07.2014г.)
2. http://developer.apple.com/library/mac/#documentation/Darwin/Reference/maiManPages/man1/rsync.1.html (14.07.2014г.)
3. COBIT 4.1
4. http://www.cyberbunker.com/web/dell-powervault-md1000.php
5. VMware_vCloud_Hybrid_Service_Data_Protection, pg.1

# Information Security of the Enterprise – Key Factor for Increasing its Competitiveness

**Zahari Goranov**

Department of „National and Regional Security“, UNWE, 1700 Sofia, Bulgaria
zgoranov27@e-dnrs.org

**Abstract.** The paper reviews the connection between the information security of the enterprise and its competitiveness. Increasing Information Security automatically increases the competitiveness of the entire enterprise

**Key words**: information security, information system, security, enterprise, competitiveness

## 1.Introduction

The problem with the protection of information security in the enterprise is now extremely topical. Information Security is protection of information and information systems from unauthorized access, use, disclosure, alteration, reading, recording and destruction. Information security is the protection of information, regardless of its form - electronic, printed or otherwise. Computer security focuses on the integrity of computer systems and networks and their processing of information. Information security management practices are the risks associated with the use, handling, storage and transmission of information and the systems and processes used for those purposes. Enterprises should increase the level of their information security, otherwise they are going to lose their competitiveness.

## 2.Definitions of Information Security and Competitiveness

What does Information Security (IS) mean?

Information security (IS) is designed to protect the confidentiality, integrity and availability of computer system data from those with malicious intentions. Confidentiality, integrity and availability are sometimes referred to as the CIA Triad of information security. This triad has evolved into what is commonly termed the Parkerian hexad, which includes confidentiality, possession (or control), integrity, authenticity, availability and utility.

Technopedia explains Information Security - information security handles risk management. Sensitive information must be kept - it cannot be changed, altered or transferred without permission. For example, a message could be modified during transmission by someone intercepting it before it reaches the intended recipient. Good cryptography tools can help mitigate this security threat. Digital signatures can improve information security by enhancing authenticity processes and prompting individuals to prove their identity before they can gain access to computer data [1].

In science there is no definition of competitiveness. Each author takes competitiveness differently. Theory of competitiveness is of paramount importance for business development. It provides a system of knowledge about how we can achieve lasting success in the fight against competitors and - higher living standards of society. According to some authors more competitive now boils down to increasing the competitiveness of its products. They believe that if a product is in demand and purchased in the market, the entity

that is offering them competitive. Information Security is an important part of the competitiveness of the enterprise (it is one of the most important factor for competitiveness). The term competitiveness is used in a bewildering variety of ways, both in the policy community and in academic research. Some equate competitiveness with the ability to achieve certain overall outcomes, such as a high standard of living and economic growth. Other definitions focus on the ability to achieve specific economic outcomes such as job creation, exports, or FDI(foreign direct investmants). Yet other definitions see competitiveness as defined by specific local conditions such as low wages, stable unit labor costs, a balanced budget, or a 'competitive' exchange rate to support a current account surplus. These different views of competitiveness have confused the public and scholarly dialogue.

The evolution of the competitiveness debate has oscillated around three ideas: market share, costs, and productivity. When the term competitiveness first gained prominence in the1980s, the public debate in the United States was dominated by fears about the seemingly [2]. Another view of competitiveness focuses on measures related to a location's costs. Work on cost competitiveness has various interpretations. Low labor costs (compensation per hour, per employee) are seen as a sign of competitiveness leading to lower unemployment, higher exports and higher FDI. Other studies examine the relationship between (labor) costs and output. Unit labor costs are often used to evaluate whether a country's balance of payments is likely to be sustainable (e.g., European Central Bank, 2008) [3].

There is another theory about competitive use of our resourses and labour. Competitiveness is defined by the productivity with which a nation utilizes its human, capital and natural resources. To understand competitiveness, the starting point must be a nation's underlying sources of prosperity. A country's standard of living is determined by the productivity of its economy, which is measured by the value of goods and services produced per unit of its resources. Productivity depends both on the value of a nation's products and services – measured by the prices they can command in open markets – and by the efficiency with which they can be produced. Productivity is also dependent on the ability of an economy to mobilize its available human resources. True competitiveness, then, is measured by productivity. Productivity allows a nation to support high wages, attractive returns to capital, a strongcurrency – and with them, a high standard of living [4].

There are also more than one definitions for computer and information security. Computer security is a branch of technology known as information security as applied to computers and networks. The objective of computer security includes protection of information and property from theft, corruption, or natural disaster, while allowing the information and property to remain accessible and productive to its intended users [5].

Information Security - this term is more commonly found not only in specialized publications. Almost everyone has experienced one or another his appearance because the information has become almost the most important part of our daily lives. A valuable things must be kept.

In the business world, information security has its own specific and versatile manifestation. In today's dynamic world gets advantage this organization has accurate and timely information that can store, process and communicate reliably. Information management is critical for internal - business processes and competitiveness of each company - this applies both to the conditions of normal business activity and in the case of unforeseen crises. Therefore, information security is becoming a decisive factor for the existence of any organization, regardless of the subject of its activities. In fact, in some areas of the business that has long been a fact [6].

## 3.Information Security as a part of the competitiveness of the enterprise

If we can not protect our information we will loose our business. Here some more definitions of information security:

- State of protection of the information environment.
- Concept of protection of programs and data from accidental or deliberate modification, destruction, disclosure and use without permission.
- Process of ensuring the protection of information technology, providing the work of information systems.
- Lack of unacceptable risk associated with leaks technical channels, unauthorized or unintended impacts on data or other resources of information system [7].

Security objectives:

- To secure the values of the system;
- To protect and ensure the accuracy and integrity of information;
- To minimize the damage resulting from the modification or destruction of information [8].

## 4.How to protect our information

To make sure that we provide high quality protection must meet the standard ISO 27001:2005. If we meet the requirements of the standard it means that much of the information is defended from our coqurents.

The ISO 27001 specifies the requirements for systems to manage information security and is applicable to all organizations, regardless of size, activities and processes. Implementation of ISO 27001 in the organization's strategic decision and aims to achieve a certain level of data protection, ensuring its confidentiality, integrity and availability. Level of protection of ISO 27001 does not cover only the risks associated with IT infrastructure and covers physical security, human resources, legal protection, compliance with legal and regulatory requirements. [9]

In the most enterprises are kept similar rules for IT security.

- Split up the Users and Resources

For an information security system to work, it must know who is allowed to see and do particular things. Someone in accounting, for example, doesn't need to see all the names in a client database, but he might need to see the figures coming out of sales. This means that a system administrator needs to assign access by a person's job type, and may need to further refine those limits according to organizational separations. This will ensure that the chief financial officer will ideally be able to access more data and resources than a junior accountant. That said, rank doesn't mean full access. A company's CEO may need to see more data than other individuals, but he doesn't automatically need full access to the system. This brings us to the next point.

- Assign Minimum Privileges

An individual should be assigned the minimum privileges needed to carry out his or her responsibilities. If a person's responsibilities change, so will the privileges. Assigning minimum privileges reduces the chances that Joe from design will walk out the door with all the marketing data.

- Use Independent Defenses

This is a military principle as much as an IT security one. Using one really good defense, such as authentication protocols, is only good until someone breaches it. When several independent defenses are employed, an attacker must use several different strategies to get through them. Introducing this type of complexity doesn't provide 100 percent protection against attacks, but it does reduce the chances of a successful attack.

- Plan for Failure

Planning for failure will help minimize its actual consequences should it occur. Having backup systems in place beforehand allows the IT department to constantly monitor security measures and react quickly to a breach. If the breach is not serious, the business or organization can keep operating on backup while the problem is addressed. IT security is as much about limiting the damage from breaches as it is about preventing them.

- Record, Record, Record

Ideally, a security system will never be breached, but when a security breach does take place, the event should be recorded. In fact, IT staff often record as much as they can, even when a breach isn't happening. Sometimes the causes of breaches aren't apparent after the fact, so it's important to have data to track backwards. Data from breaches will eventually help to improve the system and prevent future attacks - even if it doesn't initially make sense.

- Run Frequent Tests

Hackers are constantly improving their craft, which means information security must evolve to keep up. IT professionals run tests, conduct risk assessments, reread the disaster recovery plan, check the business continuity plan in case of attack, and then do it all over again. [10]

## 5.Conclusion

In conclusion we can say that information security is vital to the existence of enterprise. In modern and rapidly developing world, our company is obliged to respond to any challenge. In the modern world, technology is developing at a very great pace and we need to keep up with them. Today, effective management of any company, regardless of its size and scope of economic activity is impossible without providing opportunities for processing of all relevant information, and this in the short term for the design and management decisions. Everything is reasonably practicable only in terms of making full use of information systems. No less important for organizations to understand also that the introduction of an information system it is all over, and switching to a higher stage of organizational development and information. The introduction of any information system for management purposes is only one step in a continuous process of improving their general and in particular their information base. [11]

If our systems are working well, we can do our work better and faster. If we can protect our systems from competitors will be able to keep their information advantages. We shoul work every day to increase our IT securirty.

## References

1. http://www.techopedia.com/definition/10282/information-security-is
2. http://www.nber.org/papers/w18249.pdf
3. http://www.nber.org/papers/w18249.pdf
4. http://www.iese.edu/en/ad/AnselmoRubiralta/Apuntes/Competitividad_en.html
5. http://en.wikipedia.org/w/index.php?title=Computer_security&&oldid=30994438
6. http://cio.bg/1923_informacionnata_sigurnost_v_promishleno_predpriyatie_luks_ili_neobhodimost
7. http://tuj.asenevtsi.com/Sec2009/Sec02.htm
8. http://tuj.asenevtsi.com/Sec2009/Sec02.htm
9. http://www.lirex.bg/bg/Konsultantski-uslugi/Informatzionna-sigurnost/Penetration-Test
10. http://www.techopedia.com/2/27825/security/the-basic-principles-of-it-security
11. http://znanieto.net/index.php?option=com_virtuemart&func=downloadRequestFree&page=shop.browse&product_id=6798.

# Business Process Automation with BPMS

**Geno Stefanov**[1]

[1] UNWE, 1700 Sofia, Bulgaria
genostefanov@unwe.bg

**Absract.** Business processes can be automated and turn into executable process using a business process management system (BPMS). Not all of the business processes are suitable for automation with BPMS. This paper aims to analyze the possibilities for process automation with BPMS. The first section of the paper present the literature overview on the topic. Then the BPMS architecture is analyzed and described. Finally, the automation boundaries and execution properties are defined and analyzed.

**Keywords.** Business process, business process management, business process management systems, process automation.

## 1.Introduction

Every organization - be it a governmental body, a non-profit organization, or an enterprise - has to manage a number of processes. Typical examples of processes that can be found in most organizations include - *Order-to-cash, Quote-to-order, Procure-to-pay, issue-to-resolution, Application-to-approval.* A business process can be defined as a collection of inter-related events, activities and decision points that involve a number of actors and objects, and that collectively lead to an outcome that is of value to at least one customer [1]. Figure 1 depicts the ingredients of this definition and their relations.

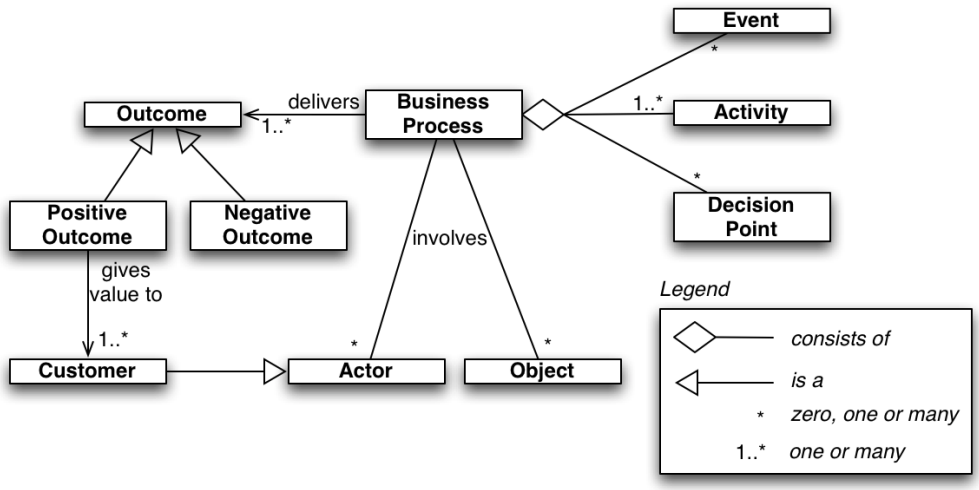



Figure 1. Ingredients of a business process

During the past decades process orientation has been an increasing trend in business improvement. Business Process Management (BPM) is a popular way of increasing process orientation in organizations, often with process automation and some kind of IT-system, Business Process Management System (BPMS). BPM is Body of principles, methods and tools to design, analyze, execute and monitor business processes [1], [2]. BPM as continuous cycle comprising the following phases (fig. 2):

- Process identification. In this phase, a business problem is posed, processes relevant to the problem being addressed are identified, delimited and related to each other. The outcome of process identification is a new or updated process architecture that provides an overall view of the processes in an organization and their relationships.
- Process discovery (also called as-is process modeling). Here, the current state of each of the relevant processes is documented, typically in the form of one or several as-is process models
- Process analysis. In this phase, issues associated to the as-is process are identified, documented and whenever possible quantified using performance measures. The output of this phase is a structured collection of issues.
- Process redesign (also called process improvement). The goal of this phase is to identify changes to the process that would help to address the issues identified in the previous phase and allow the organization to meet its performance objectives.
- Process implementation. In this phase, the changes required to move from the as-is process to the to-be process are prepared and performed. Process implementation covers two aspects: organizational change management and process automation.
- Process monitoring and controlling. Once the redesigned process is running, relevant data are collected and analyzed to determine how well the process is performing with respect to its performance measures and performance objectives.

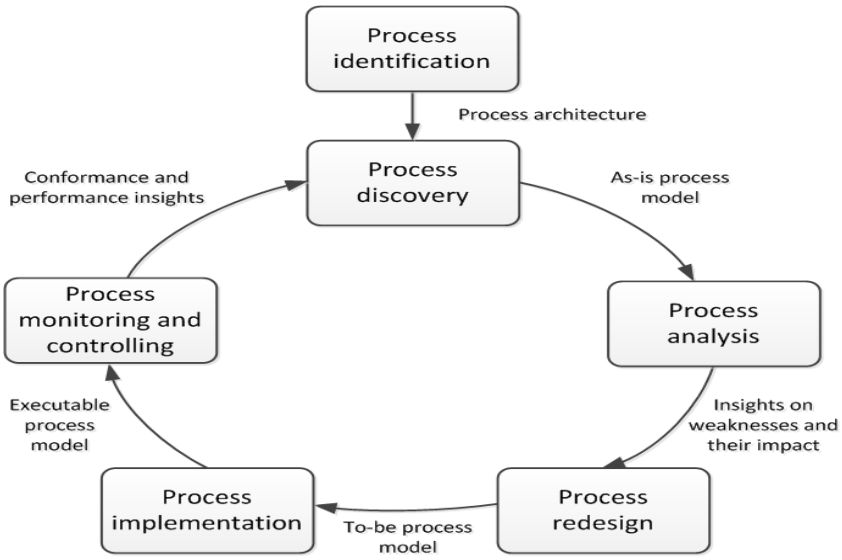



Figure 2. BPM lifecycle

The purpose of a BPMS is to coordinate an automated business process in such a way that all work is done at the right time by the right resource. A BPMS is also a standard type of software system. Vendors offer different BPMSs with a varying set of features, spanning the whole process lifecycle: from simple systems only catering for the design and automation of business processes, to more complex systems also involving process intelligence functionality (e.g. advanced monitoring and process mining), complex event processing, SOA functionality, and integration with third-party applications and social networks. Despite the variety of functionality a BPMS can offer, the core feature of such a software system resides in the automation of business processes. With a BPMS it becomes feasible to support the execution of a specific business process using the standard facilities offered by the system. However, it is essential that a business process is captured in such a way that the BPMS can deal with it, i.e. that the BPMS can support its execution.

Figure 3 shows the main components of a BPMS, namely the execution engine, the process modeling tool, the worklist handler, and the administration and monitoring tools. The execution engine may interact with external services.

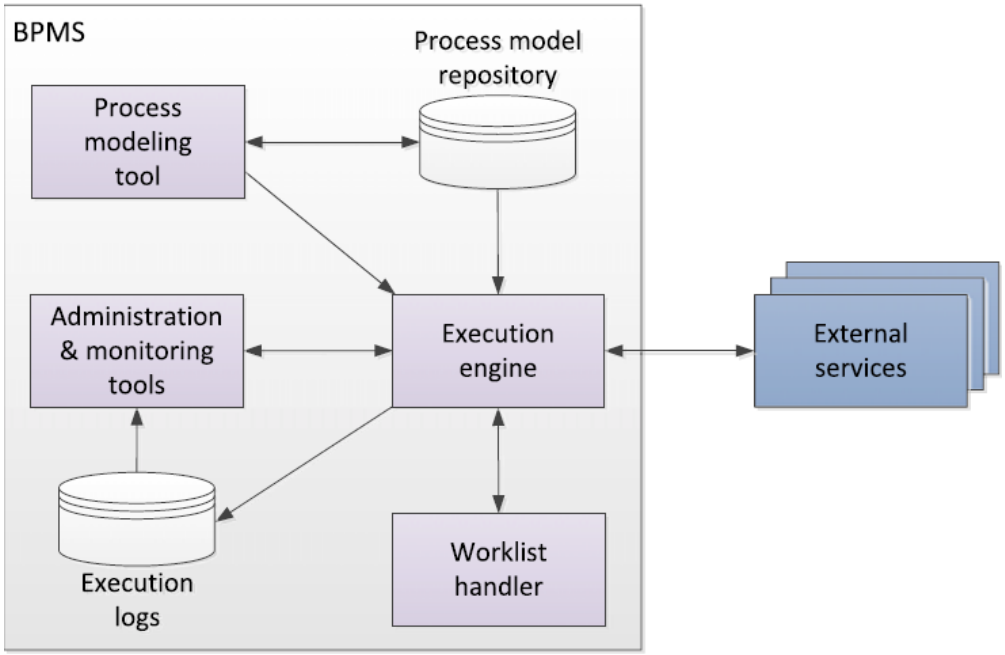



Figure 3. The architecture of BPMS

The goal of this paper is to analyze the possibilities for process automation with BPMS by defining 4 steps namely identify automation boundaries, review manual tasks, complete the process model and specify execution properties.

## 2.Process automation with BPMS

In this section we will show how to automate a business-oriented process model in order to execute it on a BPMS. We propose a five-step method to incrementally transform a business-oriented process model into an executable one, as follows (fig. 4):

1. Identify the automation boundaries
2. Review manual tasks
3. Complete the process model
4. Specify execution properties


Figure 4. Process automation steps

First, we need to identify what parts of a process can be coordinated by the BPMS, and what parts cannot. In a process there are automated, manual and user tasks. Automated tasks are performed by the BPMS itself or by an external service while manual tasks are performed by process participants without the aid of any software. A user task sits in-between an automated and a manual task. It is a task performed by a participant with the assistance of the worklist handler of the BPMS or of an external task list manager.

This difference between automated, manual and user tasks is relevant: automated and user tasks can easily be coordinated by a BPMS, while manual tasks cannot. Thus, in this first step we need to identify the type of each task. Then, in the next step, we will review the manual tasks and assess whether we can find a way to hook up these tasks to the BPMS. If this is not possible, we will have to consider whether or not it is convenient to automate the rest of the process without these manual tasks.

Once we have identified the type of each task, we need to check whether we can link the manual tasks to the BPMS, so that we can maximize the value obtained by the BPMS. Alternatively, we need to isolate these tasks so that we can automate the rest of our process. There are two ways of linking a manual task to a BPMS: either we implement it via a user task or via an automated task. If the participant involved in the manual task can notify the BPMS of the task completion using the worklist handler of the BPMS, the manual task can be turned into a user task. For example, the warehouse worker performing task "Retrieve product from warehouse" could check-out a work item of this task from their worklist to Process Automation indicate that they are about to perform the job, manually retrieve the product from the shelf, and then check-in the work item back into the BPMS engine. Alternatively, check-out and check-in can be combined in a single step whereby the worker simply notifies the worklist handler that the job has been completed.

Once we have established the automation boundaries of the process and reviewed manual tasks, we need to check that our process model is complete. Often business oriented process models neglect certain information because modelers deem it is not relevant for the specific modeling purpose, they assume it is common knowledge, or simply, they are not aware of it. It may be fine to neglect this information in a business-oriented model, depending on the application scenario. However, information that is not relevant in a business-oriented model may be highly relevant for a process model to be executed.

In the last step, we need to specify how each model element is effectively implemented by the BPMS. The implementation details are called execution properties. More, specifically, the execution properties are:

- Process variables, messages, signals and errors
- Task and event variables and their mappings to process variables
- Service details for service, send and receive tasks, and for message and signal events
- Code snippets for script tasks
- Participant assignment rules and user interface structure for user tasks
- Task, event and sequence flow expressions
- BPMS-specific properties

## 3.Conclusion

This paper presents an overview of the topic of business process management and business process management systems. The BPMS architecture is defined and analyzed.

There are many reasons for considering process automation. First, it provides workload reduction in terms of coordination: work is assigned to process participants or software services as soon as it is available. Second, it offers integration flexibility. Processes can be changed with significantly less effort as compared to legacy systems, provided they are explicitly represented via process models. Third, the execution in a BPMS generates valuable data on how processes are executed, including performance-relevant data. Finally, BPMSs improve the quality of process execution as they directly enforce rules such as separation of duties.

On the basis of the analyzed BPMS architecture and BPM itself, a four method steps are proposed - identify the automation boundaries, review manual tasks, complete the process model, specify execution properties. By executing these steps a business-oriented process model could be easily transformed into executable one, so it can be interpreted by a BPMS.

## References

1. Dumas, M., La Rosa, M., Mendling, J. and Reijers, H., Fundamentals of Business Process Management. Springer, 2013.
2. Panagacos, T., The Ultimate Guide to Business Process Management: Everything You Need to Know and How to Apply It to Your Organization. CreateSpace Independent Publishing Platform, 2012.
3. Pritchard, J. and Armistead C., Business process management – lessons from European business. Business Process Management Journal, 1999.
4. Armistead, C., Principles of business process management. Managing Service Quality, 1996.
5. Armistead, C., Pritchard, J-P. and Machin, S., Strategic Business Process Management for Organisational Effectiveness. Long Range Planning, 1999.

# Metadata Repository for Storing Multidimensional Schema Elements

**Geno Stefanov**

UNWE, 1700 Sofia, Bulgaria
genostefanov@unwe.bg

**Absract.** The multidimensional model introduces, in addition to the ER data model, dimension and fact entity, which are results of the process of dimensional modeling. Data warehouses and data marts are built on the basis of the multidimensional data model. The problem of integrating autonomously developed data marts can arise when an organization needs to merge data resides in independently developed data warehouses in order to make a better decision in a given situation. During this integration several categories of heterogeneities can arise, because of the complexity in the multidimensional data model. In order to detect these heterogeneities, a repository for storing metadata about multidimensional schema elements is built. This paper proposes an ER model of a *metadata repository* for keeping and using information about the multidimensional schema elements. The first section of the paper presents the literature overview on the topic. Then the proposed *metadata repository* is defined and explained.

**Keywords.** Data warehouse, data marts, integration heterogeneities, multidimensional metadata.

## 1.Introduction

The multidimensional model (MDM) on which the concept of DW and DM are based is built up of three basic constructive elements: the facts which are analyzed, the dimensions (coordinates of the fact) and the measures which allow for the quantitative evaluation of the facts [1]. Very often in large organizations the need for integrating independently developed DMs arises. These data marts should be based on common dimensions and facts, but in many cases different departments of one company develop their own data marts and their integration becomes a difficult task. The difficulties come out of the heterogeneities of the MDM elements and can be classified as dimension heterogeneities and fact heterogeneities from a semantic point of view. The dimension heterogeneities occur when the dimension schema structures, dimension members or the naming of semantically related dimensions have semantic discrepancies. The fact heterogeneities occur when the measures in different DMs are in different names, values (inconsistent measures), formats or even different scale. Some work has been done on the problem of resolving the conflicts occurring in data mart integration [2], [3], [4] and [8]. The main part of the DMs integration is resolving the possible heterogeneities. To the best of our knowledge there are no proposals for methods or techniques which will allow the DMs integration process to automatically identify the possible heterogeneities between the integrated MDM schemas. The author in [10] proposes a conceptual architecture of a system for detecting or identifying the possible heterogeneities between MDM schemas. One of the component of the conceptual architecture is the so called metadata repository where the MDM schema elements are stored for analyzing and determining possible heterogeneities.

The goal of this paper is to propose and define a metadata repository for MDM schema elements.

The rest of the paper is organized as follows: Section 2 introduces a formal definition for the MDM schema elements in the perspective of a DM. In Section 3 the metadata repository is presented and explained. Finally in section 4, some conclusions and future work are inferred.

## 2.Data Mart Elements

Intuitively, a DM defines one or more measure variables within fact tables, categorized by some dimensions that are organized in hierarchies of levels. Facts and dimensions consist of both their schema and the corresponding instances. For the dimension instances, we use the commonly accepted term dimension members (or members for short) [7] throughout the paper

In order to achieve the goal of the current paper, we have to analyze and define the constructing elements of one Data Mart, respectively of one MDM schema. A DM = $\{F_{1,}..., F_n, D_1, ..., D_m\}$, consists of non-empty set of *dimensions* and *fact tables* [4], [5].

Now let $\{\tau 1, ..., \tau m\}$ be a finite set of data types (e.g.integers) with their domain defined by function $dom(\tau)$.

A dimension $D \in \{D_1, ..., D_m\}$ of DM is composed of:

- The dimension schema $S_D = \{L_D, S(L_D), H_D\}$ containing finite, non-empty set of Levels $L_D = \{l1,…., l_m, l_{All}\}$, with level schema $S(L_D) = \{Sl_1,….,Sl_m\}$ and the roll-up hierarchy $H_D \subseteq L_D \times L_D$.
- The level Schema $Sl \in S(L_D)$ of some level $l_i$ is an attribute schema $(K,N_1, ……, N_k)$ with name l, key K(the dimensional attribute) and optional non-dimensional attributes $N_{1,...} N_k$. Every $N_k$ attribute $\in$ Sl has a domain $dom(N_k) = dom(\tau_k)$.
- The dimension instance $d(S_D)$ over schema $S_D$ with name d containing a set of members Vd with each $v \in Vd$ being a tuple over a level schema Sl, and a family of "roll-up" relationships between the member subsets, defined by the dimension hierarchy.

A fact table $F \in \{F_1, ..., F_m\}$ of DM is composed of:

- The fact scheme $S_f = \{ A_f , M_f \}$containing a finite, non-empty set of dimensional attributes $A_f = \{ A_{1,...} A_n\}$ and a set of measure attributes $M_f = \{ M_{1,...} M_n\}$. Each $Ai \in A_f$ is linked with a level $l \in L_D$ of the dimensions D, each $Mj \in M_f$ with a $\tau j$ . The domain of the attributes in $A_f$ is defined as $dom(Ai) = members(l)$ and $dom(Mj) = dom(\tau j)$.
- The fact instance $f(S_f)$, a set of tuples over $\{[dom(A_1) \times ... \times dom(A_n)], [dom(M_1) \times ... \times dom(M_m)]\}$. A tuple $f \in f(S_f)$ is called a "cell" or "fact". Moreover, we call the values $[f(A_1), ..., f(A_n)]$ the "coordinates" of a cell, modelling the multi-dimensional context for the measures $[f(M_1), ..., f(M_m)]$

Based on the formal definition of DM elements we will propese and define the metadata repository.

## 3.Metadata Repository

The metadata repository aims to provide suitable form for storing the MDM schema elements defined in the previous section. In order to achieve that an ER data model is proposed (fig. 1). The metadata repository is composed of 5 entities:

1. Dimensions
2. DimensionAttributes
3. DimensionAttributesType
4. Facts
5. FactAttributes

Each entity stores data about some of the MDM schema elements.

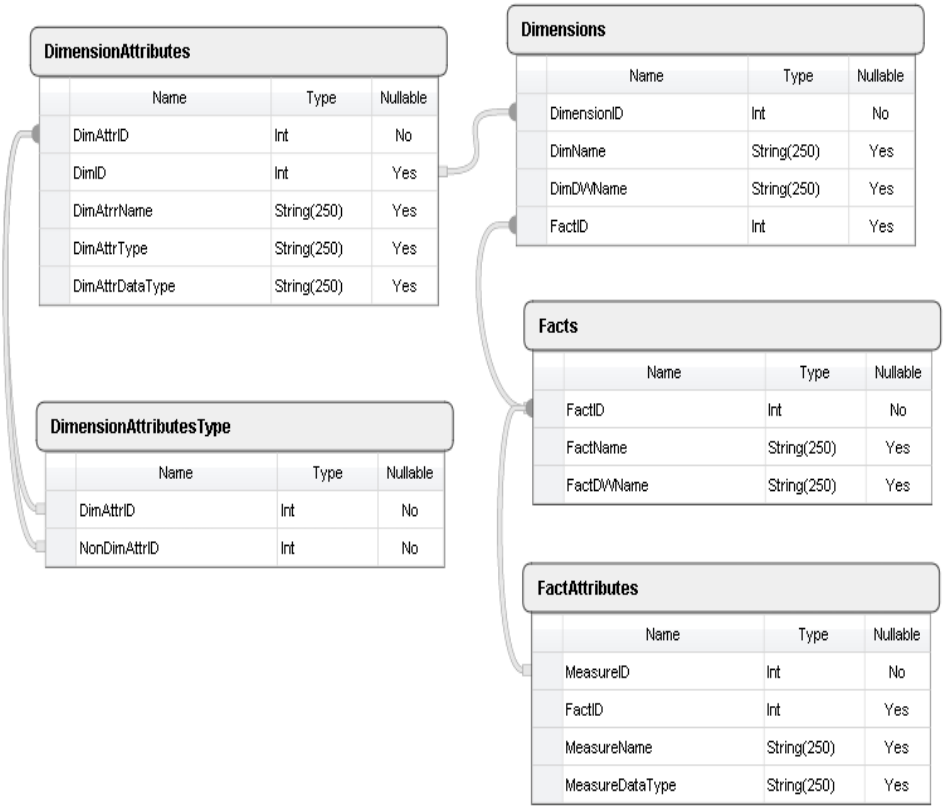



Figure 5. Metadata repository ER schema

The dimensions specific characteristics and properties which should be taken into account during the realization of the metadata repository are:

- Every dimension has a name and a relationship to some fact table, as well as belonging to exactly one data mart.
- Every dimension is composed of two attributes types –dimensional and non-dimensional attributes. The dimension hierarchy is composed of dimensional attributes. The non-dimensional attributes additionally describe some

dimensional attribute. Every attribute has its own domain described as its data type.

- One dimensional attribute may have many non-dimensional attributes, while one non-dimensional attribute can describe only one dimensional attribute.

These dimension characteristics determine the structure of the ER entities where the dimension elements will be stored.

On the other hand we should also take into account the fact tables specific characteristics and properties. They can be summarized as:

- Each fact table has a name and a belonging to a data mart.
- A fact table is composed of dimensional attributes and measures. Dimensional attributes correspond to the base level of the associated dimension. Each attribute has a corresponding domain described as a type of data

## 4.Conclusion

In this paper we have treat the problem of data marts integration and the corresponding heterogeneities. A metadata repository for storing the data mart MDM elements was proposed. By storing data about the data mart elements we will enable the automatic detection of the possible heterogeneities. This will shorten the process of data mart integration and the risk of mistakes.

The metadata repository is design as ER database and takes into account the dimension and fact table specific characteristics and properties.

## References

1. Vassiliadis, P. and Sellis, T. K.: A survey of logical models for OLAP databases. (1999)
2. Tseng, F. S. C., Chen, C-W.: Integrating heterogeneous data warehouses using XML technologies. Journal of Information Science, 31(3):209-229 (2005)
3. Torlone, R.: Two approaches to the integration of heterogeneous data warehouses, Distrib. Parallel Databases 23(1):69-97 (2008)
4. Berger, S. and Schref, M. (2008). From federated databases to a federated data warehouse system. HICSS, 0:394.
5. Golfarelli, M., Maio, D., and Rizzi, S. (1998). The dimensional fact model: A conceptual model for data warehouses. Int. J. Cooperative Inf. Syst., 7(2-3):215-247.
6. Lazarova.V. 2012. Multitenancy in Cloud Computing. In: International Conference on Application of Information and Communication Technology and Statistics in Economy and Education, October 5-6 th, 2012, Sofia, UNWE
7. P. Vassiliadis and T. K. Sellis. A Survey of Logical Models for OLAP Databases. SIGMOD Record, 28(4): pp. 64–69, 1999.
8. Diamantini, C., Potena, D., Sterfanov, G., Two approaches for resolving dimension schema conflicts(2012) In: UNITE 2nd Doctoral Symposium, 11-12 October 2012, Sofia, Bulgaria
9. M. Jarke, M. Lenzerini, Y. Vassiliou, and P. Vassiliadis. Fundamentals of Data Warehouses. Springer, 1997.
10. Стефанов, Г., Концептуална архитектура на система за определяне на хетерогенности при интеграцията на Екстракти от данни, Международна научна конференция "Информационните технологии в бизнеса и образованието", Варна, 2014.

# Warehouse Design in Natural Gas Sector's Enterprise in Bulgaria

**Ivan Andonov, Monika Tzaneva**

UNWE, 1700 Sofia, Bulgaria
Ivanandonov87bg@abv.bg, mtzaneva@unwe.bg

**Abstract.** This paper is a research in the area of natural gas sector's enterprise and aims to define what requirements are most essential in order to develop a sustainable warehouse in the process of implementation of Business intelligence in such a company. This study aims to define the scope and the sources of the data, the key requirements for wahrehouse data clansing , that should be considered during ETL development.

**Keywords.** Warehouse design, Natural gas sector's enterprise, Business intelligence.

## 1. Introduction

The development of sustainable data warehouse is not an easy thing to be done. The goal of this research is to point out the most important features in order to be created data warehouse where data should meet the following properties characterizing data quality:

- **Accuracy** – facts, values in the data warehouse should be compliant with the real – world.
- **Freshness** – data should be updated periodically in order not to be old.
- **Completeness** – there should be no missing things.
- **Consistency** – data representation should be uniform, identical.
- **Availability** – the data should be accessible.
- **Traceability** – sources of the data should be clear.
- **Clearness** – data should be understandable at the first glance.

In this paper under Data warehouse we apply the definition given by William Inmon who described it as "Collection of data that supports decision-making processes and provides the following features:

- It is subject orientated.
- It is integrated and consistent.
- It shows its evolution over time and it is not volatile." [1]

## 2. Key phases in Data warehouse prototype development

Key phases in data warehouse prototype development are related to the data warehouse lifecycle. Here it is essential what methodology would be applied in order to design this warehouse. There are two approaches – Bottom – Up and Top – Down. There are best practices that give advantages to each of these approaches depending on the case and the goal of the particular project.

Scientist Rob Mattison in his book "The data warehouse handbook" [2] describes a three phases lifecycle of the data warehouse:

• **Setting goals and planning** – This phase includes preparatory work and is based on feasibility analysis aimed at setting goals and selecting the proper approach to build warehouse.

• **Designing infrastructures** – In this phase tools and technologies are assessed, architectures solutions are compared in order to be created a plan.

• **Designing and developing data marts** – selection of the data and application development.

SAS institute applies a methodology called The Rapid Warehouse Methodology which consists of the following phases:

• **Assessment** – This is evaluation of the project i.e. Cost – Benefit analysis preparation in order to be compared estimated cost versus estimated revenue if the proper project is implemented.

• **Requirements** – Main points here are gathering of information for analysis, projects, and architecture for the systems.

• **Design** – The key element here is generation of logical and physical design.

• **Construction and final test** – Once the data warehouse is built and populated with data, it is time for users test.

• **Deployment** – It is time for training and start up.

• **Maintenance and administration** – this is one the longest phases and here upgrades and additional features could be added in order to meet new needs.

• **Review** – here the final check is done in order to be evaluated the project from economic point of view – is it profitable or not and what could be improved.

Once the methodology has been selected, it is not obligatory to follow all its phases and activities, but if the team which is in charge of data warehouse development is experienced and qualified, it is possible to choose other way of doing the things based on its knowledge. Our approach aimed to emphasize the most important steps, based on the business process specifics and available data sources of gas sector's companies Key phase of Data warehouse prototype development is the **Analysis of data sources and User requirements analysis**. Here under data sources analysis is important to be analyzed and understood available sources and to be discovered the relationships between these sources in order to get useful information. Another critical part of the job here is to assess which data should be selected and could be useful for the purposes of the decision – making processes. If multiple data sources are available, they should be integrated somehow in order to be useful.

Once the analysis of data sources is done, it is time for end users requirements analysis. It means that relevant information should be gathered in compliance with the strategic goals in order to develop a sustainable warehouse. Important part of this phase is the detailed description of all processes in the company in order to be analyzed the all possible KPIs that could be measured. Information could be selected by interviews and the tools for the interview are open – ended questions, closed questions, evidential questions and others.

After these things are done, it is time for the design of the warehouse – key elements here are Conceptual, Logical, Physical and Data – staging design. Once all is done the final part of data warehouse development is the documentation of the whole project.

## 3. Development of data warehouse prototype in natural gas enterprise

Company from the natural gas is generating a huge amount of data every day. That's why it is essential to be evaluated which one is important and could be selected in KPIs calculation and which is not. The main source of the data are the reports of the company. From decision making point of view the important thing is to be secured that the data on the base of which KPIs are calculated is correct, accurate and updated.

During the development of the data warehouse interviews with the employees of the company were conducted and annual reports were investigated in order to get the full and clear picture of which KPIs are important for such an enterprise.

Based on the data found in the annual reports, it was decided to design a star schema of the data warehouse prototype.

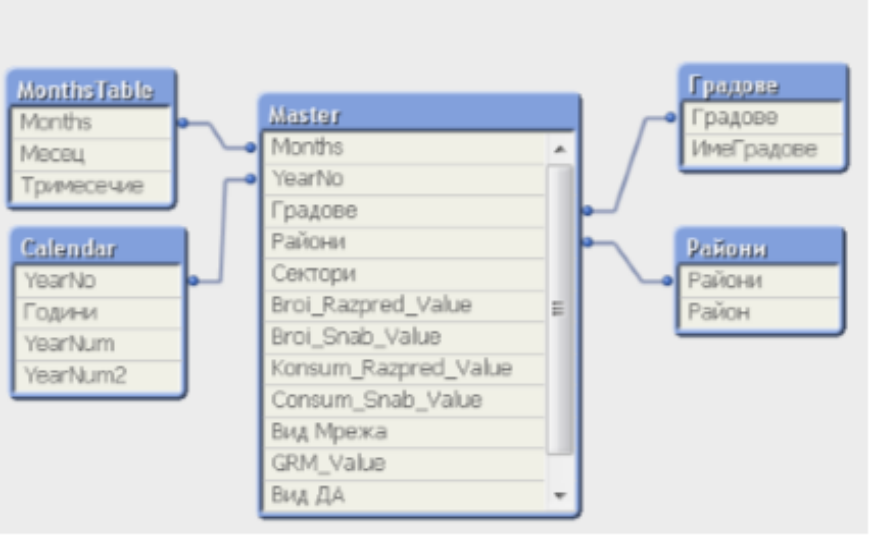



Figure 1 Star schema of the data warehouse prototype

The software that is used in order to be developed application to measure KPIs is Qlik View. This product was chosen because it is one of the leading solutions according to the Gartner Group [4].

Once the data is collected from the annual reports of the company it is time for the ETL process in order to guarantee that the data is correct, updated and consistent.

During the ETL process it is essential to check for:

- Duplicate data.
- Inconsistent values that are logically associated.
- Missing data.
- Unexpected use of fields.
- Impossible of wrong values.

Once the data is cleaned it is time for transformation which means that during this phase data is converted from operational sources format into a specific data warehouse format. For example: Different formats could be used for individual data.

Loose text may hide valuable information, that's why it should be avoided.

Loading process – it should run smoothly and should guarantee updated data.

The application in Qlik View is developed as follows – Once the data is cleaned , it is loaded in qvw files. The script of the data is shown in figure 2.

```
Investment:
CrossTable(YearNo, Investment_Value, 3)
LOAD Райони,
     Градове,
     [Вид ДА],
     [1],
     [2],
     [3],
     [4],
     [5],
     [6],
     [7],
     [8],
     [9],
     [10],
     [11],
     [12]
FROM
[..\Data\овергаз.xlsx]
(ooxml, embedded labels, table is invest);

```

Figure 2 script for loading of data in qvd file.

The next step is loading of qvd files in the final file where all KPIs will be calculated and shown in a user friendly way. The final outcome is shown in figure 3:

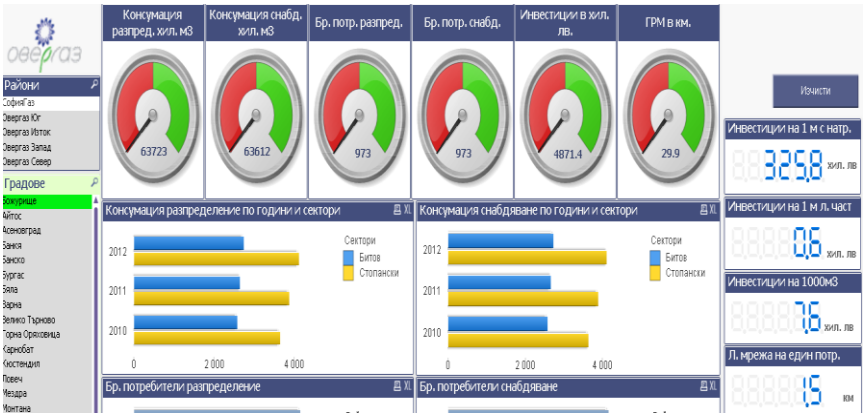



Figure 3 Application in Qlik View for KPIs measure

## 4. Conclusion

Data Warehouse prototype development is a complicated and time consuming process and when it is decided to create Data warehouse which will have to bring value to the company as a result of better quality of the analysis prepared in the BI platform.It is essential to build a team which is very familiar with the business processes in the company. The KPIs which are important from strategic point of view should also be well analyzed and all Data sources should be investigated in order not to miss anything. The next key step, after data source analysis and end user requirements analysis, is to design the project which will integrate all the data sources in order to provide valuable analysis based on current data. ETL process is another important part of creation of sustainable data warehouse. The final phase is the development of a user friendly application which will generate the value for the company that took the decision to develop sustainable data warehouse.

To sum up, important things are:

• Choosing the proper stages and activities of data warehouse design life cycle, based on business process and available data sources specifics

• Building a team of experts – with strong IT background, experts who are qualified experienced and know how job should be done.

• The core team must include experts who know very well the processes in the company in order to describe the processes from strategic point of view, what should be included in the data warehouse and what not.

When all things are done by taking into consideration all mentioned above, success is most probable thing to happen.

## References

1. Willian Inmon : Building the data warehouse, John Wiley and Sons, 2005.
2. Rob Mattison : The Data Warehouse Handbook, 2006
3. Annual reports of Overgas Inc. AD available at www.overgas.bg
4. Gartner magic quadrant as of February 2014 available at http://www.gartner.com/technology/home.jsp
5. Data Warehouse design – Modern principles and methodologies, Matteo Golfarelli and Stefano Rizzi, Mc Graw Hill, 2010.

# Model of a Real-time Application Based on SignalR Working with Big Missives of Data Which is Applicable in the Gaming Industry

**Iliya Nedyalkov**

UNWE, 1700 Sofia, Bulgaria
iliya@idnc.eu

**Abstract.** Nowadays real-time applications are vastly becoming popular. They've been used in the gaming industry for quite a while, especially within the live betting sections where each and every event (goal, point, card etc) would affect the odds presented by the bookmakers. The application has to analyze and process the upcoming missives of data and to only send the updates to the connected clients. Microsoft presented SignalR library as a method which enables rapid application development for real-time web applications. This paper will demonstrate a model of a real-time application matching the needs of the gaming industry.

**Keywords.** SignalR, Real-Time web applications.

## 1.Introduction

Nowadays real-time applications become more and more popular. They have been used in the gaming industry for quite a while; especially within the live betting sections where every situation, point or goal changes the odds of the book maker. Parsers are another type of real-time web applications which pull data from different places, analyze it, process it and display it to the users. There are other real-time applications which serve various communication channels, i.e. Skype, Viber, WhatsApp, all used by a number of people. The most popular e-commerce applications contain chats which enable the direct communication between the sellers and the customers. Microsoft presented SignalR library as a method to create real-time applications.

This paper will represent a model of a Livebetting application based on SignalR which:

1. Receives feed with odds from third-party provider
2. Analyzes the odds
3. Sends only the changed or new data to the connected clients

## 2.Gaming application overview

Every gaming site has a live-betting section. This is the most important and crucial part of any application because there are hundreds or thousands of simultaneous customers who can bet real-time. This section is meant to successfully deal with enormous number of requests and at the same time handle problems with the speed of displaying the changes of the odds' data. It should be taken into consideration that in games like tennis or basketball every point would affect all of the odds and if the system is not able to successfully update all of the info the bookies could suffer huge losses. During the weekends the Live-betting sections are extremely loaded since all the tournaments across the world are on and there are

thousands of simultaneous games in various sports, including but not limited to: football, basketball, tennis. Each and every game contains hundreds of odds for a number of scenarios.

Our SignalR application needs to receive the odds from a designated odds' provider. There are few big odds providers which deliver odds to the bookies. This is possible through API which allows the SignalR hub to pull the odds to process them and to send just the changes to the clients. In the majority of the cases API sends XML or JSON to the SignalR application which then would process the data and send only updated or new data in JSON format to its connected clients.

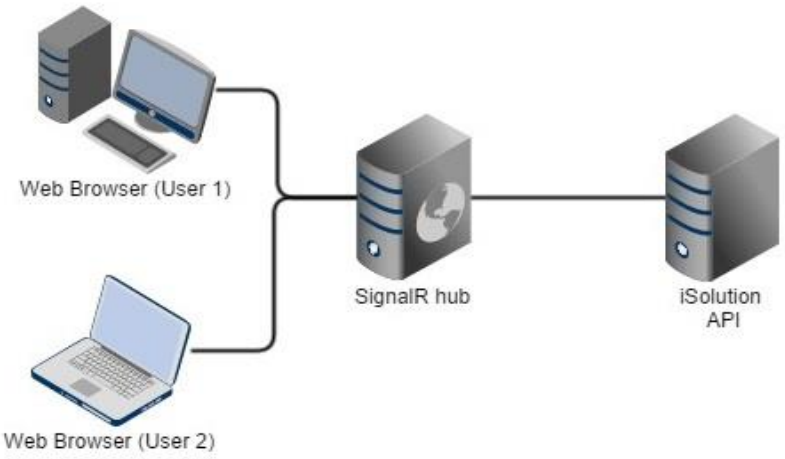



Fig. 1 Scheme of SignalR application used in the Gaming industry

## 3.How SignalR works

SignalR allows .NET developers to create real-time functionality in web applications in an extremely rapid manner. Real-time functionality means to have server code pushing content to connected clients instantly, instead of a server waiting for a client to request new data. SignalR is very usable in new type applications which needs frequent updates from the server application like live betting sections of the gaming sites, real-time gaming applications, dashboards, collaborative applications and many others.

SignalR application is based on .NET and could be console application or windows service. Web clients can be ASP.NET MVC application or a simple HTML / JavaScript page.

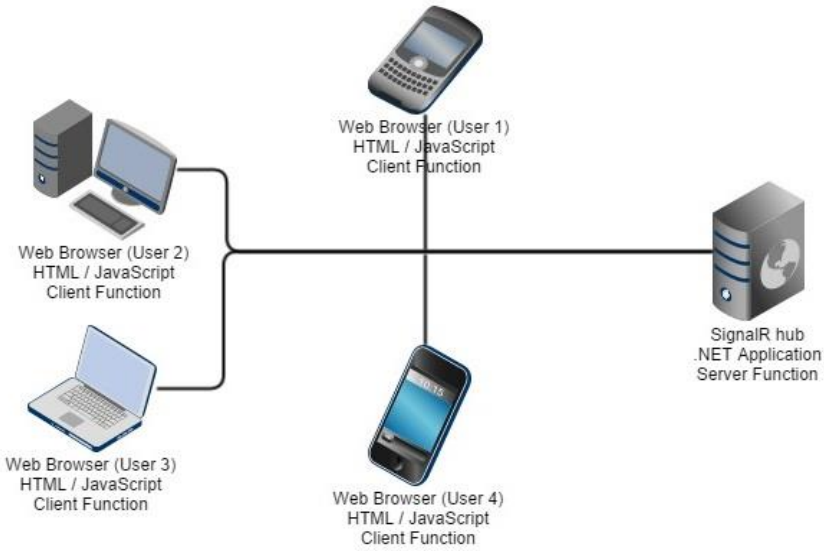



Fig. 2 Scheme of the communication between SignalR hub and its connected clients

## 4.Example of SignalR hub

Underneath is presented the source code of a simple SignalR. C# is the programming language used in the example.

```
namespace SignalRHub1
{
   [HubName("echo")]
   public class EchoHub : Hub
   {
      public void ServerMethod()
      {
         // Call the API
                  // Analyze data
            // Process data

         var caller = Clients.All;

         caller.clientmethod("Data for the clients");
      }
   }
}
```

The SignalR hub is a .NET application which acts as a server. In the afotmentioned example there is a method named ServerMethod() which can be called by the client application. The body of the method has logic which allows SignalR hub to call the Odds API, analyze data, process it and sends it to the connected clients through the "clientmethod".

## 5.Example of a client which consumes data real-time

In the listing below we present a sample of a client which communicates with SignalR hub. The client is HTML / JavaScript based.

```
<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
   <title>SignalRHub1</title>
   <script src="Scripts/jquery-1.6.4.min.js"></script>
   <script src="Scripts/jquery.signalR-2.1.1.min.js"></script>
   <script src="/signalr/hubs"></script>
</head>
<body>

   <input type="button" id="btnSend" value="Send" />

   <p id="message">

   </p>

   <script>
      $(function () {
         var hub = $.connection.echo;

         hub.client.clientmethod = function (msg) {

            $("#message").html(msg);
         }

         $.connection.hub
            .start()
            .done(function () {
               $("#btnSend").click(function () {
                  hub.server.hello();
               });
            });
      });
   </script>
</body>
</html>
```

The client application is based on HTML and JavaScript. It initializes a connection to the SignalR hub and after that it calls the server-side method ServerMethod(). In this method is the logic which allows the SignalR hub to call either the caller, group of clients or all the clients.

## 6.Conclusion

There are 2 main methods to create real-time applications. The old school one is based on requests repeated every second being send by the client to the server. The server response sends the new or updated data to the client. Even if there's no new data the communication has to be processed since there is no way of the client knowing. SignalR hub works on a different way – it calls the connected clients only in case there is a new or updated data. This decreases the levels of communication between the client and the server significantly, hence the server would work more efficiently. On this paper we described a real-time application which deals with big missives of data. The server uses the entire CPU resource to analyze and process the upcoming data and sends to the clients only updates.

## References

1. Ingebrigtsen Einar, SignalR: Real-time Application Development, 2013
2. Vespa Roberto, SignalR Real-time Application Cookbook, 2014.
3. http://www.asp.net/signalr
4. http://signalr.net/

# Regarding the Teaching of the Subject "STATISTICS" at the Universities of Economics in Bulgaria

**Andreana St. Stoykova – Kanalieva**

UNWE, Sofia, Department "Statistics and Econometrics"
an_stoykova@abv.bg

**Abstract:** The report highlights a discussion on ideas about a different approach to teaching the Statistics Subject at the Economics Universities. The thesis is presented, that Statistics is a science of a specific process of scientific knowledge, related to the study of mass phenomena, occurring in the existing universes. On the basis of the cognitive functions of analysis, the know-how and stages of conducting statistical surveys, and the statistical methods are proposed to be discussed and explained, according to the type of cognitive task, that should be solved in the course of the study.

**Keywords**: statistical survey; statistical universe; scientific knowledge process; know-how of conducting statistical surveys; statistical methods.

The question how to teach the subject "Statistics" to the students of economics is not new. Numerous discussions arise about teaching this subject, which is one of the compulsory subjects for all the students in all the universities of economics. One of the reasons is the not clarified question: "What kind of science is statistics?" on the one part, and the lack of clearly defined object of the statistics science on the other part (Many authors have written about this disputable issue. In my opinion the most important are the publications of V.Tzonev (1983,1985,1992,1996,2004), Iv. Saykova (2011,2012). I share their views.). Different opinions on this matter are available in the specialized publications). Without going into details I would like to point out that, statistics is a science investigating mass phenomena in all spheres of life presented as universes(Other authors like Friedman () think, that statistics is a methodological science. Others like V. Tzonev, Iv. Saykova consider, that it is a science with a material subject-matter, associated with studying mass phenomena, like universes. Wilde noticed, that we should "....make from statistics a science to study a process of scientific cognition.....Studying should become the heart of each statistics curriculum" (Quoted V.Tzonev, 1996:6, p.23).). In particular this is the science of designing, organizing, approbation and conducting research on occurring mass phenomena such as universes. In other words this is a science expressing a specific two-sided cognitive process: theoretical and methodological side (Although, at first glance, this definition seems to include a contradiction, this is not the case. Because the methodological nature of statistics is expressed in the fact, that 'the detected objective laws and methodological solutions become methodology for the study of mass phenomena in other areas of science and practice" (Iv.Saykova, 2011).) and as such it elaborates the theory and practical rules for the application of the statistical (aggregate) approach and the methods of testing the phenomena. The main problem is how to present the whole range of ideas, approaches, methods and techniques in such a way, as to make the students understand the essence of the statistical approach of treating the phenomena of the environment and the methodology of its practical application. This report is dedicated to this problem and how to resolve it.

## 1.

Currently the way of teaching the subject "statistics" to the students does neither motivate, nor stimulate statistical thinking. In general, it does not suggest the students the idea of realizing a specific cognitive process with specific cognitive assignments, approach, methods and means. Usually, this matter is considered in the academic courses, in such a way, that after introducing some basic concepts like universe, unit, attribute, generally the characteristic features of the variation distributions (distributions in static) and of the distributions in dynamics are studied The scientific literature is constituted in a similar way. Such an approach of treating the issues first in statics and then in dynamics, undoubtedly is reasonable, but the focus there is on the statistical methods, not on the cognitive process and the cognitive type of assignment, that should be solved by the specialist – economist or researcher and that should be solved by means of the statistical approach and its instruments. The methods are offered as books of prescriptions, supplemented with a set of standard statistical software programs, which facilitate the calculation procedures. Unfortunately in this case almost nothing is mentioned about the philosophy of the statistical approach and its cognitive capabilities(Such understanding of the subject of statistics is due to many not enough literate definitions of statistics. such as the example in Wikipedia, stating "Statistics is a mathematical discipline, that studies the extraction of information through analysis and interpretation of empirical data using the probability theory.... Statistics arised in association with the empirical science and differs from most branches of mathematics in its application orientation."(*underlined by me*)). The overall methodology to carry out such type of research is not discussed. The fact should also be added, that in some cases the problems of statistics are treated through the prism of probability estimated, without even mentioning, that they refer only to research, based on data obtained from representative samples. Actually the statistical knowledge is presented in a similar way in most of the statistics textbooks in foreign languages. Only some of the courses are an exception. Actually, in general the question about how to teach the "Statistics" subject to the students - economists should be discussed.

On the background of what has been said it should be noted that a number of authors in Bulgarian statistical literature(Whose ideas are shared by the author of this report.), such as V. Tsonev (1996; 2004; 2007a; 20076) Iv. Saykova (1994a;2006) A. U. Totev (1976), T. Kanaliev and others advocate a new approach to the teaching of the "statistics" subject in general and in each of its components, but so far their appeals and proposals have remained without any significant response (Other proposals are available for teaching the statistics discipline such as.D.Arkadiev (2001), E.Shkodrev(2005) etc. They will not be discussed here due to differences with the views of the author of this report.).

## 2.

What's missing in teaching statistics so far and how it can be overcome?

The statistics curricula still contain numerous obsolete views, hindering the students to cover the whole range of possible practical tasks, that use the statistical approach and implement the statistical cognitive process. These views refer to various aspects of this cognitive process. Without going into details, some of them will be referred below:

First, there is no introductory, initial comparative characteristics of the knowledge, acquired by means of conventional and statistical thinking. It is not demonstrated how statistical thinking differs from conventional thinking. In many cases the fact is not mentioned at all or just hinted, that the universe is the only form of studying mass

phenomena, that exist as sets of units. Yes, the concept "statistical universe" is introduced, but it is not enough to understand the essence of the statistical approach. The fact has not been clarified sufficiently, that the estimates obtained from research at the level of universes are valid only for universes as a whole, and not for the specific case. It is difficult for the students to realize the place of empirical frequency distribution as the only form of studying the characteristics of the universe and its characteristics as a whole. When listing the characteristics of the statistical universe it is hardly ever noted, that it has its own size. The only exception is the case, when treating problems of universes in representative sampling surveys. There isn't any broader interpretation of the concept of statistical universe, to enable the coverage of numerous important phenomena of our environment, by applying the aggregate approach. The place of the concept "statistical structure" is not clarified as well, despite of talking a lot about structures, changes in structures, disproportions etc. Most often this concept is associated with the relative frequency distribution of units in a universe. In this case the question is raised why two concepts are used to name the same thing? Where is the difference? In this regard the concept "statistical sub-universe" is also missing among the introductory basic concepts. The question about the statistical structures remains without explanation for the students (Ref. T. Kanaliev (1992;2003)). Obviously when defining basic concepts it is necessary to use a broader basis, in order to cover more concepts, to enable the future economists to apply a statistical approach of testing on a wider range of phenomena from reality.

Second, despite the numerous scientific publications and some written textbooks (see. Tzonev (1984a ) , Saykova (2011), Saykova, and Todorova (1994a) Stoykova-Kanalieva and Georgieva (1998) ets., the issue on the steps of statistical survey is still treated one-sidedly. The opinion is still widespread, that statistical survey has three steps: observation, grouping and analysis. Such an attitude to this question is not in conformity with the overall logics on a survey and does not enable the students to regard the statistical survey as an implementation of a cognitive process. What do I mean? Actually the technology of each survey consists of: a) a step of planning and preparing the survey in two aspects: theoretical-methodological and organizational - technical, along with provision of resources; b) step approbation of the concepts, foreseen in the above-mentioned preparatory step and c) step of actual implementation of the survey. The step of planning and preparing the survey is a very important part of it. As we know it is the starting point, the basis from which the researcher (specialist-economist) begins to study a specific phenomenon from reality. Here, in compliance with the necessity of an information the universe, the researcher's problem, the aim, tasks and size of the survey are defined in the theoretical and methodological part. The theoretical concepts of the object of the survey and the mode of operation and the instruments of the study are set out. A significant stage of this step is to identify the sources of data for the survey. As we know they could be diverse. Their choice depends on the aims and the tasks of the survey, on the cognitive functions of the survey, the management level, where the results would be presented, of its characteristic features and the essence of the studied universe and its components. In most cases the statistical observation is the most important source. Practically the statistical observation belongs to the three steps of the statistical survey, whereas at the preparatory step the observation program is developed, defining the universes, units and features, to be observed, the methods of the observation is worked out, defining the methods and the forms of registration of the primary data, time and place of the observation, selecting and preparing the observation staff, etc. At the step of pilot survey an experimental test is carried out of the quality of the elaborated methods of observation, of the preparation degree of the observation staff, etc. The observation itself as an action is a component of the implementation step of the survey. Studied through the prism

of the place of the statistical observation, obviously it has stages and elements, belonging to all the three main steps of statistical survey and could not be treated by itself as s separate step. The same refers to the issue of statistical grouping. The formation of the groups according to the meanings of one, two, three or more attributes, the creation of models of grouping tables, etc. belong to the preparatory step, and the approbation and the process of referring the data on the individual units to separate groups and their counting are elements of the other steps of statistical survey. The same refers to the statistical analysis. In other words, the statistical survey, as an overall cognitive process, consists of three steps (Nowadays, some authors (Ref. Iv. Saykova, 2011) speak of five stages, which actually broaden these three stages and make them more specific.), which however are not the traditionally mentioned ones and have much more stages and elements, than the usually discussed ones.

Third, in some cases, such as recurrent survey, or making use of data from previous surveys, some of the basic steps or some of their stages and elements are missing. This is a problem, that must be discussed with the students. Otherwise, when they encounter such a situation in practice, they would not be prepared for it. Moreover, in practice, the specialists economists rarely conduct overall statistical surveys. However they always encounter already estimated indicators, as a result of previous statistical surveys. It is important for them to be able to understand, interpret and use the data.

Fourth, when clarifying the essence and the contents of the statistical empirical frequency distributions the issue of the generalized statistical characteristics (the parameters) and their estimates is considered within comparatively narrow limits. Really we speak about estimates only when the survey is representative in nature. This in turn, limits also the presentation of the components of the estimate error to two: systematic and probable component and does not give a real notion about the possible other components of the error and their sources in the various types of statistical surveys. The question of statistical estimate should be placed on a broader basis. For all types of statistical survey we should talk about estimate. In view of the real aspects of practice, a broader interpretation of the concept “estimate” would enable the classification of the components of the error and their sources in all types of statistical surveys and each step of generation of statistical information. At present the students are left with the impression, that in the thorough studies in the course of generating statistical data, a systematic error is encountered in rare cases, and in representative surveys – just a probable error. Something that does not reflect reality. Special attention should be paid to the problem of errors in interpretation of the obtained results. As it is known, they can occur accidentally, due to misunderstanding of the obtained results and misleading interpretation, or consciously, for presentation of more pleasing results.

Fifth, one of the most important problems associated with teaching of the subject "statistics" is how to present the statistical methods. As already mentioned, the manner of presentation of the methods, is accepted by the students, and probably by the readers of academic statistical literature as well, as a certain fragmenting and prescriptive presentation of the matter. The methods are presented as prescriptions how to solve problems and as calculation procedures. At the same time, the type of cognitive task, that each of the methods is required to decide, isn't mentioned anywhere. Out of sight of the students remains the idea, that through these methods a cognitive process is realized, that they are part of an overall process for conducting the survey and the methods are the only tools for the implementation of this process. Presented in such a way, the matter of statistics seems difficult to study, with many formulas and detached from its essential part.

My profound inner conviction is, that the logic of exposing the methods should follow the idea of the type of cognitive problem solving, faced by the economist. Essentially what

should he learn about the mass phenomenon, subject to the statistical survey. If this is the task of fact-descriptive (situational) type, the analysis will cover the estimation of the characteristic features of statics or dynamics, without any descriptive nature, unlike the mean levels, the estimators of variance, of structure, the general estimators of changes and speed of changes of the phenomena in time, and growths, growth rates and growth, indices etc. The student should be aware of what do these characteristics show and how to interpret them. If the tasks are explanatory (diagnostic), then the methods are from the group of testing of hypotheses, analysis of dependencies, analysis of a steady trend, seasonal and cyclical fluctuations, analysis of internal regularities in the development of the phenomena, analysis of growth and differences etc. If the economist should solve an assignment of prognostic type, he should know and use different methods of forecasting and approbation of the correctness of the developed prognoses. The decision-making methods, in conditions of uncertainty (Favoring these methods is probably a result of some not enough substantiated and difficult to accept definitions of the nature of statistical science like:"Statistics is a science of managing uncertainty"(N.Fisher, 2001)), which recently have been promoted in a number of books, serve the optimizing function of the analysis and suggest previous descriptive, explanatory and predictive surveys. In general there should be continuity in the presentation of the methods and upgrading of knowledge, whereas the link is to decide what type of cognitive task is pursued. Only in this way each method will find its place and meaning in the system of knowledge of the students.

Sixth, I would not have passed the fact, that some of the statistical methods are presented in a way, that can hardly be bound by the idea of statistical universes and distributions. Such is, for example, the case with the well-known regression analysis, that is usually considered as a multidimensional aggregate, where the dependence is estimated, on the basis of one, two or more regression equations. In fact in this case the task is to find a model of accompanying changes in the sequence of universes, ordered by increasing values of explanatory attributes (Ref. V.Tzonev (1996)). The same refers to the survey of the laws of the development of the phenomena in dynamics (Ref. Z.Sugarev (1977,1978)). The essence and the cognitive capabilities of the very popular correlation analysis have remained without explanation. Usually, it is not mentioned, that it does not study and does not answer the question about the nature the tested relation. It is hardly mentioned, that the coefficients measure the approximation degree between the effect and the determining factors under quite restrictive conditions – normal distribution of the residuals. This is the reason why it is inadequate to use the correlation coefficients in non-linear relationships, like most of the relations in economics (Ref. Iv.Saykova et al. (2002)). I would focus on the disputable index theory, in which most statisticians do not like to recognize the simple fact, that it is a matter of a statistical analysis of changes in universes, caused by factors (Ref. V.Tzonev (1977;1997), T. Kanaliev (1977;1978;2005)) etc.

Seventh, the courses in statistics should focus especially on the interpretation of the obtained results. How should the students interpret the obtained indicators? Now, the results are interpreted partially (dispersed) after each method, without covering the obtained total result. Moreover such interpretations use specific terms like "statistically significant", "statistically reliable" etc. without interpreting in such a way, to be understood by non-statisticians. In fact this is one of the major problems in teaching statistics to specialist – economists. They should be able to understand the resulting summary characteristics, to compare them with other similar ones and should be able to detect emerging trends, objective laws and inconsistencies. In general when teaching statistics to the students - economists appropriate consideration should be given to the interpretation of the meaningful content of the result of the conducted statistical survey. The students should also be taught to

understand the data published in the specialized literature, the indicators and all the information products.

The problems discussed refer mainly to the education of students - economists. As far as the education of the students- statisticians is concerned, in addition to the above, there are a lot of other questions to be discussed, which will not be covered in this report due to the restrictions of the presentation. I am convinced, that there should be a difference in the way of presentation of the statistical matter in the introduction course in statistics for students statisticians and non-statisticians. Obviously it should be quite different. The reason thereto is, that the economists-non-statisticians and statisticians play different roles in the implementation of the statistical cognitive process. Therefore the introductory textbooks in statistics for specialists economists and for statisticians should have different contents with different focuses (A similar opinion is also supported by V.Tzonev and I.Saykova in a number of their publications (). Their concerns are completly backed up by the author of the report.).

The issues and the ideas, covered by this report, are some of all the studied ones in statistics. They are just an effort to improve the process of acquiring knowledge in statistics by the students.

## References

1. Аркадиев, Д. (2001) Някои размисли и виждания върху преподаването на статистика за икономисти. - Сб. Актуални проблеми на обучението на студентите по статистика в България и взаимодействието между статистическата теория и практика.С.,с. 19-31. Arkadiev. D. (2001) Some Views and Ideas on Teaching Statistics to Students of Economics – Collected Works: Current Issues of Teaching Students of Statistics in Bulgaria and Interrelation between Stistical Theory and Practice . Sofia, p.31.
2. Къналиев, Т. (2005) Съвкупностният (статистическият) подход на изследване и някои дискусионни въпроси на индексната теория , сп. Статистика, кн.1 ( Kanaliev, T. (2005) The Aggregate (Statistical) Approach in studying some debatable problems of the Index Theory , Statistics, 1)
3. Къналиев Т. (2003) Има ли подходящ измерител на структури и на структурни изменения в съвкупност? , сп. Статистика, кн.2 (Kanaliev, T. (2003) Is there a suitable Measurement of Structures and Structural Changes of Aggregates? Statistics, 2)
4. Къналиев, Т. (1992) Относно понятиета статистическа структура и измерителите на структура в статика, сп. Статистика, кн.5 (Kanaliev, T. (1992) On the Notion of Statistical Structure and Static Structural Measures, Statistics, 5)
5. Къналиев, Т. (1978) Относно построяването на индексните формули чрез последователно съблюдаване на съвкупностния подход, сп. Статистика, кн.1 (Kanaliev, T. (1978) About Construction of Formulae of the Indexes through Consecutive Observation of the Population Method, Statistics,1)
6. Къналиев, Т., (1977) За последователен съвкупностен подход при дефиниране съдържаниеуто на понятието “индекс” в индексната теория, Статистика, кн.2 (About the consistent General Method for determination of the Content of the Concept “INDEX NUMBER” in the Index Theory, Statistics,2)
7. Стойкова-Къналиева, А., Георгиева, М. Обща теория на статистиката, Изд. Мартилен, София (Stoykova-Kanalieva, A., Georgieva, M. General Theory of Statistics, PH “Martilen”, Sofia)
8. Сугарев,Здр. (1982) За периодните статистически съвкупности, Статистика, кн.2 (Sugarev, Zdr., 1982 On Periodical Statistical Aggregates, Statistics, 2)
9. Сугарев,Здр. (1977) Развитие във времето на статистически съвкупности и структури, Статистика, кн.5 ( Sugarev, Zdr., (1977) Development of Statistical Populations and Structures in Time, Statistics, 5)

10. Съйкова, Ив. (2012) Методологически и теоретични проблеми в статистическата наука, сп. Статистика, кн.1-2 (Saykova, Iv. (2012) Methodological and Theoretical Problems in Statistical Science, Statistics,1-2)
11. Съйкова, И. (2011) Проектиране на статистическото изследване, В:Живот отдаден на статистиката, Университетско изд. Стопанство, София (Saykova, Iv. (2011) Projecting Statistical Study, In: Life Dedicated to Statistics, Economics University PH, Sofia)
12. Съйкова, Ив., Стойкова-Къналиева, А.,Съйкова, Св. (2002) Статистическо изследване на зависимости, Университетско издателство" Стопанство",София (Saykova, Iv. Stoykova-Kanalieva, Saykova, Sv. (2002) Statistical Survey of Dependences, Economics University PH, Sofia)
13. Съйкова, Ив. (2006) Потребността от промяна при преподаването на статистиката във ВУЗ, В: Статистиката и предизвикателствата на информационното общество,Изд. Наука и икономика, Варна Saykova, Iv. (2006) Necessary Changes in Teaching Statistics to University Students In: Statistics and Challenges of Information Society, Science and Economics PH, Varna)
14. Съйкова,Ив., Тодорова,С. (1994a) Статистическото изследване (постановка, методи оценка на резултати), Издателска къща "Люрен", София (Saykova, Iv. (1994a) Statistical Survey (Views, Methods, Estimates of the Results, Luren PH, Sofia)
15. Съйкова, Ив. (1994) Преподаването на статистика за икономисти,сп. Статистика, кн.5 (Saykova,Iv. (1994) Teaching Statistics to Economists, Statistics, 5)
16. Съйкова, Ив., Къналиев, Т. Проблеми при подготовката на статистическите изследвания, Научни трудове на ВИИ"К.Маркс", т.II София (Saykova, Iv., Kanaliev, T. Problems in Elaborating Statistical Surveys, Scientific Works of the Higher Economics Institute "K. Marx", Vol. II, Sofia
17. Тотев, А. Ю (1976) Теория за масовото проявление на явленията Теория на статистиката. Теория на признаците, сп. Статистика, кн.3 Totev, A. Y (1976) Theory of Mass Manifestation of Phenomena. Theory of Statistics. Theory of Attributes. Statistics, 3)
18. Цонев, В. Съйкова. Ив., (2008) Статистиката- наука за проектиране, организиране и провеждане на изследвания на масови явления, сп.Статистика, кн.1 (Tzonev,V.,Saykova, Iv (2008) Statistics – A Science about the Design, Organization and Carrying out Mass Phenomena Investigations)
19. Цонев, В., Съйкова, Ив. (2007 а) Какво не достига днес в обучението на статистиците, за да бъдат те конкурентноспособни и ефективни в работата си? ( И защо кандидат студентите проявяват слаб интерес към специалността "Статистика"?), сп.Статистика, кн.2 (Tzonev,V., Saykova, Iv (2007a) ( What is still missing in the Current Education of Statisticians in order to be Competitive and Effective in their Work? And Why Student Candidates show little Interest in the Course in Statistics?, Statistics, 2)
20. Цонев, В., Съйкова, Ив (2007 6). Предизвикателствата на времето и подготовката на бъдещите поколения статистици, сп.Статистика, кн.4 (Tzonev,V., Saykova, Iv.) (2007 6)Challenges of the Time and the Education of Future Generation of Statisticians, Statistics, 4)
21. Цонев, В (2004) Отново на въпроса каква наука е статистиката и за подготовката на ново поколение проектанти на макроизследвания, Статистика, кн.5 (Tzonev,V., (2004) Once again on the issue what is "Statistics as a Science" and on the Training of a New Generation of Designers of Macro-studies, Statistics,5)
22. Цонев, В. (1997) Теорията на индексите и нейната статистическа алтернатива, сп. Статистика, кн.6 (Tzonev,V., (1997) The Index Theory and its Statistical Alternative, Statistics,6)
23. Цонев, В. (1996) Разгръщането на философията и логиката в понятийното пространство на съвкупностите като конкретноисторически процес, Статистика, кн.5,6 (Tzonev,V., (1996) The Unfolding of Philosophy and Logic into the Collectives' Conceptual Space as a Specific Historical Process, Statistics,4,6)
24. Цонев, В (1992) Общата теория на статистиката като философия и логика, Статистика, кн.5 (Tzonev,V., (1992) General Theory of Statistics as Philosophy and Logic, Statistics, 5)
25. Цонев, В. (1985) Понятието съвкупност и статистическото мислене, сп. Икономическа мисъл , кн.5 (Tzonev,V.,(1985) The concept of "Population"and Statistical Thinking, Economic Thought , 5)

26. Цонев, В (1984а) Предварителното моделиране на обекта в емпиричните изследвания, Социологически проблеми, кн.1 (Tzonev,V., (1984a) Preliminary Modeling of the Object in the Empirical Survey, Sociological Problems, 1)
27. Цонев, В (1984,6) Структура на статистическото мислене, сп.Статистика, кн.6 (Tzonev,V., (1984,6) Structure of Economic Thought, Statistics,6)
28. Цонев, В. (1983) Понятието съвкупност и общата теория на статистиката, сп. Статистика, кн.5 (Tzonev,V., (1983) The concept of "Population" and the General Theory of Statistics, Statistics,5)
29. Цонев, В. (1977) За двете нива на статистическия диагностичен анализ, Статистика, кн.5 (Tzonev,V., (1977) On the Two Levels of the Statistical Diagnostic Analysis, Statistics,5)
30. Шкодрев, Е. (2005) (Не)научните понятия в езика на (не)науката статистика В: Статистиката и предизвикателствата на информационното общество през XXI век (327-337), Наука и икономика, Варна (Shkodev, E. 2005. (Non)Scientific concepts in the Language of (Non)Scientific Statistics, In: Statistics and the Challenges of Information Society in XXI century (327-337), Science and Economics, Varna.
31. Fisher, N.I. (2001) Crucial issues for statistics in the next two decades, International Statistical Review,69,1, 3-4. International Statistical Institute
32. Friedman, J.H. (2001) The Role of Statistics In Date Revolution, International Statistical Review 69, 1, 3-4. International Statistical Institute
33. Wild, C. (1994) Embracing the Wilder View of Statistics" American Statistician,48,163-171
34. http://bg.wikipedia.org/wiki/статистика (accessed on 24.11.2014)

# Utilizing the Study of the Social-Economic Factors Influence on the Contemporary Bulgarian Human Values by the Use of IBM SPSS Amos Software

**Alexander Naidenov**,

Statistics and Econometrics Department, UNWE, Sofia, Bulgaria,
anaidenov@gmail.com

**Abstract.** The study of the Bulgarian human values main drivers requires the usage of latent variables in order to represent the hidden 'forces' that constitute these complex structures. When the standard statistical techniques cannot provide the adequate tools for that kind of study we need to use advanced methods such as structural equation modeling (SEM). Because of the demanding nature of the latter it is obvious that specialized software should be applied in order to solve the estimation procedure issues. One of the state-of-the-art solutions for the SEM application is the IBM SPSS Amos software package.

**Keywords:** social-economic factors, human values, structural equation modeling, specialized software, IBM SPSS Amos

## 1. Introduction

While interacting with each other we usually ask ourselves the question 'Why does he or she act like this?' One possible answer to this rather philosophical question is 'Because of his or hers value system'. Although there are many theories about what the value system is and the meaning of the basic human values, the famous scientist Shalom Schwartz summarizes [4]:

- Values are beliefs tied to emotion.
- Values are a motivational construct and refer to the desirable goals that people strive to attain.
- Values transcend specific actions and situations. They are abstract goals of abstract nature which distinguishes them from concepts like norms and attitudes, which usually refer to specific actions, objects or situations.
- Values guide the selection or evaluation of actions, policies, people, and events. They serve as standards or criteria.
- Values are ordered by importance relative to one another. People's values form an ordered system of value priorities that characterize them as individuals. This hierarchical feature of values also distinguishes them from norms and attitudes.

So we can conclude that the basic human values are '…desirable, trans-situational goals, varying in importance, that serve as guiding principles in people's lives'[5].

Using the sound foundation of a great number of multinational surveys (67 countries with more than 60 000 respondents) Schwartz defines 10 basic human values that constitute the value system. They are described in following table:

**Table 1.** Basic human values description

| Basic human value | Description |
|---|---|
| *Self-Direction* | Independent thought and action; choosing, creating, exploring. |
| *Stimulation* | Excitement, novelty and challenge in life. |
| *Hedonism* | Pleasure and sensuous gratification for oneself. |
| *Achievement* | Personal success through demonstrating competence according to social standards. |
| *Power* | Social status and prestige, control or dominance over people and resources. |
| *Security* | Safety, harmony and stability of society, of relationships, and of self. |
| *Conformity* | Restraint of actions, inclinations and impulses likely to upset or harm others and violate social expectations or norms. |
| *Tradition* | Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide the self. |
| *Benevolence* | Preserving and enhancing the welfare of those with whom one is in frequent personal contact. |
| *Universalism* | Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature. |

Each one of these basic human values interacts 'positively' with its 'neighbor' values and 'negatively' the opposite ones. To clarify that Schwartz builds a diagram of the theoretical model of the relations among the values [6], as this shown on Fig.1.

As it can be seen in the Figure 1, the values with close meanings also form so called value dimensions (openness to change, self-transcendence, conservation, self-enhancement) which are those of a higher order.

In order or measure those values in a real situation Schwartz builds so called the Portrait Values Questionnaire (PVQ). A shortened version of the latter, comprising 21 items (questions), is included in the wide-scope international European Social Survey (ESS). Using 2 or 3 items per each value it is possible to measure the extent to which each respondent identifies himself or herself with the person described in each one of the items. Each portrait description is evaluated by the use of the ordinal scale: *1 - Very much like me, 2 - Like me, 3 - Somewhat like me, 4 - A little like me, 5 - Not like me, 6 - Not like me at all* [2]. By the basic human values measuring we can build a picture of the human values prioritization for a given person, group or society as a whole.

Bulgaria has been participating in the ESS since the third wave of the survey i.e. 2006. The latest wave is the sixth one (2012) and by the use of the online available data in SPSS format it is possible to build a detailed portrait of the Bulgarian value system. For more information on the results see the Section 2 of this paper.

But answering the question about the main reasons for someone's behavior is not enough. It is also interesting to find out which are the main drivers (factors) that influence the prioritization of given values at the expense of other ones.

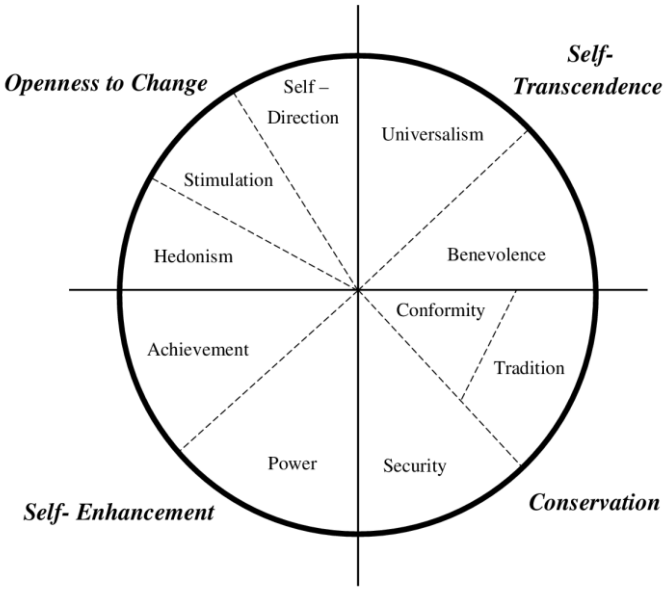



**Fig.1.** Theoretical model of the relations among the values

Schwartz [4], Baychinska [9], Garvanova [10], Inglehart [3] and Paunov [11] studied a great variety of factors that drive the formation of the human values system priorities. Considering their studies and the available indicators in the ESS questionnaire we can define 6 factor groups that can describe the 'structure' of the Bulgarian value system. The factor groups and the variables (characteristics) that constitute them are as follows:

**Table 2.** Factor groups and variables

| Factor group | Variables |
|---|---|
| *Demographic* | Sex, age, number of household members, subjective health status. |
| *Personal economic* | Employment last 7 days, cohabitation, children in the household, years of education, place of residence, place in the social hierarchy. |
| *Public economic* | Satisfaction with present state of economy in country and satisfaction with the national government. |
| *Personal social* | Satisfied with life as a whole, personal happiness, optimistism for the future, sense of direction in life. |
| *Public social based on trust* | Trust in: parliament, legal system, police, politicians, political parties. |
| *Public social* | Country overall democracy, health services status, free and fair elections, free media, citizens equality in court. |

Many studies prove that main socio-economic factors are not independent ones but intercorrelated with each other [8]. It is obvious that someone's age, education, employment status, social status, etc. define his or hers life satisfaction, trust in countries authorities, etc.

Therefore if we can build and estimate a model of the relationship between the human value system and the factor groups, considering the human basic values interconnections and factor groups' dependencies, we will be able to clarify the main reasons for the given values prioritization. Unfortunately in this case the use of the 'standard' statistical procedures and methods is not appropriate. The main reason for this is the 'latent' nature of the studied phenomena which means that we cannot observe immediately these variables (human values and factor groups) in the real world. In order to study them we have to measure observed variables defined in a questionnaire (e.g. ESS) and then to reveal the 'true' nature of the 'forces' behind them. An adequate solution to this issue is the usage of the structural equation modeling (SEM) methodology. The latter offers the possibilities to account for the 'hidden' nature of the studied phenomena considering the complex intercorrelations between them and the influence of the unknown events. The main advantages of the SEM methodology can be summarized to following [7]:

- SEM is a highly flexible and comprehensive methodology. The latter is appropriate for investigating achievement, economic trends, health issues, family and peer dynamics, self-concept, exercise, self-efficacy, depression, psychotherapy, and other phenomenon.
- Traditional methods specify a default model whereas SEM requires formal specification of a model to be estimated and tested. SEM offers no default model and places few limitations on what types of relations can be specified. SEM model specification requires researchers to support hypothesis with theory or research and specify relations a priori.
- SEM is a multivariate technique incorporating observed (measured) and unobserved variables (latent constructs) while traditional techniques analyze only measured variables. Multiple related equations are solved simultaneously to determine parameter estimates with SEM methodology.
- SEM allows researchers to recognize the imperfect nature of their measures. SEM explicitly specifies error while traditional methods assume measurement occurs without error.
- Traditional analysis provides straightforward significance tests to determine group differences, relationships between variables, or the amount of variance explained. SEM provides no straightforward tests to determine model fit. Instead, the best strategy for evaluating model fit is to examine multiple tests e.g. chi-square minimum (CMIN), root mean square error of approximation (RMSEA), goodness-of-fit index (GFI), etc.
- SEM resolves problems of multicollinearity. Multiple measures are required to describe a latent construct (unobserved variable). Multicollinearity cannot occur because unobserved variables represent distinct latent constructs.
- A graphical language provides a convenient and powerful way to present complex relationships in SEM. Model specification involves formulating statements about a set of variables. A diagram, a pictorial representation of a model, is transformed into a set of equations. The set of equations are solved simultaneously to test model fit and estimate parameters.

## 2. Application of the IBM SPSS Amos for the study of the social-economic factors influence on the contemporary Bulgarian human value system.

The application of SEM methodology for the study of the social-economic factors influence on the contemporary Bulgarian human value system is an uneasy task. The main reason for this is the complex modeling nature of the SEM, especially when there is also a

need for missing data imputation and model parameters estimation using asymptotically distribution free (ADF) method (ADF estimation is used due to the ordinal nature of the observed variables.). Because of the demanding nature of the SEM procedures it is obvious that specialized software should be applied in order to solve the estimation issues. One of the state-of-the-art solutions for the SEM application is the IBM SPSS Amos software package. Once an educational program for SEM in the 90s, now it is a stand-alone powerful software tool for design and visualization of complex phenomena relationships, building of a theoretical model of a simultaneous dependencies, automatic composition of input covariance matrix, model adequacy and consistency testing, data imputation, etc. [1]

The latest available version of IBM SPSS Amos software package at the time of writing this paper is the 22nd. The main interface of this product consists of two windows: input and output. At the *input window* the user builds the model (visually or as a source code), chooses the data and methods for estimation (OLS, WLS, ML, ADF, Bayesian, etc.) and adjusts data (imputation, specification search, multiple-group analysis, etc.). A screenshot of the input is presented at Fig.2.

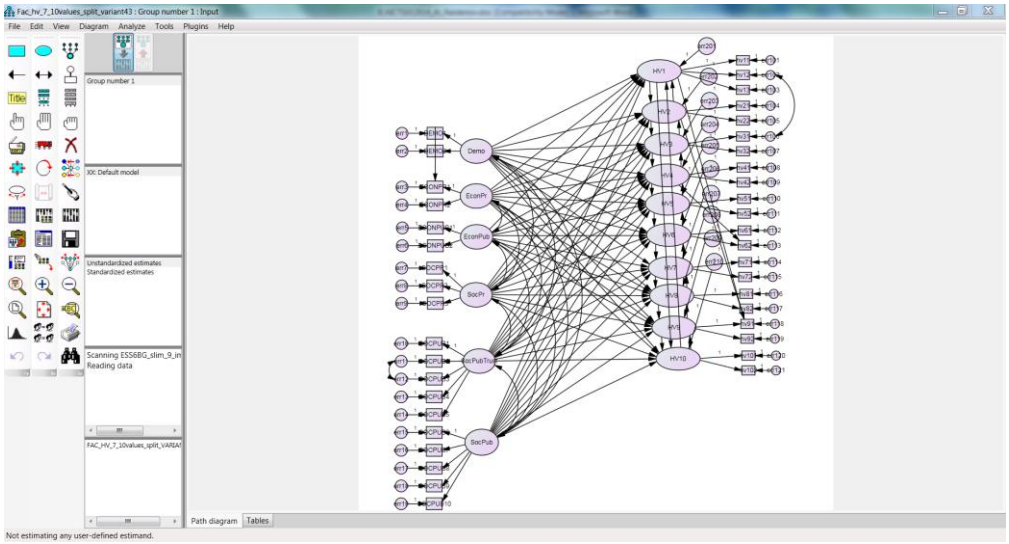



**Fig.2.** Screenshot of IBM SPSS Amos software input window

The *output window* shows the results of the model parameters estimation. For the sake of easy orientation in the 'rich' software output there is a tree-like list that describes the main sections of the output. A screenshot of the Amos output is given at Fig.3.

Although the software package has many advantages it suffers from some drawbacks too. The main cons are: the high computer hardware requirements especially when the user wants the results in foreseeable time while building a complicated model with a vast sample size. There is also a lack of the possibilities for the usage of so called polychoric correlation which is usually suitable for correlation measurement of ordinal type of variables.

By the means of the user-friendly graphical tools the model of the social-economic factors influence on the contemporary Bulgarian human value system is build and can be seen in Fig.4. The overall model consists of three parts: factor, dependent and structural (The ellipses in the model are used to visualize the latent variables and error terms and the rectangles picture the measured variables.).


| | DEMO1 | SOCPUB10 | SOCPUB9 | SOCPUB8 | SOCPUB7 | SOCPUB6 | SOCPUB5 | SOCPUB4 | SOCPUB3 | SOCPUB1 | SOCPUB2 | SOCPR3 | SOCPR1 | SOC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DEMO1 | 288,031 | | | | | | | | | | | | | |
| SOCPUB10 | -1,134 | 6,152 | | | | | | | | | | | | |
| SOCPUB9 | ,411 | 1,714 | 7,922 | | | | | | | | | | | |
| SOCPUB8 | ,602 | 2,666 | 2,778 | 7,706 | | | | | | | | | | |
| SOCPUB7 | -2,655 | ,920 | ,918 | 1,197 | 4,730 | | | | | | | | | |
| SOCPUB6 | -2,493 | 1,154 | 1,602 | 2,579 | 1,117 | 5,995 | | | | | | | | |
| SOCPUB5 | 1,415 | 1,242 | ,696 | 1,456 | ,917 | 1,361 | 3,809 | | | | | | | |
| SOCPUB4 | 1,271 | 1,312 | ,628 | 1,559 | 1,048 | 1,503 | 3,113 | 3,753 | | | | | | |
| SOCPUB3 | ,782 | 1,535 | ,945 | 1,984 | 1,348 | 1,717 | 2,602 | 2,876 | 8,156 | | | | | |
| SOCPUB1 | ,006 | 1,218 | ,830 | 1,775 | 1,220 | 1,510 | 2,829 | 3,098 | 3,088 | 4,960 | | | | |
| SOCPUB2 | -,080 | 1,847 | ,537 | 1,720 | ,973 | 1,419 | 2,595 | 2,735 | 4,501 | 3,207 | 5,556 | | | |
| SOCPR3 | -9,032 | ,644 | ,382 | ,828 | ,943 | ,863 | ,862 | ,882 | 1,002 | 1,064 | ,638 | 7,355 | | |
| SOCPR1 | -8,549 | 1,259 | ,574 | 1,576 | 1,846 | 1,696 | 1,167 | 1,307 | 1,378 | 1,563 | 1,167 | 3,597 | 7,161 | |
| SOCPR2 | -11,020 | ,842 | ,648 | 1,195 | 1,530 | 1,102 | 1,108 | 1,056 | 1,231 | 1,337 | ,931 | 3,868 | 4,865 | |
| ECONPUB1 | -3,396 | 1,256 | ,419 | 1,236 | 1,394 | 1,377 | 1,249 | 1,325 | 1,214 | 1,493 | 1,389 | ,989 | 2,202 | |
| ECONPUB2 | -1,412 | 1,434 | 1,222 | 2,312 | 1,729 | 2,050 | 1,666 | 1,858 | 2,086 | 2,452 | 1,766 | 1,376 | 2,265 | |
| ECONPR1 | -13,866 | ,359 | -,334 | ,647 | ,181 | ,522 | ,621 | ,608 | ,062 | ,957 | ,262 | 3,603 | 3,372 | |
| ECONPR2 | -5,584 | ,673 | ,392 | ,588 | ,675 | ,487 | ,672 | ,666 | ,607 | ,864 | ,549 | 2,804 | 2,627 | |
| DEMO2 | -9,723 | -,047 | ,098 | -,036 | ,100 | ,120 | -,071 | -,115 | -,047 | -,082 | -,119 | ,225 | ,017 | |
| hv31 | 2,655 | -,054 | ,154 | ,063 | ,145 | ,049 | -,033 | -,011 | ,107 | -,124 | -,118 | ,135 | -,077 | |
| hv102 | -3,130 | ,032 | -,033 | -,044 | ,102 | -,008 | ,159 | ,111 | ,050 | ,242 | ,024 | 1,141 | ,825 | |
| hv101 | -6,363 | -,016 | -,115 | -,098 | ,168 | ,082 | ,116 | ,119 | -,106 | ,214 | ,002 | 1,567 | 1,079 | |
| hv92 | -10,088 | ,131 | -,195 | -,129 | ,187 | ,100 | -,033 | -,041 | -,320 | ,017 | -,169 | 1,107 | ,975 | |
| hv91 | -7,233 | ,037 | -,026 | -,129 | ,174 | ,185 | ,069 | ,082 | -,082 | ,135 | -,068 | 1,300 | ,969 | |
| hv82 | -9,641 | ,055 | -,090 | ,004 | ,282 | ,096 | ,017 | -,020 | -,299 | ,012 | -,193 | 1,081 | ,827 | |
| hv81 | -10,031 | ,100 | -,145 | -,029 | ,306 | ,124 | ,044 | ,043 | -,162 | ,133 | -,040 | 1,248 | 1,090 | |
| hv72 | -5,370 | ,013 | ,005 | ,006 | ,189 | ,111 | ,106 | ,110 | ,077 | ,170 | ,049 | 1,146 | ,865 | |
| hv71 | -3,736 | -,048 | ,045 | ,011 | ,057 | ,030 | ,033 | ,037 | ,081 | ,106 | -,017 | 1,070 | ,784 | |
| hv62 | -,682 | ,050 | -,015 | -,091 | ,274 | ,165 | ,063 | ,115 | -,064 | ,136 | ,025 | ,491 | ,432 | |
| hv61 | -7,268 | -,215 | -,339 | -,258 | ,203 | -,072 | -,159 | -,168 | -,452 | -,100 | -,239 | ,529 | ,587 | |
| hv52 | ,355 | -,226 | ,127 | -,102 | ,029 | -,083 | -,013 | -,022 | ,121 | ,004 | -,046 | ,253 | -,056 | |
| hv51 | 1,000 | -,219 | ,235 | -,008 | ,139 | -,092 | ,002 | -,005 | ,014 | -,006 | -,137 | ,219 | ,015 | |
| hv42 | 3,123 | -,035 | ,205 | ,078 | ,112 | ,045 | ,032 | ,019 | ,147 | -,015 | ,023 | ,202 | ,053 | |
| hv41 | 4,874 | ,012 | ,238 | ,209 | ,138 | ,056 | ,068 | ,090 | ,186 | ,032 | ,066 | -,188 | -,263 | |
| hv32 | 2,659 | -,094 | ,187 | ,034 | ,212 | -,029 | ,055 | ,089 | ,260 | ,037 | ,063 | ,140 | -,009 | |
| hv13 | ,689 | -,188 | ,195 | ,015 | ,036 | ,098 | -,015 | -,040 | ,068 | ,001 | -,116 | ,415 | ,170 | |
| hv12 | 1,094 | -,215 | ,041 | -,049 | ,070 | -,103 | -,126 | -,092 | ,005 | -,096 | -,133 | ,329 | ,023 | |

**Fig.3.** Screenshot of IBM SPSS Amos software output window

The factor groups FG1…FG6 (latent variables) and the observed (measured) factor variables (MV19) are placed in the left part of the scheme and describe the correlations (covariations), pictured with two-way arrows, between the factor part elements (exogenous variables).

The 'dependent' part includes the latent (endogenous) variables for the 10 basic human values defined by Schwartz (HV1…HV10) and their 'measurement' indicators by the 21 observed items (MV20…MV40). As it can be seen all relationships between the values are visualized by the one-way arrows which express the dependence of one variable from another (a.k.a. regression weights).

The structural part shows the influence of the latent factor groups on the dependent latent human values structures.

In both factor and dependent parts of the overall model error terms are included (E1…E40). Their function is to account for the influence of unknown or not-observed events and also the possible measurement errors.

Indeed the model in Fig.4. is the final form of the investigated one and the steps to achieving it are as follows:

- *Step 1* - Building the theoretical model of all possible relationships between model variable, considering the theory and studies of the basic human values and the socio-economic factors;
- *Step 2* - Model specification by defining the equations 'behind' the theoretical model.
- *Step 3* - Data provision usually by random sampling design with minimum of 10 subjects per model parameter;
- *Step 4* - SEM requirements fulfillment check - sufficient sample size, no outliers, multivariate normality, no missing data, etc.;
- *Step 5* - Model identification i.e. to check if unique values can be found for parameter estimation. Degrees of freedom (df) are calculated and compared to zero. If df>0 then the model is an identified;
- *Step 6* - Parameter estimation by using the abovementioned methods by minimizing the differences between the population covariance matrix and the covariance matrix derived from the hypothesized model;

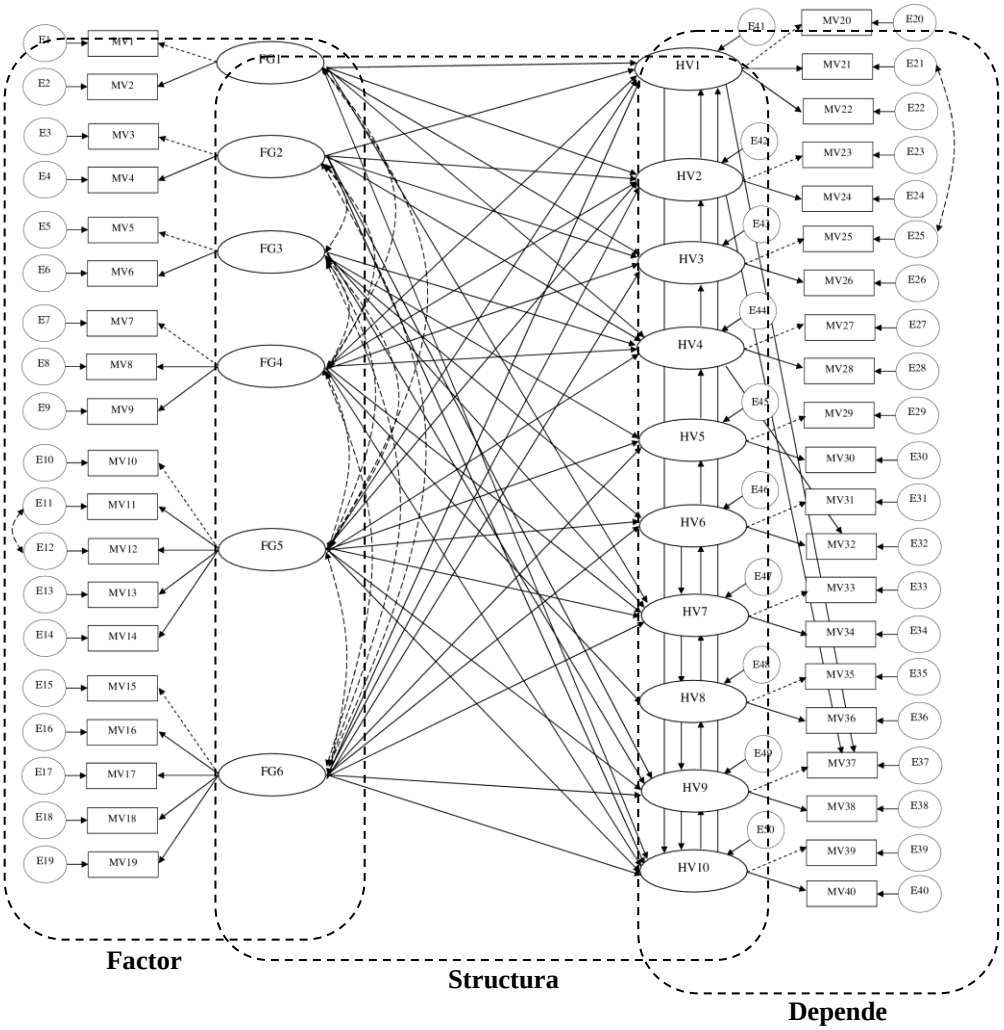



**Fig.4.** Graphical model of the social-economic factors influence on the contemporary Bulgarian human value system

- *Step 7* - Model fit assessment by using the fit indexes such as CMIN, RMSEA, GFI, etc.;
- *Step 8* - Model re-specification in case of not good enough fit, usually done by the use of so called modification indexes;
- *Step 9* – Model finalization and results interpretation.

Combining the data from the ESS wave 6, the theories of the basic human values and socio-economic factors, the SEM methodology and the IBM SPSS Amos software tool we obtain the desired model of influence (Fig.4). The main characteristics of the final model (re-specified several times using Amos modification indexes) are: CMIN/DF=6,594, RMSEA=0,049 and GFI=0,910. These show that the model can be accepted as adequate and the results can be correctly interpreted.

In short we can conclude that all factor groups influence the basic human values of Bulgarians. Some human values are influenced only by one factor (e.g. hedonism (HV8) depends on personal social factors only) but others are affected by at the most five of factor

groups (e.g. self-direction – HV10). The key drivers of the Bulgarian basic human values prioritization are personal social and public economic factors. However the latter do not have unidirectional influence e.g. in case of universalism value they have a negative effect but contrariwise the benevolence is positively affected by them.

## 3. Conclusion

In order to investigate the reasons for the basic human values prioritization in Bulgaria we need to use constructs of a higher level such as the latent variables. These 'hidden forces' measured by observed variables and influenced by unknown events cannot by investigated by the means of the 'standard' statistical methods. That is way the application of the structural equation modeling is required. The utilization of the study of the social-economic factors influence on the contemporary Bulgarian human values by SEM methodology raises the need for the usage of specialized software. One of the best available solutions for conducting SEM is IBM SPSS Amos. Although it provides the required software basis for the analysis, powered by user-friendly interface, it suffers from higher hardware requirements especially in the case of complex model estimation using a big data set.

## References

1. Arbuckle, J.: *IBM SPSS Amos 22 User's Guide*, IBM (2013).
2. European Social Survey: *ESS6 - 2012 Documentation Report*, European Social Survey (2014).
3. Inglehart, R.: *Changing Values among Western Publics from 1970 to 2006*, West European Politics (2008).
4. Schwartz, S.: *Basic Human Values: An Overview*, The Hebrew University of Jerusalem (2008).
5. Schwartz, S.: *Basic Human Values: Theory, Measurement, and Applications*, Revue française de sociologie (2006).
6. Schwartz, S., Boehnke, K.: *Evaluating the structure of human values with confirmatory factor analysis*, Journal of Research in Personality (2004).
7. Suhr, D.: *The Basics of Structural Equation Modeling*, University of Northern Colorado (2006).
8. Torgler, B., Valev, N.: *Gender and public attitudes toward corruption and tax evasion*, Contemporary Economic Policy (2010).
9. Байчинска, Кр: *Ценностите на българина и българската култура в прехода към демократично общество*, Монография (2000).
10. Гарванова, М.: *Трансформация в ценностите на съвременния българин*, Издателство "За буквите – О писменехъ" (2013).
11. Паунов, М.: *Ценности и ценностни приоритети на българите: особености, сходства и различия с други европейски държави*, Университетски издателство Стопанство (2009).

# Electronic Services at Public Institutions

**Katia Kirilova**

[1] UNWE, Department of Public Administration, 1700 Sofia, Bulgaria
kstrahilova@abv.bg

**Abstract.** The paper discusses the basic principles of administrative services. On this basis are analyzed the characteristics of administrative services. This is the basis for the search of approaches to the computerization of services. Part of this concept is the standardization of administrative and electronic services. In conclusion are formulated guidelines for development of electronic services at public institutions.

**Keywords.** Information Technologies, Electronic services, Public Administration.

## 1.Introduction

One of the serious problems facing public institutions is the introduction of e-government and computerization of administrative processes. In this connection it is significant and the question of expanding the number and quality of electronic services. In this respect, in recent years are necessitated a number of changes, mostly related to computerization and introduction of the principles of e-governance, which is discussed below.

## 2.Principles of administrative services

The main task of any administrative authority is carrying out administrative services to citizens and businesses. This is an inherent feature of any one administration (Fig. 1).


Fig. 1 Performed administrative services

The basic principles of administrative services are enshrined in the Administration Act, the Administrative Procedure Code and the Regulation on administrative services [3]. They are:

- Equal access to government services and information for administrative services;
- Various forms of access to administrative services and information about them;
- Polite and responsive;
- Coordination and interaction with all parties interested in the improvement of the administrative service in the administration, which ensures uniformity of results under the same circumstances;
- Reliable feedback;

- Implementation of commitments of administrative services to developed and announced rules;
- Quality of services provided.

## 3.Characteristics of administrative services

The concrete application of these principles is reflected in the characteristics and peculiarities of the different administrative services [1]. In this connection it should be pointed out that the administrative service may be:

- Issuance of individual administrative acts, which certify facts of legal significance;
- Issuance of individual administrative acts which recognize or deny the existence of rights or obligations;
- Performing other administrative actions that are legal interest for natural or legal person;
- Consultations represent a legitimate interest for a natural or legal person on administrative law, which are given by law or that are associated with the issuance of an administrative act or perform other administrative service;
- Expertise, represent a legitimate interest to natural or legal person when the normative act provides for their implementation as obligations of the administration of a public authority or by an authorized organization.

## 4.Standardization of administrative and electronic services

To enable the implementation of e-governance and the provision of administrative services electronically is necessary to carry out standardization of administrative and electronic services. This ceteris paribus creates opportunities for:

- Introduction of unified lists names of the administrative services;
- Uniformity in the way of description and metadata description of administrative services;
- Transfer of descriptions of administrative services between administrative authorities;
- Introduction of concept "version" of service maintenance descriptions versions and more.

According to the Ordinance on Administrative Services names of administrative services [3] in all informational materials, as well as in all electronic databases are used and filled in accordance with the List of Unified Names of Administrative Services (LUNAS). It is supported by the Administration of the Council of Ministers and published on the website of the Council of Ministers. Any administrative authority declares for entry in LUNAS names provided by its administrative services, and changes or deletions in them (Fig. 2).

The maintained list of Unified Names of Administrative Services contains information and description of these types of services (Fig. 3):

- Services provided by all administrations;
- Administrative structures accountable to the National Assembly;
- Central Administration;
- District Administration;
- Municipal Administration;
- Internal administrative services;
- Deleted services.

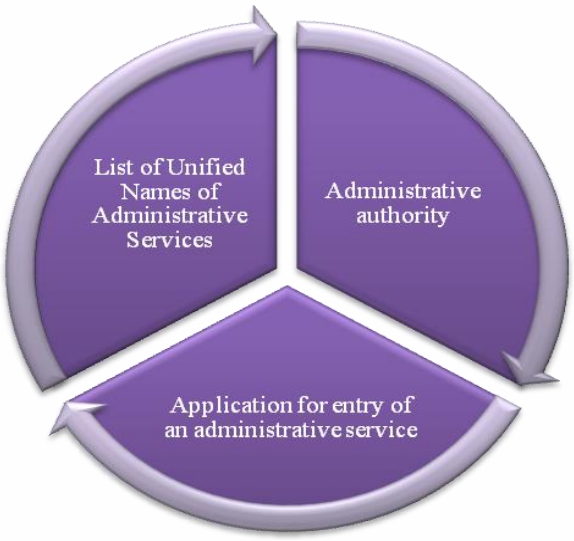



Fig. 2 Entry of an administrative service in LUNAS

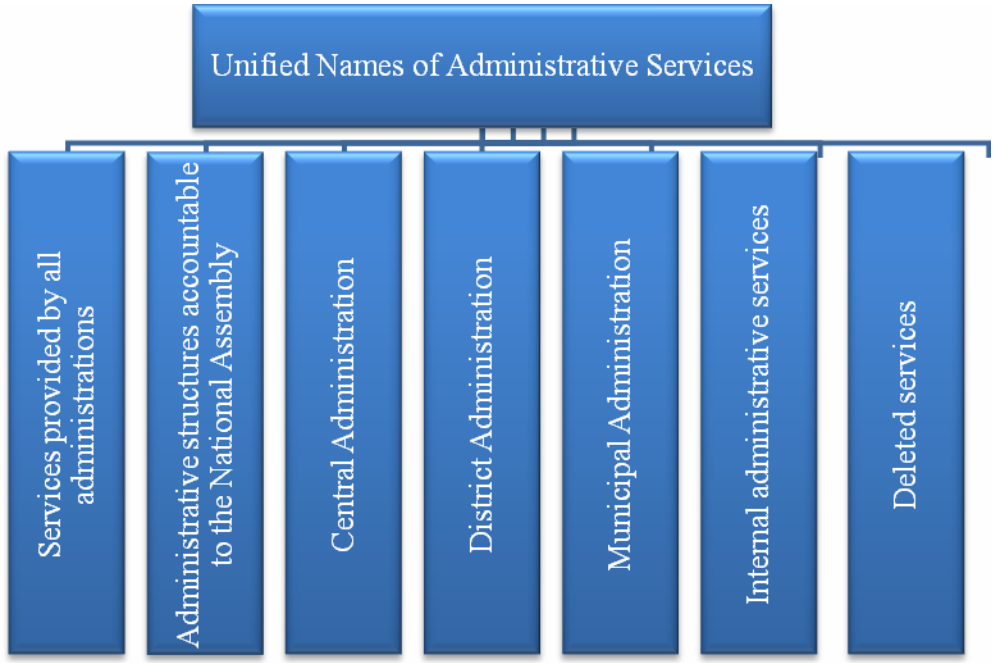



Fig. 3 Unified Names of Administrative Services

The current version of the list is available at the following Internet address www.government.bg/fce/001/0211/files/SUNAU_SSAO-2013_1405.xls (Fig. 4) [2].

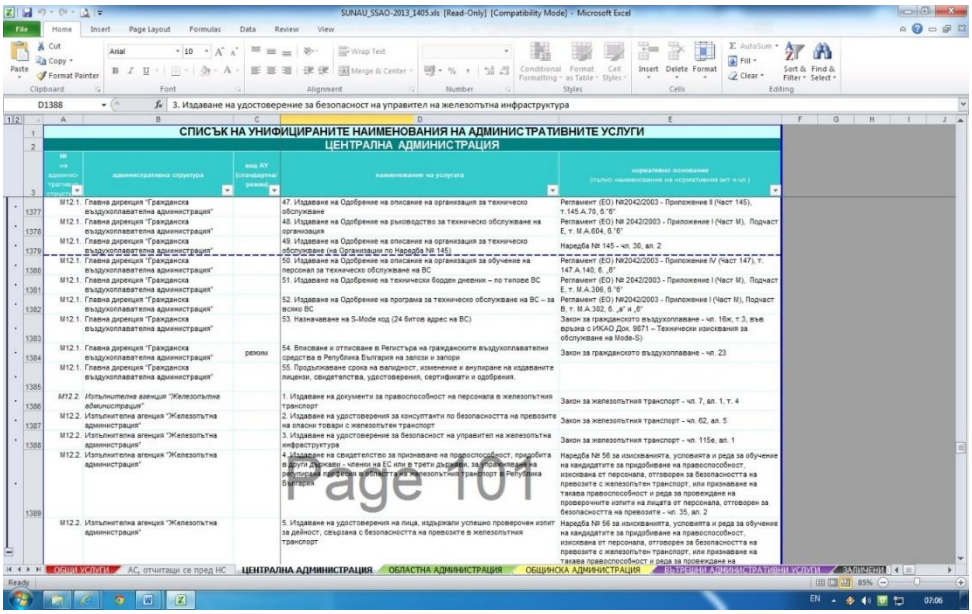



Fig. 4 List of Unified Names of Administrative Services

These and other problems show that in coming years many have to work to modernize the administration by introducing e-governance [1]. So over time we can successfully achieve the objectives of the related administration, reaching the ultimate goal "related administration" [4], [5].

## 5.Conclusion

Finally, it should be noted that the administration should begin implementation of systems for comprehensive e-governance. For this to be possible it is necessary to establish and implement common information models. Their computerization will eventually include a whole series of actions related to the introduction and use of databases, pass entirely into electronic registers and registration procedure, modeling for exchanging information, forming a complex electronic administrative services, etc.

## References

1. Kirilov, R., E-government in the public sector - challenges and paradoxes, International Conference on Application of Information and Communication Technology and Statistics in Economy and Education, October 5 – 6 th, 2012 UNWE, Sofia, Bulgaria,.
2. www.government.bg/fce/001/0211/files/SUNAU_SSAO-2013_1405.xls
3. www.lex.bg
4. http://ec.europa.eu/isa/documents/isa_2_proposal_en.pdf
5. http://ec.europa.eu/digital-agenda/

# Methodological Aspects of Development of New Electronic Services at State and Local Administrations

**Katia Kirilova**

UNWE, Department of Public Administration, 1700 Sofia, Bulgaria
kstrahilova@abv.bg

**Abstract.** The paper addresses the issues of the development of new electronic services. The paper presents a methodology and appropriate methods for the development of services. Analyzes are made in the meaning of the approaches of creating a vision for a new electronic service and implementation of a new web service on an existing concept. The paper gives information about the characteristics of the different stages of the life cycle. In conclusion are formulated guidelines for development.

**Keywords.** Methodology, Electronic services, Local Administration.

## 1.Introduction

Computerization of services in the public sphere shows the current approaches to modernization of governance levels and receipt of administration at the service of citizens and businesses. To enable the realization of this priority is necessary to develop traditional and new electronic services. The creation of an online service is consistent and purposeful process of converting consumer demand and objective necessity in a specific technological solution. For this purpose previously necessary to develop methodological tools and its application in concrete examples.

## 2.Methodological issues in the creation of new electronic services in public administration

In terms of the development of the information society are constantly changing information needs and the needs of citizens and businesses. Consumers increasingly have expectations when searching the Internet to find the ability to perform one or another administrative service remotely. Proof of this is the growing number of individuals who use the Internet for the purpose of online services, including in its interaction with public organizations, as well as increasing the share of people using electronic signatures [1].

Development of new electronic services by public sector organizations is necessary due to the ever-changing information needs of people and businesses. In the most general terms, we can consider that the methodology for the development of new electronic services can be analyzed in two ways:

- Creation of a concept for a new electronic service;
- Implementation of a new web service on an existing concept.

The first aspect reflects the main marketing aspects of consumer expectations for the realization of new electronic services and the second stage provides basic information aspects of the implementation of online services. Fig. 1 shows the main components of the first aspect in the creation of a concept for a new electronic service.

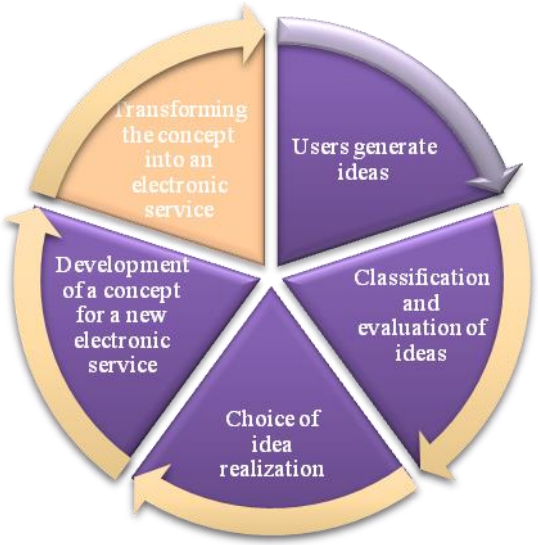



Fig. 1 Main components of the creation of concepts for new electronic services

Usually the generation of ideas based on ongoing processes in the environment and the wishes and needs of users (citizens and businesses). Practice shows that any change in the environment can be used as a source of ideas because the expectations and needs of users change. Classification and evaluation of ideas is necessary because not all ideas can be adopted and lead to the development of the concept. Some rejected if it meets the following organization of public policy and its possibilities. In this respect it appears a selection of ideas, they must comply with the objectives, policies and possibilities of public organization. Once you have selected the ideas necessary to develop a concept for a new electronic service. When it is carried out preliminary marketing study and develop different concepts for each idea is selected some of them, should be considered the expected effectiveness of the new electronic service for public organizations.

## 3.Transform the concept into a new e-service

Of special interest to the theory and practice is a process of transforming the concept in a specific electronic service. According to some authors, this process is very complicated and long [1]. For the purposes of this study may suggest the following sequence of steps (Fig. 2).

The first task of the process of transforming the concept of service is **planning to implementation and scheduling**. This is the first step in a series of activities in the provision of online services and information. As a rule, modern man expects to find any information on the Internet. In this sense, as the network is useful in nature, so the presence of "parasitic" information in it is constantly growing;

**Identification of key users of the new electronic service**. This was a kind of analysis of the relevant processes, since the determination of groups of stakeholders associated with measures to safeguard security. Along with this wide range of stakeholders can lead to some technical problems when working with the system, such as the reduction of its rapid action.

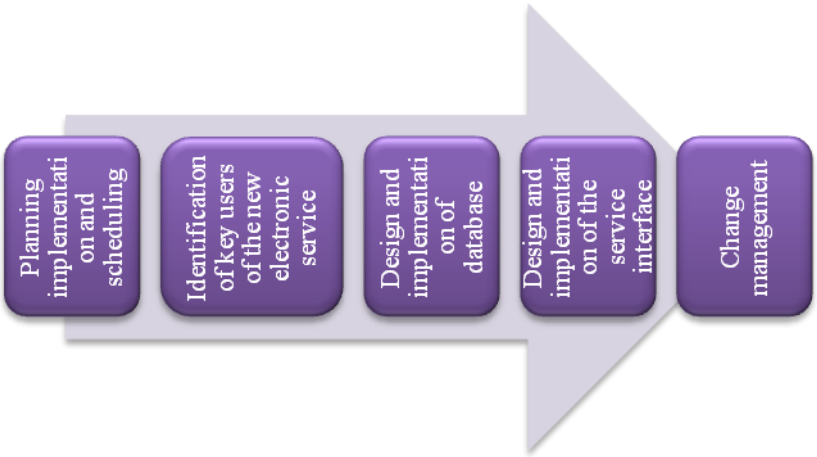



Fig. 2 Milestones in the implementation of new electronic services

**Design and implementation of the database**. Here, the focus is on the design of the tables in the relational structure and the relationships between them. Then the focus moves on the appropriate environment for the realization of the database and it is built. Choice of media for the realization and the characteristics of the stages in the creation of a database is widely analyzed in the literature [2].

**Design and implementation of the service interface**. This step is the design of variants of channels for access to the database of the institution (Web, e-mail, telephone, etc.). At this stage of the implementation of online services takes place and the programs, a set of modules for management of user requests, help system (Help) and security modules.

**Change management**. This task is linked to the creation of documentation to describe the progress of the project and description of the tasks performed for a specific period. Each task is described by the expected outcome, actions needed to achieve the final result, the planned date for completion of the task and the actual completion date of the task. Naturally, the development of information systems in public administration and in the implementation of any software project may occur changes planned state. To be possible to manage this process needs to draw procedure is used to determine the mechanisms for filing assessment of resource availability, decision making, monitoring of possible changes in the scope of the project and all related activities and results . Procedure for change management in transforming the concept into a new electronic service:

- Sets out the procedural requirements for change control, the ways in which accepts or rejects a proposal, based on an assessment impact on the schedule, budget, scope and objectives of the project;
- Determines the manner in which members of the team developing the project and interested parties may request changes;
- Record information for analysis and evaluation of the project.

Part of the procedure for change management is the documentation of monitoring and controlling changes (Fig. 3).

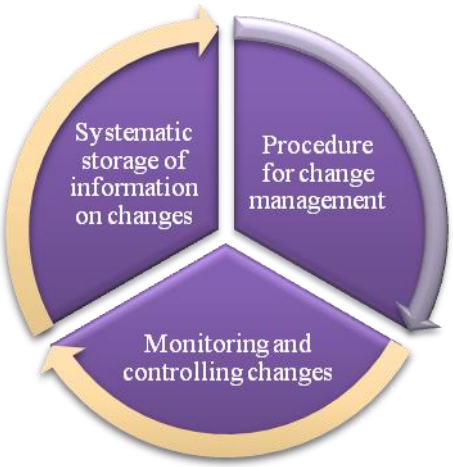



Fig. 3 Procedure for change management

Maintaining this documentation provides and requires a change in the project becomes a particular method of monitoring and controlling changes. Maintaining such documentation purposes:

- To create conditions for pre-identifying changes in the scope or other unplanned activities to monitor their impact;
- To create conditions for the protection of the results that have been approved and signed as appropriate to achieve the objective;
- To ensure that the new tasks and requested changes are economically justified and that the affected targets are identified and modified accordingly;
- To create conditions for continuing with new tasks and their assignment to perform the relevant parties;
- To ensure that the cost and budget of the project.

Changes are subject to acceptance by the developer and sponsor. They should be reflected in the concept of a new electronic service.

## 4.Conclusion

The process of creating online services is one of the toughest tasks facing any public institution. One of the possible application of the proposed approach on the development of new electronic services in the field of local administration in the development and maintenance of electronic services offered by municipalities. In this subject area (local administration), there are specifics that may lead to a modification of the above methodological aspects.

## References

1. Kirilov, R., Information Technologies and Student practices, 3 rd ICAICTSEE, December 6-7 th, 2013, University of National and World Ecomony, Sofia, Bulgaria, ISBN 978-954-644-586-5.
2. Milev, P., "Technological issues of implementing software for media monitoring", "Trakia Journal of Sciences", Vol. 11, Suppl. 1, 2013, Trakia University, ISSN 1313-7069, 2013.
3. http://www.nsi.bg

# Information Problems at Verification of Public Projects

**Rosen Kirilov**

UNWE, 1700 Sofia, Bulgaria
rosenkirilov@mail.bg

**Abstract.** The aim of this paper is to provide basic information about the problems in the verification of public projects. They are displayed based on the experience of the project "Student practices". Characteristic of this project is the electronic verification of expenditure incurred by universities. The project can be used as a good basis for improvement of the platforms in this direction. Finally, conclusions and guidelines for the future development of the software are drawn.

**Keywords.** Information technologies, project, electronic verification of expenditures.

## 1.Introduction

Project BG051PO001 - 3.3.07-0002 "Students practice" [1] focuses on practical training of students in real working environment. Its ultimate objective is to increase the percentage of students employed on a permanent contract. For the period 06/2013 - 12/2014 University of National and World Economy [2] has concluded the contract of 4200 students for practical training and verified their costs electronically.

## 2.Electronic verification

Electronic verification of expenditure on public projects is based on the following principles (Fig. 1):

- Full compliance of expenditure with existing regulations;
- Complete line between tax and expenditure;
- Correspondence between underlying and reported costs;
- Correspondence between electronic and scanned paper documents.

The main roles of the participants in the process of verification are (Fig. 2):

- Expert financial - accounting services;
- Expert in scanning and uploading system scanned financial documents;
- Verifier.

In praktiki.mon.bg system is attached approach for double verification of costs incurred by two independent bodies (Fig. 3).

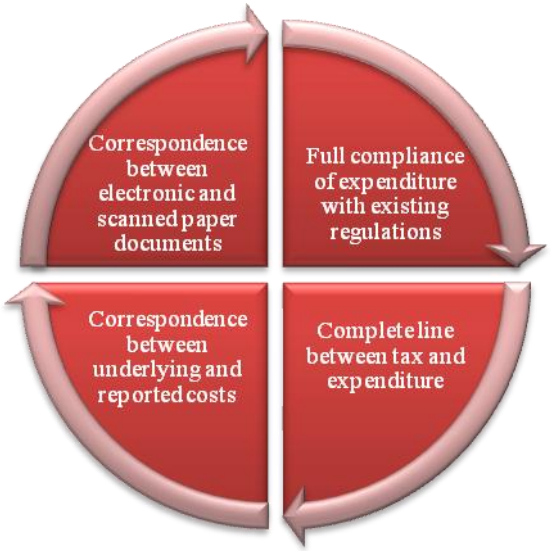



Fig. 1 Basic principles of the electronic verification

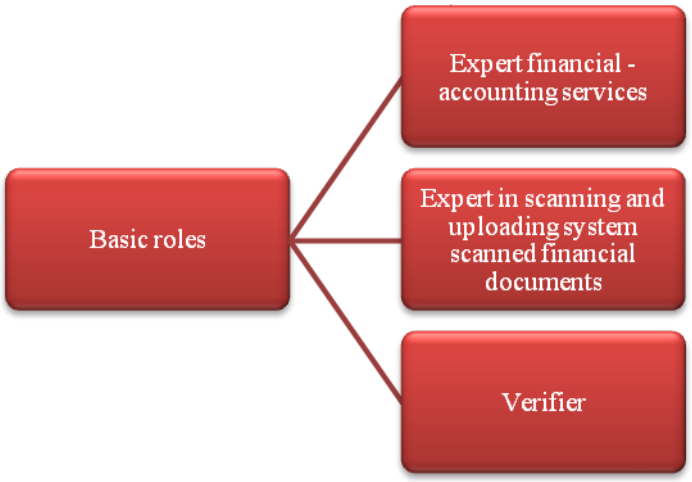



Fig. 2 Roles of participants


Fig. 3 Verifiers

## 3.Information problems

A result of the analysis may state the following information problems in the electronic verification of expenditure on public projects (Fig. 4):

- Maintaining data integrity in the system;
- Presence of management procedure changes;
- Presence of management procedure versions of documents;
- Electronic notification of changes.

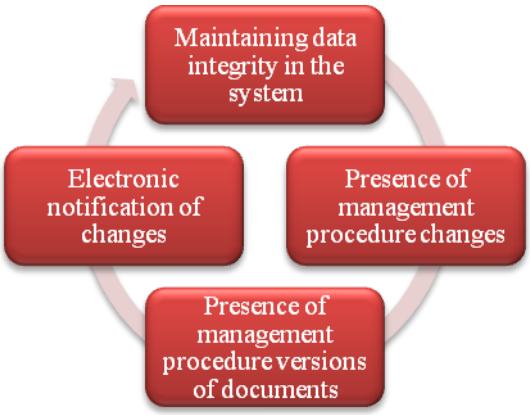



Fig. 4 Information problems

## 4.Interface on system for electronic verification

Fig. 5 and Fig. 6 presents the main screen of the electronic verification of expenditure.

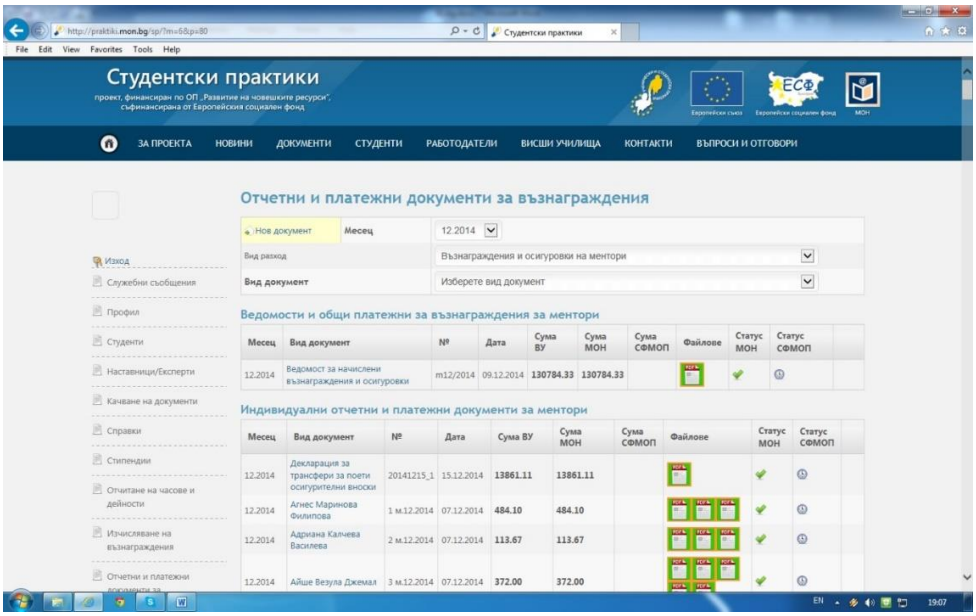



Fig. 5 User interface – 1

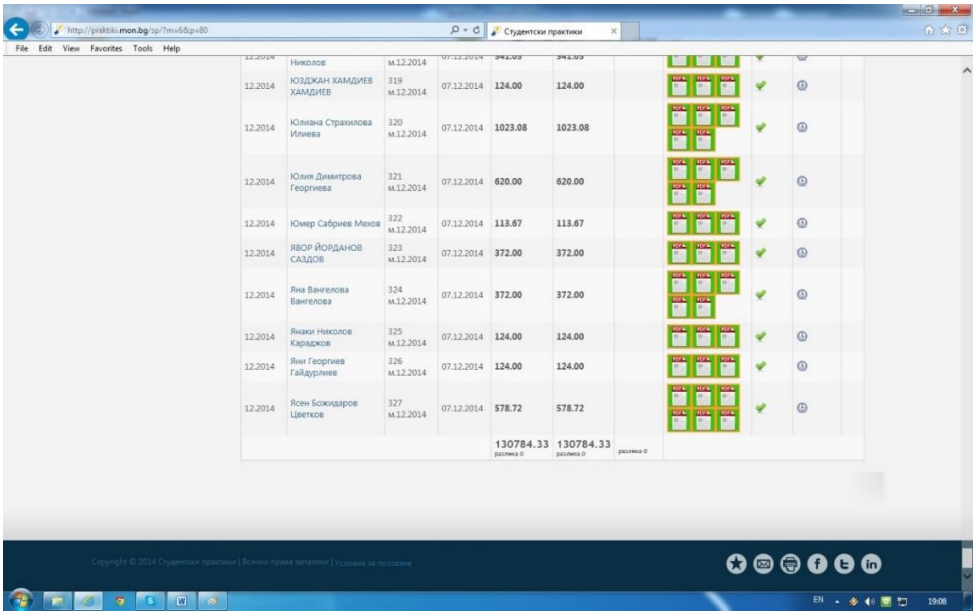



Fig. 6 User interface – 2

## 5.Conclusion

Finally, it should be noted that the application of this approach is relatively new. As such it has certain disadvantages. Such difficulties are mainly corrections of documents and payments. They are the basis for future development of electronic verification of expenditure. Prospects in the future are the majority of public projects to be verified in a similar way.

## References

1. http://praktiki.mon.bg
2. http://www.unwe.bg

# Information Technologies for Career Development of University Staff at UNWE

**Rosen Kirilov**

UNWE, 1700 Sofia, Bulgaria
rosenkirilov@mail.bg

**Abstract.** The purpose of this paper is to present the structure and basic functionality of the system implemented for the development of academic staff in the University of national and world economy. The system is available at web address kariera.unwe.bg. The paper consistently introduces the concept of the system, its modules and work. Finally, conclusions and guidelines are drawn for the future development of the presented software.

**Keywords.** Information technologies, software, career development

## 1.Introduction

The career development of academic staff in the University of national and world economy is based on current legislation. It includes a set of laws, regulations and requirements defining academic growth. The system reflects the wishes of the computerization process of maintaining data publishing activities of the academic staff and their inclusion in the so called papers for assessing the quantitative requirements.

## 2.Concept of the system

At the University of national and world economy there are developed regulations on the implementation of a system of quantitative and qualitative requirements. This system is applied in the evaluation of candidates for certain academic positions. Quantitative requirements for individual positions are summarized in the so called report cards. When someone is applying for an academic position, the readiness of candidates is verified on the basis of these papers by committee on quantitative requirements (Fig. 1).

## 3.Main processes

In order to clarify the actions that users perform, here are some of the main processes [1]:

- Registration of users in the system. The registration of users in the system is one of the first processes. It is connected with the authorization of users to work with individual modules;
- Completion of data for teachers. Once users are registered after completing their basic data;
- Defining a paper for the development of teaching staff. Defining a paper for the development of the teaching staff can be part of the policy of the management of each university;
- Detailed development of the components of the paper for the development of teaching staff. Once defined paper for the development of the teaching staff it should proceed to the detailed design of the components of the paper. Here is essentially performed and the association of individual components of the paper to

the regulatory requirements for the development of teaching staff;

- Evaluation of the performance requirements of the paper for the development of teaching staff. One of the key moments for the system is the assessment of the requirements for the development of teaching staff. At its essence is a starting point for the transition from an academic position to another.

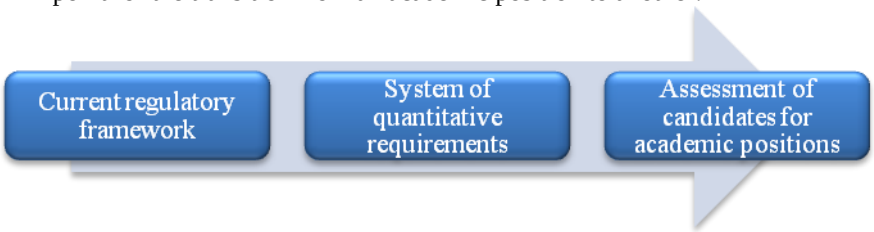



Fig. 1 Assessment of candidates for academic positions

## 4.User interfaces

It is important to indicate the positions of the main menu of the system, namely (Fig. 2):


Fig. 2 Main menu of the system

- Educational activities;
- Research;
- Personal activities;
- Papers for development;
- Plan;
- Reports;
- Exit.

One of the largest items in the menu system is "Research". It is most interesting in terms of data collection for publishing activities of the academic staff (Fig. 3).

Also interesting is the "Papers for development" part of the system. It generates papers for assessment of the quantitative requirements of a candidate (Fig. 4).


Fig. 3 Research


Fig. 4 Papers for development

## 5.Conclusion

Finally, it should be noted that the implemented system reflects on current regulations in the University of national and world economy. It provides opportunities to unify and standardize the collection of information for publication activity of the teaching staff. On this basis, the system can provide a variety of information and facilitate decision making information.

## References

1. http://kariera.unwe.bg

# Technological Approaches for Researching Content of Web Forums

**Plamen Milev**

UNWE, 1700 Sofia, Bulgaria
pmilev@unwe.bg

**Abstract.** The paper presents some technological features of crawling pages with topics of web forums. The purpose of crawling is an examination of the content of the forums. The navigation through the web pages is realized by the so called web spider and this process is modeled and optimized based on the specifics characteristic of this group of sources of information. The paper presents a method for such an automated crawling of pages of web forum and in conclusion its advantages are highlighted.

**Keywords.** Internet, forum, web crawler.

## 1.Introduction

The automated crawling of pages with topics of web forum has been developed through several interrelated processes. First, we launch a process to investigate pages with themes. This is a process that refers to a given web source. Then, we initialize the program module for crawling pages with themes, which is actually the search engine. We load the scheme with instructions for the given web source. Based on this scheme, we load the initial web address from where we should start crawling. Finally, we perform HTTP request to this address. If error occurs due to the poor response of the server, this error will be reported and we finalize the process. If the HTTP server response is correct, we analyzes the resulting HTML. This code searches for a list component described in the scheme with instructions of the forum. If we do not recognize a list component, we finalize the process of recording data error. On detection of a list component, we proceed to extract the addresses of all categories that should be crawled.

## 2.Automated crawling of web forum pages

Before extraction of all web addresses of categories that should be crawled, we record errors that have occurred in the process of searching for these addresses. Then, we go on with processing each of the extracted web addresses. We process only those iterations of addresses that have not yet been processed. When the number of unvisited addresses are exhausted, the process is finalized. When the address is not visited so far, we make HTTP request to it. If some problem in the remote server occurs, we report the error and finalize the process. In the case with correct response from the remote server, we go on and analyze the returned HTML. We search for list components. In the absence of such components, we finalize the process. In the presence of a list component, we extract all web addresses with topics that should be crawled and we report the errors that have occurred in the process of searching for these addresses. Thus, the loop described in Fig. 1 is repeated until all required categories of web addresses are crawled.

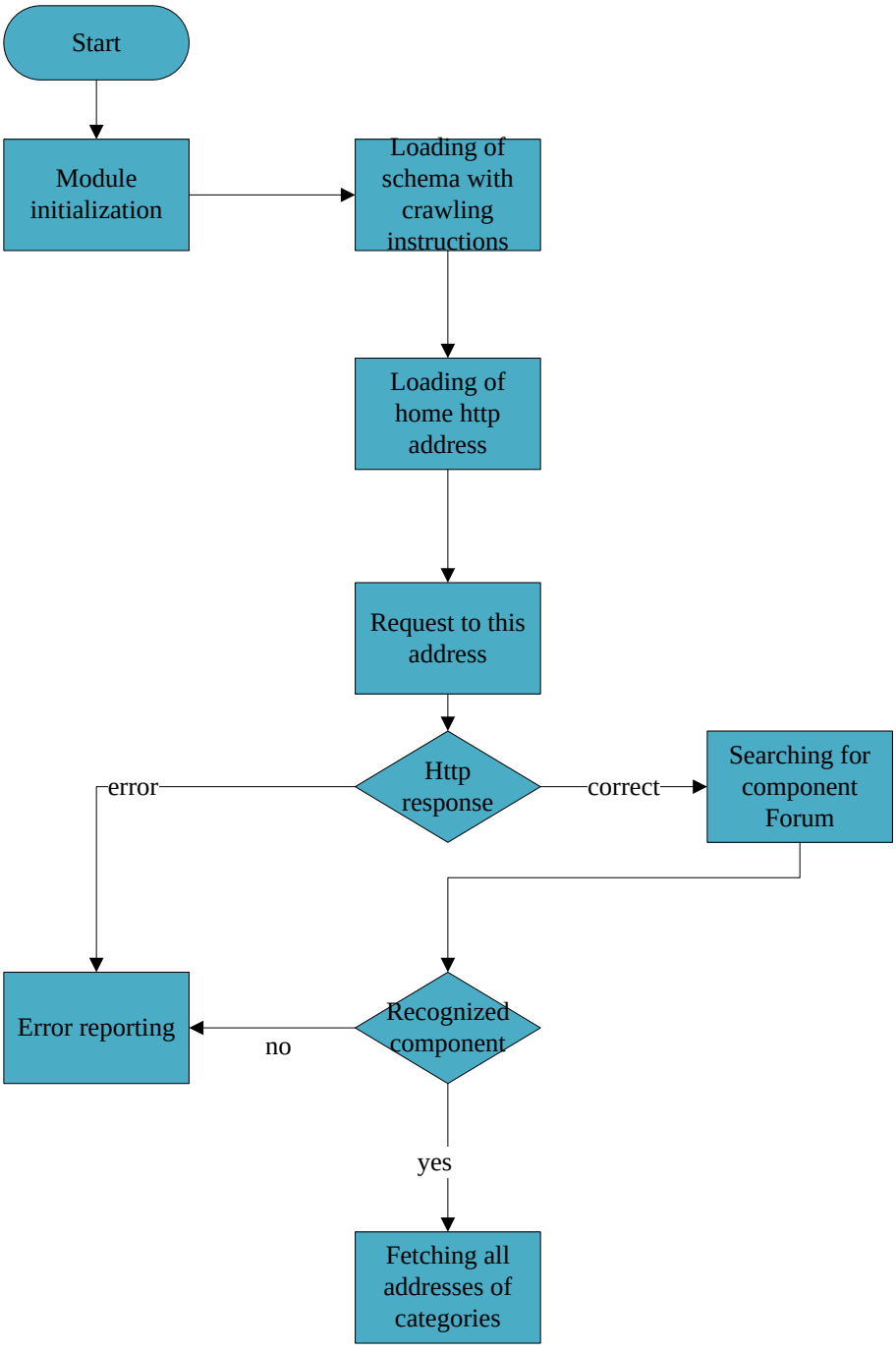



Fig. 1. Crawling of web forum pages with categories

Fig. 2 illustrates the process of paging within the method for automated crawling of pages with topics of web forum.

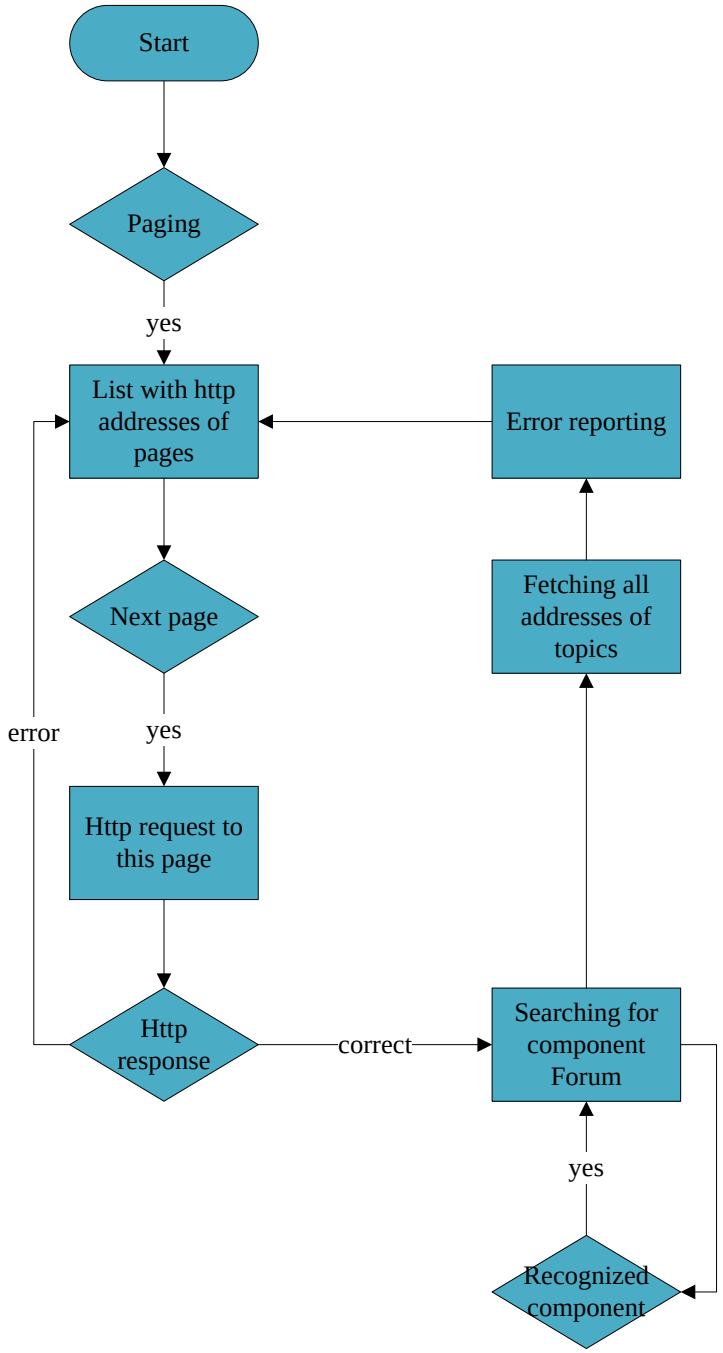



Fig. 2. Crawling of web forum pages with topics

We search for paging component. In the absence of such a component the process is finalized. If there is a component for paging, we attempt to extract web address of the next

page to be crawled. If there is no such an address, the process is finalized. If there is such an address, we execute HTTP request to it. If some problem in the remote server occurs, an error is reported. In a correct answer, we analyze the resulting HTML. In the absence of a list component, an error is reported. In the presence of such a list, we extract the data. The whole cycle continues until the point where there is no next web page address and we report the errors that have occurred in the process of searching for these addresses. We go on with consistent work with each of the extracted web addresses. We process only those iterations of addresses that have not yet been requested. When the number of unvisited addresses are exhausted, the process is finalized. When the address is not visited, we execute HTTP request to it. If some problem in the remote server occurs, the error is reported and the process is finalized. In a correct response to the remote server, we analyze the returned HTML. We search for list components. In the absence of such component, the process is finalized. In the presence of a list component, we extract all web addresses with topics that should be crawled and we report the errors that have occurred in the process of searching for these addresses. The cycle described in Fig. 2 is repeated until all required web addresses with topics are crawled.

## 3.Conclusion

In conclusion it can be said that the method for automated crawling of pages with topics of web forum presents an algorithm to extract all the lists of topics available in pages of categories of forums. Implemented in this way, the method has the following technological advantages:

- Optimized using of limited network resources due to crawl only unvisited pages of web forum;
- Ability to detect components of paging and to retrieve lists topics from all pages on the web forum;
- Uniform analyzing of HTML of a web forum for the determination of pre-defined component.

# Advantages of Using Server-side Applications for Implementing Media Monitoring Solutions

**Plamen Milev**

UNWE, 1700 Sofia, Bulgaria
pmilev@unwe.bg

**Abstract.** The paper discusses issues related to the functionality of software for media monitoring. The main objective of this paper is to present methodological solutions of implementation of such software. Leading part in the paper takes issue with keyword search in the information flow and streamlining the search results. On this basis appropriate technological solutions are proposed. The main results of the work of the software solution show feasibility of the proposed methodological approach. The conclusion outlines some trends in development of media monitoring systems in the context of the critical role of modern technological capabilities.

**Keywords.** Server-side application, media monitoring.

## 1.Introduction

Media monitoring is practically achieved by a combination of technologies including audio and video recording, high speed text scanners and text recognition software and human readers and analysts. The automation of the process is highly desirable and can be partially achieved by deploying data mining and machine learning techniques. The paper discusses technological issues of implementing software for media monitoring of online information sources. Understandings in the paper about the nature of this type of media monitoring are illustrated in Fig. 1. According to this paper a media monitoring software should implement the following options:

- Maintaining a large number of online information sources (all known if it is possible);
- Ability to automatically search in publications of these information sources according to our interests (interests of our clients);
- Creating a newsletter with wanted publications found according to established criteria.

## 2.Architecture of server-side applications for implementing media monitoring solutions

According to this paper a complete architecture of software for media monitoring of online information sources should contain the following components (Fig. 2):

- Web clients, which check for new publications respectively in web sites groups (different groups of web sites, arranged according to certain criteria). The use of several web server aims to provide a resource for media monitoring software when working with a large number of information sources;
- Application server, which manages analyze and retrieval of web content. This application server also organizes the work of all available web servers and synchronizes their actions;

- Database server, which stores the downloaded from information sources content that is going to be indexed by media monitoring software;
- Application server, which indexes the downloaded from information sources content. At this level are implemented algorithms for indexing that allows optimal information search from the client part of media monitoring software;
- Database server, which stores all the indexed publications. Client search operations are performed by that server;
- Web server of the client part of the system that manages the content that is displayed on the web site;
- Application server, which is an intermediate level between the client website and the database with indexed publications. This application server provides search by key words in publications indexed.


Fig. 1 Understanding of media monitoring

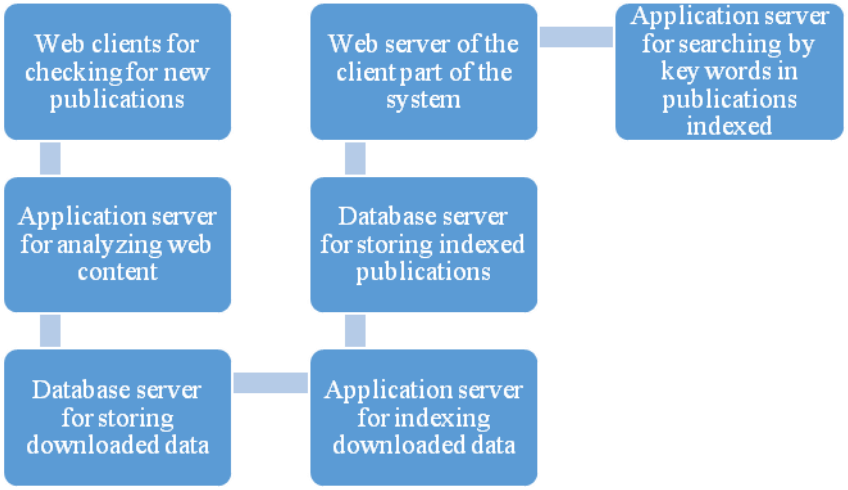



Fig. 2 Software architecture

For the purposes of our study, the retrieval of the information must be appropriately combined with a suitable analyze, such as to enable the presence of the extracted contents in a structured way. There are many different information sources with their distinctive designs and ways of presenting information, but Fig. 3 illustrates the general scheme of a media website that consists of a home page and subpages with different elements. Information unit in this case is the publication with its details. These details are usually the author, title and text of the publication, the date it was published on the website, photo or galleries of photos, etc. The aim of the module for analysis and retrieval of web content consists mainly to

recognize individual elements of publications, so that subsequently they may be indexed and processed separately. The algorithm for analyze and retrieval of web content can be described in the following steps:

- Request to the home page of the information source with publications;
- Determination of the menu with links to categories of publications;
- Defining a menu with links to subcategories, if any;
- Determination of the page area with a list of links to publications;
- Request to detailed page of a publication in the list;
- Determination of the base area of the publication within the page;
- Determination of the area of each of the essential elements of the publication;
- Retrieval of individual contents in the form of HTML code;
- Redefinition of the rules for the contents of this information source.

This model does not know in advance the addresses (links) of the pages with publications. For this reason, the initial starting point is always the home page of the web site. The next item that interests us is the location of the menu with the categories of publications. Categorization itself is not a focus of this paper. What is important for us is the proper distinction between information structures. Following the steps of the presented algorithm lead us to the extraction of web content that forms one publication. Ways in which these contents could be represented in HTML form are very indefinite. For this reason we actually need a specific approach. Our system should implement a basic functionality that is common to all the detailed pages of publications. This implementation, however, provides methods for setup of specific web content for each component separately extracted. The use of these methods is not required. In many cases, the base implementation should be sufficient for a proper interpretation of the HTML code to the desired structure. Additional functionality will be used when there is a need for correcting errors in incorrectly structured HTML, clearing up some of the code tags, removing unnecessary fragments extracted from HTML (as some banners), ignoring JavaScript. The presented approach can be defined as an object-oriented, because of the potential to redefine the way of extracting the resulting text from web content specific to object-oriented platforms that use the techniques of inheritance.

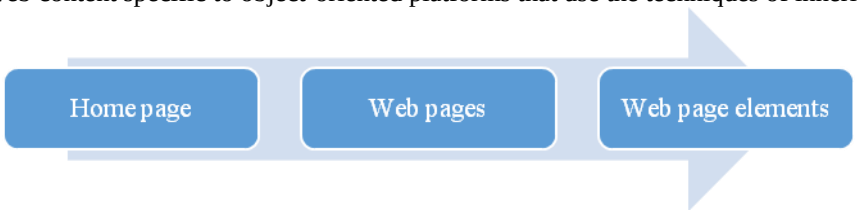



Fig. 3. Structure of web site with wanted publications

## 3.Conclusion

The model for analyze and retrieval of web content presented in the paper can be characterized by the following potential advantages:

- Ability to extract structured information from different sources in a defined general way;
- Programming of rules for behavior in analyze of web content that is used for various web design structures;
- Methods for redefining certain areas of the web pages for specific treatment of HTML fragments.

# Relevance of Data Warehouses in Municipalities in Bulgaria

**Ina Lyubenova**

UNWE, 1700 Sofia, Bulgaria
Department 'Information Technologies and Communications',
ina.lyubenova@gmail.com

**Abstract.** The report examines the data warehouses in the municipalities of Bulgaria and how their relevance influence the services that they offer. It has been shown the importance of properly designed data warehouses and the need of methodology to follow, when building a data warehouse for a municipality. Conclusions are made and guidelines are identified for improving municipal processes using business intelligence systems.

**Keywords.** Data warehouse, business intelligence systems, local administration, municipality

A data warehouse is a system used for reporting and data analysis. Integrating data from one or more disparate sources creates a central repository of data, a data warehouse (DW). Data warehouses store current and historical data and are used for creating trending reports for senior management reporting such as annual and quarterly comparisons.[2]

The data stored in the warehouse is uploaded from the operational systems (such as marketing, sales, etc., shown in the figure to the right). The data may pass through an operational data store (a staging area) for additional operations before it is used in the DW for reporting.[2]

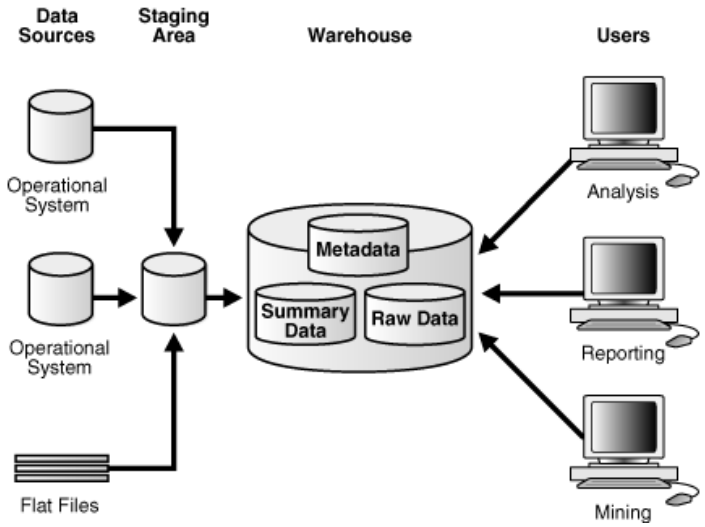



Figure.1. A Data Warehouse with data sources, staging area and users.

In Figure 1, the metadata and raw data of a traditional OLTP system is present, as is an additional type of data, summary data. Summaries are very valuable in data warehouses because they pre-compute long operations in advance, which could be extracted very quick and easy later using queries to the Data Warehouse.[3]

Often BI applications use data gathered from a data warehouse (DW) or from a data mart. A data warehouse contains a copy of analytical data that facilitates decision support. However, not all data warehouses serve for business intelligence, nor do all business intelligence applications require a data warehouse.[1]

But after all their connection is very strong and the development of a BIS, almost certainly means that a Data Warehouse is also developed.

In order to determine the features of the Data Warehouses in municipalities in Bulgaria, we are going to use the results from a research made in the municipalities in Bulgaria. For the determination of their features a questionnaire is designed and an empirical research is conducted. The study was conducted on a representative sample of 54 municipalities in Bulgaria. Figure 2 shows the proportion of municipalities in the sample to all municipalities in Bulgaria. Studied municipalities accounted for approximately 22% of the population. According to many statistical sources, samples with a volume of more than 30 units are defined as larger samples and their results should be considered reliable in terms of the population of all 264 municipalities in Bulgaria.

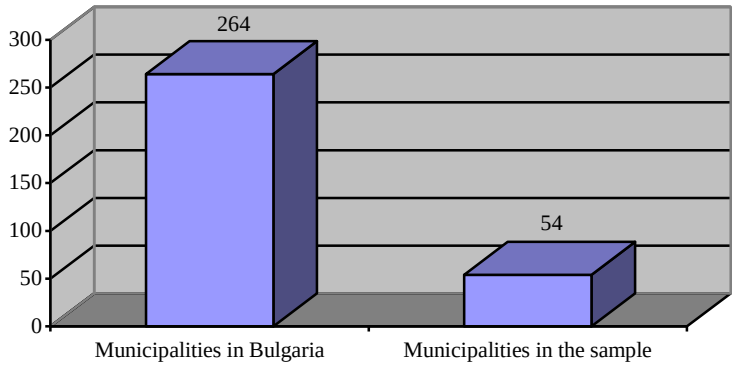



Figure 2. Municipalities in the sample compare to all the municipalities in Bulgaria

All the municipalities in the sample of the present study were selected randomly through a lottery selection and a function that generates random numbers. For the purposes of the study a stratified sample was used, and therefore the sample consists of municipalities from all 28 regions of Bulgaria. In this sense, referring to an administrative district in the Republic of Bulgaria, the possibility of each municipality to be in the sample is equal compared to other municipalities in this district. The way in which the sample is formed, gives us reason to believe the data obtained from the questionnaires is representative and we could treat the statistical results from the sample to the entire population of the municipalities in Bulgaria. Table 1 shows a list of the random chosen municipalities in the sample with their population according to the National Statistical Institute (NSI) 31.12.2010. The total number of residents in the surveyed municipalities is 3,418,906, which is almost 50% of the total population of the Republic Bulgaria.

Table 1. Municipalities in the sample and their population, NSI (www.nsi.bg)[4]

| DISTRICT | MUNICIPALITY | POPULATION |
|---|---|---|
| Blagoevgrad | Blagoevgrad | 76812 |
| | Petrich | 55408 |
| Burgas | Burag | 206700 |
| | Tzarevo | 9411 |
| Varna | Aksakovo | 21972 |
| | Dolni Chiflik | 19141 |
| | Dalgopol | 14204 |
| Veliko Tarnovo | Gorna Oriahovitza | 48049 |
| | Zlataritza | 4584 |
| | Svishtov | 48351 |
| Vidin | Vidin | 64989 |
| Vratza | Vratza | 74648 |
| | Oriahovo | 12069 |
| Gabrovo | Gabrovo | 66321 |
| | Drianovo | 10315 |
| | Triavna | 12094 |
| Dobrich | Kavarna | 15657 |
| | Krushari | 5118 |
| Kardgali | Ardino | 12282 |
| | Kardgali | 68406 |
| Kiustendil | Kocherinovo | 5811 |
| | Kiustendil | 61944 |
| Lovech | Lovech | 52308 |
| | Yablanitza | 6334 |
| Montana | Berkovitza | 19256 |
| Pazardgik | Pazardgik | 120422 |
| | Rakitovo | 15418 |
| Pernik | Breznik | 7506 |
| Pleven | Belene | 10671 |
| | Dolni Dabnik | 14230 |
| Plovdiv | Karlovo | 53656 |
| | Plovdiv | 347611 |
| | Hisaria | 12768 |
| Razgrad | Razgrad | 53918 |
| | Zavet | 11110 |
| Ruse | Borovo | 6511 |
| Silistra | Dulovo | 28634 |

| | | |
|---|---|---|
| | Tutrakan | 16637 |
| Sliven | Nova Zagora | 41199 |
| | Tvarditza | 14180 |
| Smolian | Devin | 12963 |
| | Zlatograd | 12112 |
| Sofia (capital) | Stolichna | 1259446 |
| Sofia | Botevgrad | 33529 |
| | Kostinbrod | 17448 |
| Stara Zagora | Bratia Daskalovi | 9625 |
| | Kazanlak | 75509 |
| Targovishte | Antonovo | 6395 |
| | Targovishte | 59855 |
| Haskovo | Svilengrad | 23867 |
| | Dimitrovgrad | 55869 |
| Shumen | Varbitza | 10318 |
| Yambol | Elhovo | 16332 |
| | Yambol | 75742 |
| Total | | **3418906** |

The questionnaire was divided into several main sections, each of which consists of a different number of questions. Total number of questions in the questionnaire was 50. All questions in the survey are intended to clarify the current level of use of information technology in local government. Sections in the questionnaire are:

1. Employees in Information Technology (9 questions);
2. Hardware (7 questions);
3. Software (7 questions);
4. Databases (14 questions);
5. Website, online services (7 questions);
6. Business intelligent systems, early warning systems (6 questions);

We are going to look at the one dimensional frequency distribution of Question 6.1. Does your municipality have a BIS? The answers are presented in Figure 3. Using a one dimensional frequency distribution of quality features we can only see the frequency distribution of the various meanings of the features. In this case we can see that 7 municipalities from the survey have a BIS, which is 13% of the surveyed municipalities, while 47 municipalities don't have a BIS, which is 87% of the surveyed municipalities.

The results from the survey show that very few municipalities in Bulgaria have a BIS.

As the trends are BIS to increasingly step into the organizations, public and private, we need to figure out how to make that transfer easier for our local municipalities.

In the light of this matter we are going to look at the results from another question form the survey that took place among the municipalities in Bulgaria.

One dimensional frequency distribution of Question 4.1.Does your municipality have a Database built for its activities? is presented in Figure 4. Using a one dimensional frequency distribution of quality features we can only see the frequency distribution of the various meanings of the features. In this case we can see that 33 municipalities from the survey don't have a Database built for their activities, which is 61% of the surveyed

municipalities, while 21 municipalities have a Database built for their activities, which is 39% of the surveyed municipalities.

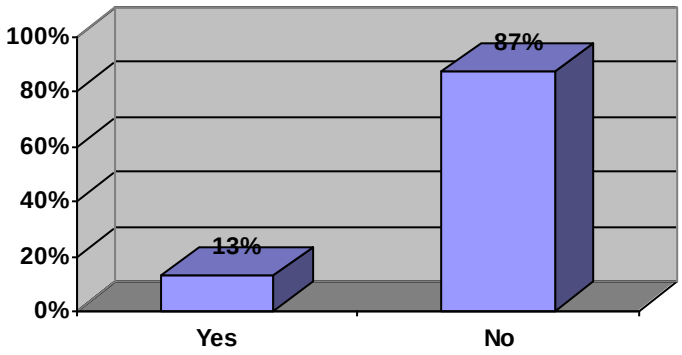



Figure 3.One dimensional frequency distribution of Question 6.1.Does your municipality have a BIS?

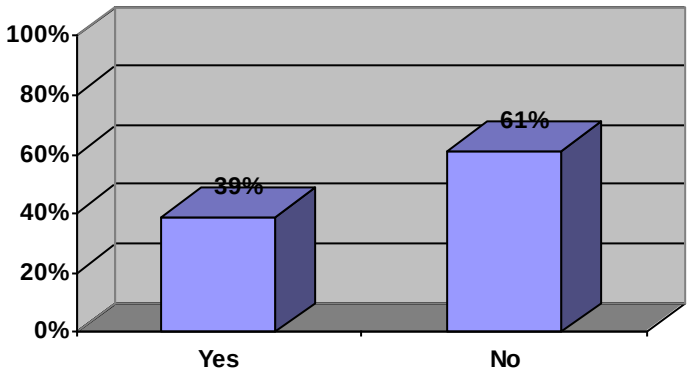



Figure 4. One dimensional frequency distribution of Question 4.1.Does your municipality have a Database built for its activities?

As we can see a significant part of the municipalities in Bulgaria already have Databases built. So they could be used as a base for building Data Warehouses, which are an essential part of building BIS.

## References

1. http://en.wikipedia.org/wiki/Business_intelligence
2. http://en.wikipedia.org/wiki/Data_warehouse
3. Oracle 9*i* Data Warehousing Guide Release 2 (9.2), Available at: http://docs.oracle.com/cd/B10500_01/server.920/a96520/concept.htm
4. www.nsi.bg

# Data Warehouses as a Factor for a Proper Functioning of the Local Administration

**Ina Lyubenova[1], Rosen Kirilov[2]**

[1]UNWE, 1700 Sofia, Bulgaria
Department 'Information Technologies and Communications',
ina.lyubenova@gmail.com

[2]UNWE, 1700 Sofia, Bulgaria
Department 'Information Technologies and Communications',
rosenkirilov@mail.bg

**Abstract.** The report analyzes the data warehouses in the municipalities of Bulgaria. The functions that data warehouses perform are presented and also how they reflect on the proper management and functioning of the local administration. In the conclusion there are suggestions for more efficient management of the data warehouses that allows the municipalities to function in a better environment.

**Keywords.** Data warehouse, business intelligence systems, local administration, municipality

## 1.Introduction

As we saw in our previous report a significant part of the municipalities in Bulgaria already have Databases built. So they could be used as a base for building Data Warehouses, which are an essential part of building BIS. Now we are going to examine the features of the Data Warehouses that will help in the proper functioning of the local municipalities.

## 2.Understanding Data Warehouse

In order to understand the Data Warehouses we need to know that

- The Data Warehouse is that database which is kept separate from the organization's operational database.
- There is no frequent updation done in data warehouse.
- Data warehouse possess consolidated historical data which help the organization to analyse it's business.
- Data warehouse helps the executives to organize,understand and use their data to take strategic decision.
- Data warehouse systems available which helps in integration of diversity of application systems.
- The Data warehouse system allows analysis of consolidated historical data analysis.

## 3.Data Warehousing

Data Warehousing is the process of constructing and using the data warehouse. The data warehouse is constructed by integrating the data from multiple heterogeneous sources. This data warehouse supports analytical reporting, structured and/or ad hoc queries and decision making. Data Warehousing involves data cleaning, data integration and data consolidations.

## 4.The Environment for Data Warehousing

The environment for data warehouses and marts includes the following:

- Source systems that provide data to the warehouse or mart;
- Data integration technology and processes that are needed to prepare the data for use;
- Different architectures for storing data in an organization's data warehouse or data marts;
- Different tools and applications for the variety of users;
- Metadata, data quality, and governance processes must be in place to ensure that the warehouse or mart meets its purposes.[1]

Today, the most successful companies are those that can respond quickly and flexibly to market changes and opportunities. A key to this response is the effective and efficient use of data and information by analysts and managers. A “data warehouse” is a repository of historical data that are organized by subject to support decision makers in the organization. Once data are stored in a data mart or warehouse, they can be accessed.[1]

## 5.Using Data Warehouse Information

There are decision support technologies available which help to utilize the data warehouse. These technologies helps the executives to use the warehouse quickly and effectively. They can gather the data, analyse it and take the decisions based on the information in the warehouse. The information gathered from the warehouse can be used in any of the following domains:

- **Tuning production strategies** - The product strategies can be well tuned by repositioning the products and managing product portfolios by comparing the sales quarterly or yearly.
- **Customer Analysis** - The customer analysis is done by analyzing the customer's buying preferences, buying time, budget cycles etc.
- **Operations Analysis** - Data Warehousing also helps in customer relationship management, making environmental corrections.The Information also allow us to analyse the business operations.

## 6.Data Warehouse Features

Bill Inmon is sometimes also referred to as the "father of data warehousing"; his design methodology is based on a top-down approach and defines data warehouse key features in these terms

- Subject oriented - The data in a data warehouse is categorized on the basis of the subject area and hence it is "subject oriented".
- Integrated - Data gets integrated from different disparate data sources and hence universal naming conventions, measurements, classifications and so on used in the data warehouse. The data warehouse provides an enterprise consolidated view of

data and therefore it is designated as an integrated solution.

- Non-volatile - Once the data is integrated\loaded into the data warehouse it can only be read. Users cannot make changes to the data and this practice makes the data non-volatile.
- Time Variant - Finally data is stored for long periods of time quantified in years and has a date and timestamp and therefore it is described as "time variant".[3]

## 7.Differences between Data Warehouses and Operational Databases

The following are the reasons why Data Warehouse are kept separate from operational databases:

- The operational database is constructed for well known tasks and workload such as searching particular records, indexing etc but the data warehouse queries are often complex and it presents the general form of data.
- Operational databases supports the concurrent processing of multiple transactions.Concurrency control and recovery mechanism are required for operational databases to ensure robustness and consistency of database.
- Operational database query allow to read, modify operations while the OLAP query need only**read only** access of stored data.
- Operational database maintain the current data on the other hand data warehouse maintain the historical data.

The data in Data Warehouses (DW) was obtained from multiple sources and stored in various databases in different formats and structures. As a result, DW is the most important and expensive part of BI because it takes data from these sources and unify it. DW is the central repository of data in a business organization. DW contains a wide range of data that represents a homogeneous picture of business conditions at some point in time. Its aim is to create an infrastructure database that is always online, it contains all the information systems for online transaction processing (OLTP), but is structured so that it is fast and efficient for queries and analysis. Development of DW includes establishing a system for extracting data from the underlying transaction operating systems. DW provides system managers flexible access to data. DW term usually refers to a combination of many different databases throughout the organization. This distinguishes it from Data Marts, which are one or several databases, designed to assist managers in making decisions. Data Marts are smaller and are focused on a particular problem or department, such as marketing, sales, production, etc. [2]

## 8.Conclusion

For the successful building of Data Warehouses in the local municipalities, a proper way need to be found to integrate the sources that are already built, in order to combine them in one common Data Warehouse. This will lead to building more BIS in the local municipalities.

## References

5. http://en.wikipedia.org/wiki/Data_warehouse
6. http://fbm.uni-ruse.bg/d/im/White_Paper_2007_Bg.pdf
7. http://www.mssqltips.com/sqlservertip/2976/comparing-data-warehouse-design-methodologies-for-microsoft-sql-server/

# Using Information from Extended Events in Microsoft SQL Server for Identifying Problematic Queries

**Mitko Radoev**

Department of Information Technologies and Communications, UNWE, Sofia, Bulgaria

**Abstract.** The query execution is a central issue in any Database Management System. The aim is to perform the queries as fast as possible. There are many different resources and tools to improve the performance of the queries. Improvement of the performance, however, should be targeted mainly at the problematic queries. In recent versions of Microsoft SQL Server a new tool was added for gathering information - Extended Events. This study aims to present the capacity of this tool and to propose practical approaches to identify the problematic queries, using collected information.

**Keywords:** problematic queries, query performance, Extended Events

## 1.Introduction

SQL Server Extended Events is a general event-handling system. It was first introduced in SQL Server 2008. Extended Events is a tool that can be configured to monitor and capture different types of data, including performance data.

This is not the only source of data for query execution. In Microsoft SQL Server there are multiple execution-related Dynamic Management Objects /DMO/ such as:

- sys.dm_exec_query_stats,
- sys.dm_exec_sql_text,
- sys.dm_exec_query_plan

and many others views and functions.

Another possible source is Data Collector. Query Statistics data collection actually contains data from DMOs.

In many cases the information obtained from the DMOs or Data Collector is not enough to diagnose the problem. Then the solution is to use Extended Events.

### 1.1 Goals

The main goals of this study are to present the capacity of Microsoft SQL Server Extended Events and to propose practical approaches to identify the problematic queries, using collected information.

The following tasks are being performed in order to achieve these goals:

- Introducing the main concepts of Extended Events;
- Presenting the tools, related to Extended Events;
- Proposing approaches for identifying problematic queries.

### 1.2 Problematic queries

Problematic queries should be sought in two directions:

- Queries that run for a long time and require a lot of resources - such requests block server and consume a significant part of its resources and thereby worsen the performance of the other queries that are waiting to be executed.

- Queries that do not use too many resources but perform very often.

## 2.Extended Events concepts

Extended Events is a highly scalable and highly configurable tool that allows users to collect as much information as necessary to identify and troubleshoot a performance problem. This system uses very few resources. SQL Server Extended Events uses existing concepts, such as an event or an event consumer and introduces new concepts. [1]

### 2.1 Extended Events objects

Extended Events is realized in packages which are containers of objects like events, targets, and actions. The following figure shows objects and their relationships.

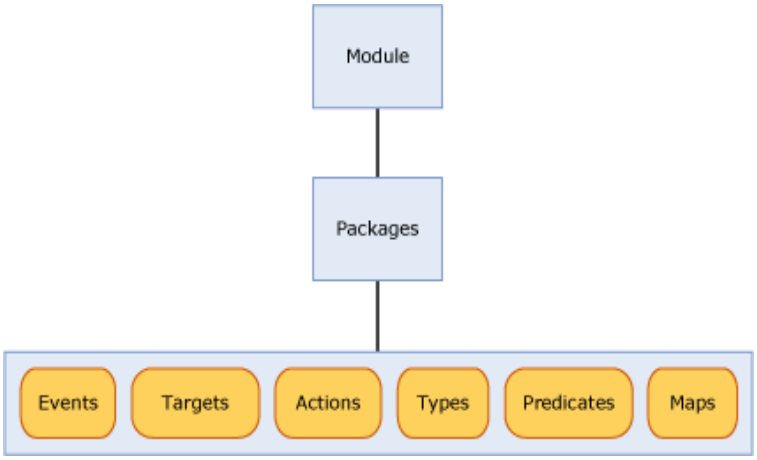



Fig. 2. Extended Events objects [2]

### 2.2 Event types

There are four types of events. They are presented in the following table:

Table 1. Event types

| Event type | Typical events | Used for |
|---|---|---|
| **Admin events** | Events such as error reports | System administration |
| **Analytic events** | Events such as lock acquisition and SQL statements starting and completing | Analysis about system performance and health |
| **Debug events** | Events that fire when threads and processes start and stop | Diagnose and solve engine-related problems |
| **Operational events** | Events that fire when databases being attached, detached, started, and stopped | Managing the SQL Server service and databases |

Analytic events describe program operation and are typically used in performance investigations.

# 3 Extended Events related tools

The following tools can be used for creating and managing SQL Server Extended Events sessions and for obtaining session data and metadata by using Transact-SQL statements:

- Data Definition Language (DDL) statements,
- Dynamic management views,
- Catalog views,
- System tables.

## 3.1 Extended Events related DDL statements

Table 2. Extended Events related DDL statements

| Name | Description |
|---|---|
| **CREATE EVENT SESSION** | Creates an Extended Event session object |
| **ALTER EVENT SESSION** | Starts or stops an event session or changes an event session configuration |
| **DROP EVENT SESSION** | Drops an event session |

## 3.2 Extended Events related Dynamic management views

Table 3. Extended Events related Dynamic management views

| Name | Description |
|---|---|
| **sys.dm_xe_packages** | Lists all the packages registered with the extended events engine |
| **sys.dm_xe_objects** | Returns a row for each object that is exposed by an event package |
| **sys.dm_xe_object_columns** | Returns the schema information for all the objects |
| **sys.dm_xe_map_values** | Returns a mapping of internal numeric keys to human-readable text |
| **sys.dm_xe_sessions** | Returns information about an active extended events session |
| **sys.dm_xe_session_targets** | Returns information about session targets |
| **sys.dm_xe_session_events** | Returns information about session events |
| **sys.dm_xe_session_event_actions** | Returns information about event session actions |
| **sys.dm_xe_session_object_columns** | Shows the configuration values for objects that are bound to a session |

### 3.3 Extended Events related Catalog views

Table 4. Extended Events related Catalog views

| Name | Description |
|---|---|
| **sys.server_event_sessions** | Lists all the event session definitions |
| **sys.server_event_session_events** | Returns a row for each event in an event session |
| **sys.server_event_session_actions** | Returns a row for each action on each event of an event session |
| **sys.server_event_session_targets** | Returns a row for each event target for an event session |
| **sys.server_event_session_fields** | Returns a row for each customizable column that was explicitly set on events and targets |

### 3.4 Extended Events related System tables

Table 5. Extended Events related System tables

| Name | Description |
|---|---|
| **trace_xe_event_map** | Contains one row for each Extended Events event that is mapped to a SQL Trace event class |
| **trace_xe_action_map** | Contains one row for each Extended Events action that is mapped to a SQL Trace column ID |

## 4.Identifying problematic queries

When identifying the problematic queries our efforts are focused on three groups of queries:

- Queries that run for a long time,
- Queries that require a lot of resources,
- Queries that perform very often.

When using Extended Events, collected data can be used for identifying the problematic queries. Events data is collected to a file. The data can be loaded into a table or query it directly, using this system function:

**sys.fn_xe_file_target_read_file** [3]

Another possible approach is to query the data from the following Extended Events related dynamic management views:

**sys.dm_xe_sessions**

**sys.dm_xe_session_targets st** [4]

### 4.1 Identifying slow-running queries

Discovering the slow-running queries is crucial because it has the biggest impact on the end-users opinion about the information system.

One can identify the slow running queries by analyzing a session output file and sort the trace output on the duration data column. An example is presented in Figure 2:

```
WITH xEvents AS
(SELECT object_name AS xEventName,
CAST (event_data AS xml) AS xEventData
FROM sys.fn_xe_file_target_read_file
('C:\EventsData\QueryTest.xel', NULL, NULL, NULL))
SELECT
xEventName,
xEventData.value('(/event/data[@name=''duration'']/value)[1]','bigint') Duration,
xEventData.value('(/event/data[@name=''physical_reads'']/value)[1]','bigint') PhysicalReads,
xEventData.value('(/event/data[@name=''logical_reads'']/value)[1]','bigint') LogicalReads,
xEventData.value('(/event/data[@name=''cpu_time'']/value)[1]','bigint') CpuTime,
xEventData.value('(/event/data[@name=''batch_text'']/value)[1]','varchar(max)') BatchText,
xEventData.value('(/event/data[@name=''statement'']/value)[1]','varchar(max)') StatementText,
xEventData.value('(/event/data[@name=''query_plan_hash'']/value)[1]','binary(8)') QueryPlanHash
FROM xEvents
ORDER BY Duration DESC;
```

Fig. 3. Identifying slow-running queries

### 4.2 Identifying queries that perform a large number of physical reads

To identify queries that require a lot of resources, one can just change the sort order.

```
WITH xEvents AS
(SELECT object_name AS xEventName,
CAST (event_data AS xml) AS xEventData
FROM sys.fn_xe_file_target_read_file
('C:\EventsData\QueryTest.xel', NULL, NULL, NULL))
SELECT
xEventName,
xEventData.value('(/event/data[@name=''duration'']/value)[1]','bigint') Duration,
xEventData.value('(/event/data[@name=''physical_reads'']/value)[1]','bigint') PhysicalReads,
xEventData.value('(/event/data[@name=''logical_reads'']/value)[1]','bigint') LogicalReads,
xEventData.value('(/event/data[@name=''cpu_time'']/value)[1]','bigint') CpuTime,
CASE xEventName WHEN 'sql_batch_completed' THEN
xEventData.value('(/event/data[@name=''batch_text'']/value)[1]','varchar(max)')
WHEN 'rpc_completed' THEN
xEventData.value('(/event/data[@name=''statement'']/value)[1]','varchar(max)')
END AS SQLText,
xEventData.value('(/event/data[@name=''query_plan_hash'']/value)[1]','binary(8)') QueryPlanHash
FROM xEvents
ORDER BY PhysicalReads DESC
```

Fig. 4. Identifying queries that perform a large number of physical reads

### 4.3 Identifying queries that perform very often

```
WITH xEvents AS
(SELECT object_name AS xEventName,
CAST (event_data AS xml) AS xEventData
FROM sys.fn_xe_file_target_read_file
('C:\EventsData\QueryTest.xel', NULL, NULL, NULL))
SELECT
xEventName,
COUNT(*) as TotalExecutions,
SUM(xEventData.value('(/event/data[@name=''duration'']/value)[1]','bigint')) as TotalDuration,
SUM(xEventData.value('(/event/data[@name=''cpu_time'']/value)[1]','bigint')) as TotalCpuTime
FROM xEvents
GROUP BY xEventName,
ORDER BY TotalExecutions DESC;
```

Fig. 5. Identifying queries that perform very often

## References

1. Microsoft Corporation. Extended Events. http://technet.microsoft.com. [Online] http://technet.microsoft.com/en-us/library/bb630282.aspx.
2. Troubleshooting Performance Problems in SQL Server 2008. Sunil Agarwal, Boris Baryshnikov, Keith Elmore, Juergen Thomas, Kun Cheng, Burzin Patel. s.l. : Microsoft Corporation, March 2009, SQL Server Technical Article.
3. Fritchey, Grant. SQL Server 2012 Query Performance Tuning. 3rd. s.l. : Apress, 2013. ISBN: 978-1-4302-4204-8.
4. Delaney, Kalen. Microsoft SQL Server 2012 Internals. s.l. : Microsoft Press, 2013..
5. Vieira, Robert. Professional Microsoft SQL Server 2008 Programming. s.l. : John Wiley & Sons, 2009.

# Using the Data Collector in Microsoft SQL Server for the Query Performance Analysis

**Mitko Radoev**

Department of Information Technologies and Communications, UNWE, Sofia, Bulgaria

**Abstract.** Query performance is a major problem for every database. The efforts of the developers of the information systems are aimed at achieving optimal performance. However, the actual results do not always go to plan. Therefore, it is necessary to analyze the performance of the queries within information system already in operation. Depending on the results of this analysis, actions to improve the performance of the queries are performed. There are many tools for collecting and analyzing information for the execution of the queries in Microsoft SQL Server. Such relatively new tool is the Data Collector. This study aims to present the capacity of this tool and suggest options for using the collected information in the analysis of the performance of the queries.

**Keywords:** query performance, performance analysis, Data Collector

## 1.Introduction

The Data Collector is a component of Microsoft SQL Server that collects different sets of data. The Data Collector stores the collected data in a relational database or rather data warehouse. The Data Collector provides one point for data collection across database servers. This collection point can obtain data from different sources and is not limited to performance data. (1)

Data Collector was introduced in SQL Server 2008 and is available in the Standard, Enterprise, Business Intelligence, and Web editions. SQL Server Integration Services must be installed.

System behavior can be either tracked in real time in the form of graphs or captured as a log for offline analysis. The preferred mechanism on production servers is to use the log (2).

The main goals of this study are to present the capabilities of Data Collector and to suggest options for using the collected information in the query performance analysis.

## 2.Data Collector Terminology

- **Target**: An instance of the Database Engine in an edition of SQL Server that supports Data Collection.
- **Target type**: The type of target, which has certain characteristics and behavior.
- **Data provider**: A data source, specific to a target type that provides data to a collector type.
- **Collector type**: A logical wrapper around the SSIS packages that provide the actual mechanism for collecting data and uploading it to the management data warehouse.
- **Collection item**: An instance of a collector type.
- **Collection set**: A group of collection items.
- **Collection mode**: The manner in which the data is collected and stored.

• **Management Data Warehouse**: A relational database used to store collected data.

## 3.Data Collector concepts

The data collector is integrated with SQL Server Agent and Integration Services.

• **SQL Server Agent** is used to schedule and run collection jobs.

• **Integration Services** are used to execute packages that collect data from individual data providers.

The following figure shows how the data collector fits in the overall strategy for data collection and data management in Microsoft SQL Server.

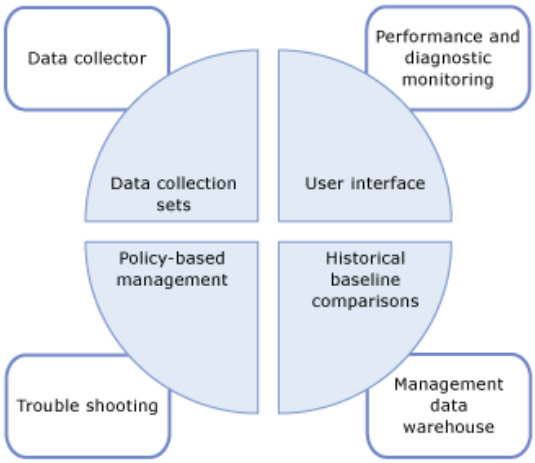



Fig. 6. Data collector in the overall strategy for data collection and data management (1)

## 4.Predefined data collection sets

Data Collector collects performance information and saves it in a data warehouse. It has three built-in data collecting specifications designed to collect the most important performance metrics. The information collected by default is about disk usage, query statistics, and server activity. Default data collectors are presented in the following table 1.

Table 6. Default data collectors

| Name | Description |
|---|---|
| **Disk Usage** | Collects information about disk space used by data and log files for all databases on the SQL Server instance, growth trends, and average day growth. By default, the Disk Usage data collector collects data every 60 seconds and uploads data into the data warehouse every 6 hours. |
| **Query Statistics** | Collects information about query statistics, activity, execution plans and text on the SQL Server instance. It does not store all executed statements, only 10 worst performing ones. |
| **Server Activity** | Collects information about resource usage (CPU, memory, disk I/O, and network), SQL Server waits, instance activity (batch requests, logouts, SQL compilations and recompilations, transactions, user connections, logins), and operating system activity. |

The Query Statistics data collection set has a single collection item that collects information about query statistics, activity, execution plans, and code on the SQL Server

instance. By default, it collects data every 10 seconds and uploads it in the cached mode every 15 minutes. (3)

## 5.Query Statistics Reports

SQL Server includes three built-in reports, corresponding to the predefined data collections:

- Disk Usage Summary Report
- Query Statistics History Report
- Server Activity History Report

Each of these reports provides more detailed information in the form of sub-reports. In addition to these standard reports, custom reports can be created using the SQL Server Business Intelligence Development Studio, or with any of the various Reporting Services tools. (4)

Query Statistics History Report presents the Top Queries that use the most resources, such as:

- Total CPU
- Duration
- Total I/O
- Physical Reads
- Logical Writes

The table below the graph provides detailed statistical information about the execution of queries.

This report contains two sub-reports:

- Query Details sub-report
- Query Plan Details sub-report

The Query Details sub-report presents the text of the query. It shows a graph of the query plans and average execution time.

Query Plan Details sub-report contains the information about the query plan. There is also a hyperlink that opens a graphical representation of the execution plan for the current query.

Query Statistics History Report is especially useful in analyzing the performance of queries because it contains detailed information about each query using significant resources over a period of time.

## References

1. Microsoft Corporation. Data Collection. MSDN. [Online] 2014. http://msdn.microsoft.com/en-us/library/bb677179.aspx.
2. Fritchey, Grant. SQL Server 2012 Query Performance Tuning. 3rd. s.l. : Apress, 2013. ISBN: 978-1-4302-4204-8.
3. Petrovic, Milena. SQL Server performance monitoring with Data Collector. SQL Shack. [Online] 2 19, 2014. http://www.sqlshack.com/sql-server-performance-monitoring-data-collector/.
4. McGehee, Brad. SQL Server 2008: Performance Data Collector. simple talk. [Online] 12 29, 2008.https://www.simple-talk.com/sql/learn-sql-server/sql-server-2008-performance-data-collector
5. Vieira, Robert. Professional Microsoft SQL Server 2008 Programming. s.l. John Wiley&Sons, 2009.

# Dashboard Design for Dynamic Management of Database Performance

**Veska Mihova**[1]

[1] UNWE, 1700 Sofia, Bulgaria
vmihova @unwe.bg

**Abstract.** With the growth of managed and stored data, a need for monitoring and management systems arises. The companies are increasingly interested in performance monitoring of the systems which manage their business. This report is a result of current and previous research of systems' performance and database productivity in a system. This report will propose design of a dashboard for dynamic management of database performance.

**Keywords.** Database, database monitoring, database administration, database performance.

## 1.Introduction

The solution of the problem with the systems' productivity was created, but in some aspects there are serious weaknesses provoking the search for better solutions.

Monitoring systems include operation, monitoring database productivity, monitoring critical points of systems, and detection of errors and system problems. The problem is that they follow the current state, i.e. when the problem is ongoing. A stable decision would be a decision that allows monitoring predicting upcoming problems.

Previous researches present an entire concept for management of database performance. [5] The idea covers tasks and concepts that are very similar to those of business intelligent systems - data collection, storage history, producing new values based on the accumulated history, forecast. The previous papers take the idea of the architecture of business intelligent systems and develop it further to serve the idea of future forecasted environment. This paper will propose design of a dashboard for monitoring the future forecasted environment.

## 2.Definition of a dashboard

A business intelligence dashboard is a data visualization tool that displays the current status of metrics and key performance indicators (KPIs) for an enterprise. Dashboards consolidate and arrange numbers, metrics and sometimes performance scorecards on a single screen. They may be tailored for a specific role and display metrics targeted for a single point of view or department. The essential features of a BI dashboard product include a customizable interface and the ability to pull real-time data from multiple sources.

Digital dashboards allow managers to monitor the contribution of the various departments in their organization. To gauge exactly how well an organization is performing overall, digital dashboards allow you to capture and report specific data points from each department within the organization, thus providing a "snapshot" of performance.[1,3]

Benefits of using digital dashboards include:[4,2]

- Visual presentation of performance measures

- Ability to identify and correct negative trends
- Measure efficiencies/inefficiencies
- Ability to generate detailed reports showing new trends
- Ability to make more informed decisions based on collected business intelligence
- Align strategies and organizational goals
- Saves time compared to running multiple reports
- Gain total visibility of all systems instantly
- Quick identification of data outliers and correlations

Dashboards are unique. The design of each dashboard is driven by the business and its needs and culture. What may work for one business may not work for another. There are general guidelines that are available when initially developing the tool or when looking to improve current performance metrics.

Planning and researching a good design is crucial for dashboards. A good information design will clearly communicate key information to users and makes supporting information easily accessible.

## 3.Dashboard design for dynamic management of database performance

A dashboard design can be created through a variety of tools. In this research, it is taken into account that a dashboard can be created through the QlikView tool.

### 3.1. Page „Dashboard“

This page is designed for administrators of the application. Conditionally the screen in this page is divided into two parts - the right to visualize number of transactions per day, and in the other part, the key indicator is the execution time of queries. There are some visual elements that clearly define their message without taking a lot of space. The collected information has to be visualized in a limited area of the screen.

The information on the dashboard is represented visually as a combination of graphics and it is an overview of the status of the monitored databases. There are two main types of graphics used on the screen - Gauge Chart and Line Chart.

The first Gauge Chart diagram shows the average number of transactions per day. Its value is changed by the selections per day or per hour.

The next diagram is a Line Chart diagram showing the number of transactions per day. This type of diagram is selected, because the trend can be seen per days, in which days of the month there are increases and decreases.

The other Gauge Chart diagrams show the average execution time in seconds per day, maximum transaction execution time in seconds, minimum transaction execution time in seconds and thresholds of execution time in seconds. The values are changed by the selections per year, per month, per day or per hour. So the workload of the system can be monitored in certain periods.

With the help of the visual values, the state of the indicators is monitored, to see whether there are large deviations from the specified critical threshold. That way, the days when there were critical values, can be determined, and the reasons for the deviation can be analyzed.

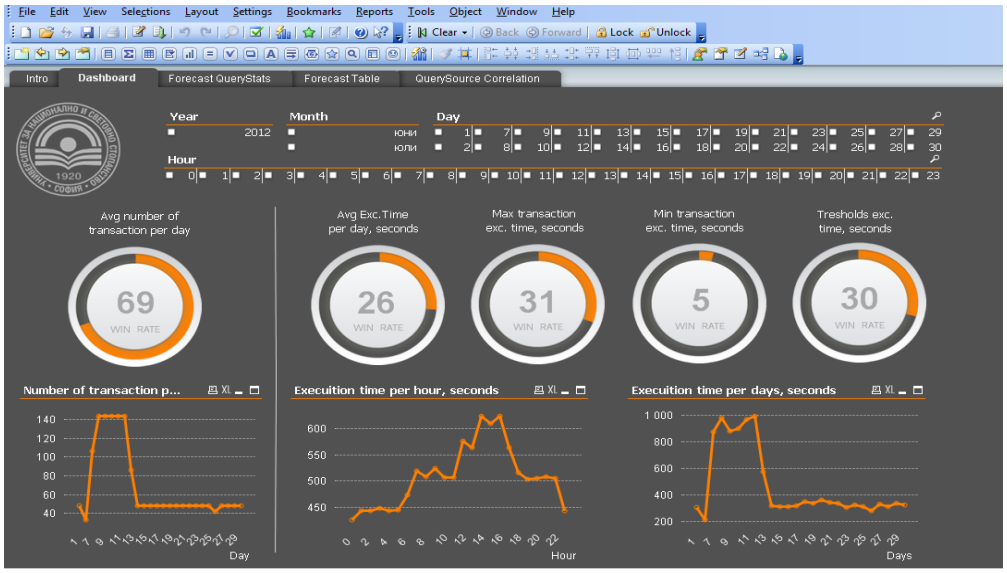



Figure 7 Page „Dashboard“

The second Line Chart diagram shows the execution time in seconds per hour. This type of diagram is selected, because the trend can be seen per hour, in which hour of the day there are increases and decreases.

The third Line Chart diagram shows the execution time in seconds per days. This type of diagram is selected, because the trend can be seen per days, in which days of the month there are increases and decreases.

### 3.2. Page – „Forecast QueryStats“

The page "Forecast QueryStats" displays the values for execution time and the forecasted values for execution time of registered business problems in the system, put against the identified critical thresholds.

The data on this page is visualized through two Line Chart diagrams. The graphic allows for a quick comparison between the base statistics and the projected ones, making sure that they reach the critical value.

The First Line Chart chart shows average execution time per day, measured in seconds. The quantitative values are - basic execution time per day, predicted execution time per day and the critical value of the execution time per day. The trend lines per days, are easily tracked.

The second Line Chart diagram shows the average execution time per hour, measured in seconds. The quantitative values - basic execution time per day, predicted execution time per day and the critical value of the execution time per day.

### 3.3. Page „Forecast Table“

The page "Forecast Table" displays information about the number of rows in the sources of data recorded for monitoring, and their forecast data. Both values are measurable, quantity is compared with quantity, this is the reason for the selection of the type of graph - Bar Chart. On this page the data is visualized through two Bar Chart diagrams.The first Bar

Chart shows the average number of rows to the table per hour.The second Bar Chart shows the average number of rows in a table per day.

On the page there is also a List box - Table, which allows an additional filter by source. Individual tables and the number of rows in them, are easily tracked.

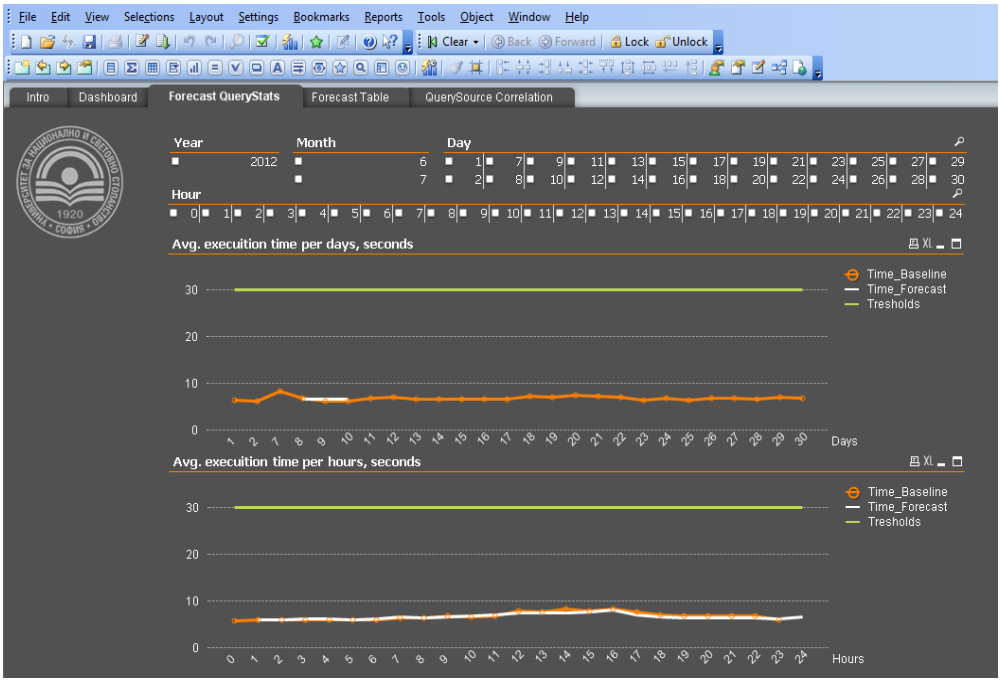



Figure 8 Page „Forecast QueryStats“.

### 3.4. Page „QuerySource Correlation“

This page gives a detailed analysis of the correlation between data sources and applications. This is the so called and applied in previous studies - correlation analysis. In the analysis, the coefficients of correlation between one or more pairs of variables are compared, in order to determine statistical correlations between them. In this case, the two variables for which correlation coefficients are calculated are: number of records in a table data source and time of query execution. For a graphical representation of this, a Line chart and a Pivot table are selected.

The first diagram shows the trend of the correlation coefficient by day. In tabular form, the data is grouped by application, by source and by days. Based on the displayed values, it can be seen that the correlation coefficient is in the range between 0 and 0.3, i.e., the correlation is weak. In order to be able to make a selection in the data source, two additional filters of the page are placed.

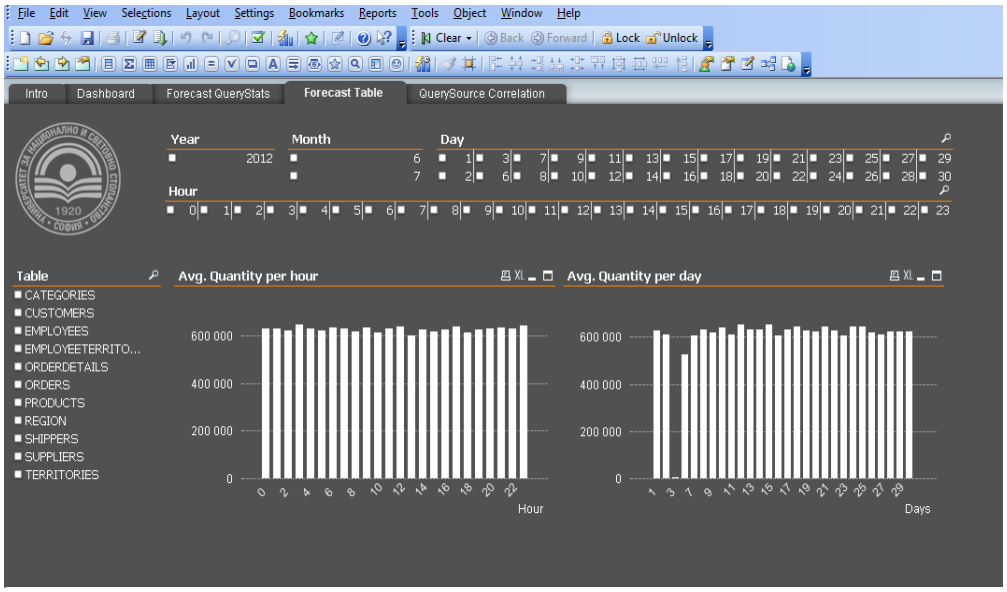



Figure 9 Page „Forecast Table“

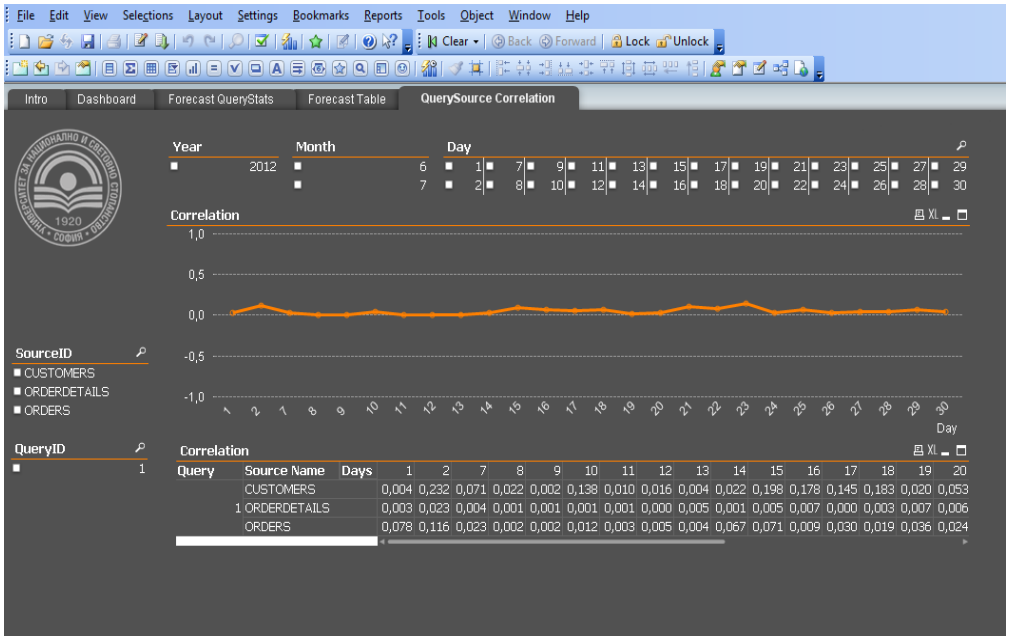



Figure 10 Page „QuerySource Correlation“

## 4.Conclusion

In summary, in this paper a variant of dashboard design for monitoring of current and future state of a database is presented.

In the beginning the author presents the definition of a dashboard and benefits of using digital dashboards. The author proposes design of four dashboard pages in help of database administrators. The proposed solution gives the database administrator enough time for reaction to an occurring problem with the application's performance and respectively the database's performance.

## References

1. http://www.dashboardinsight.com/articles/digital-dashboards/fundamentals/what-is-a-dashboard.aspx#sthash.vHPsWUWO.dpuf
2. Briggs, Jonathan. "Management Reports & Dashboard Best Practice". Target Dashboard. Retrieved 18 February 2013.
3. http://searchbusinessanalytics.techtarget.com/definition/business-intelligence-dashboard
4. http://www.solarwinds.com/solutions/sql-database-monitor.aspx. Performance Monitoring & Management Solutions.
5. Mihova, V., Monitoring the future state of a database, UNITE 2nd Doctoral Symposium R&D in Future Internet and Enterprise Interoperability, 11-12.10.2012, Sofia.

# Automated Taxonomy Generation for Intelligent Tagging and Search

**Maria Marzovanova**

UNWE, 1700 Sofia, Bulgaria
mmarzovanova@unwe.bg

**Abstract.** Intelligent tagging and search is an approach for organizing unstructured data that includes mechanisms for automated text tagging based on metataxonomy. Metataxonomy is a specific extended taxonomic structure that is developed over a traditional manually created taxonomy involving different kinds of relationships between the nodes and adding terms from online sources. The purpose of this paper is to outline aspects for research on developing a methodology for automated generation taxonomy that will be kind of infinite self-learning process.

**Keywords.** Tagging, search, extended taxonomic structure, text documents, automated taxonomy generation.

## 1.Introduction

Unstructured data needs to be read and categorized in order to be useful and this is a well-known fact. Another fact is that the most organized and structured way to do it is using a taxonomy. Taxonomies allow the pieces of information to be arranged in terms of location and storage mechanism in advance. In this way, a considerate taxonomy can be a relief for information management systems.

The process of creating a taxonomy manually consists of two main stages. First, the taxonomist have to collect a large set of terms. Only terms, which are bound to the domain of the taxonomy, and whose usage frequency is not so high that they have lost their specific meaning, can be candidate participants for the taxonomy.

The second stage is the art stage. All those terms have to be arranged considering and understanding the relations between one another. The main source for the decisions made during this stage is the human analytical mind.

To bring the problem and the solution today, we can say, that because unstructured data is growing every second and expanding its content, a taxonomy created to bring order to it, is already unusable, the moment it is finished. So manually creating taxonomies is out of the question.

## 2.Automated taxonomy generation

What came to replace the taxonomist in the process mentioned above are the systems for automated taxonomy generation. The automation appears on both stages, dividing the process in a few more steps. Taxonomy generation systems have the following workflow:

- Splitting a certain text source in separate words and phrases,
- Determining the terms connected to the taxonomy domain among all candidate elements,
- Arranging the terms into a taxonomy.

All of these steps have been subject of multiple studies and discussions. There are

different algorithms developed and explained. Back in 2002, Blaschke introduces a methodology to extract ontology-like structured knowledge directly from the literature, based on statistical treatment of documents [2]. In 2003 Krishnapuram working on the subject again, discusses the problems of automated taxonomy generation. Their article considers a few approaches for creating taxonomies and gives detailed view of the involved issues [5].

Internet and social networks, as a specific aspect, are the main causes of the growth of the information nowadays. On the other hand they are the most powerful source that can be used to bring order to the resultant unstructured piles of knowledge. One of the first insights on that idea is authored by Sanchez in 2004 [8], who proposes a method for deriving a taxonomy using web resources. A few years later, social tagging becomes the source for automated taxonomy generation in the works of Schwarzkopf [7].

What they all have in common is that they are trying to substitute the human analytical mind using the power of text-mining, mathematical calculations and knowledge extracted from diverse databases.

## 3.Automation in creating metataxonomies

A metataxonomy is a structure that starts as a traditional taxonomy of terms and evolves to a rich structure with diverse meaningful relation types and wide range of vocabulary. Most of the processes included in that evolution are automated, except for one, the creation of the basic taxonomy. In the aspect of intelligent tagging and search necessities, when referring to machine created taxonomy, we can't establish our limits to one time execution of a generation process.

In order to keep a system for intelligent tagging and search adequate, it's important to develop and keep its metataxonomy as updated as possible. If we take a look on the separate procedures constituting the generation process, we can determine, which of them need to be improved in order to suite that specific need.

Taxonomy generation starts with extraction of keywords. Our aim here is to pile up the wider range of related term as possible. For example, Yutaka Matsuo [6] explains an algorithm for keyword extraction that applies to a single document. His methodology involves extracting frequent terms, then he generates co-occurrence distribution of each term and the other terms and makes his conclusions based on that distribution. If probability distribution of co-occurrence between a term and the frequent terms is biased to a particular subset of frequent terms, then the term is likely to be a keyword. The degree of biases of distribution is measured by the $\chi 2$-measure. He selects the terms with high values as keywords. He works with one document with the idea to make the algorithm simple and not domain-independent. The freedom involved in intelligent tagging and search adopts the idea of domain-independent taxonomy. This method on the other hand can be improved to use more than one document in order to get wider range of keywords.

The next step is calculating similarity. Arranging keywords in a taxonomic structure is based on the relation between these keywords. The main insight here is to present this relation in numbers, based on the frequency of each two terms appearing together in literature, and that is called similarity. There are different kinds of similarity that are considered appropriate – cosine similarity, google similarity distances (sNGD [3] and aNGD [4]).

For example, Cosine similarity is calculated with the formula:

$$cosine\ similarity = \frac{n_{x,y}}{\sqrt{n_x} \cdot \sqrt{n_y}}$$

where Nx and Ny are the count of articles where each of the terms appears separately, and Nxy is the number of the articles where the two terms appear together. One word appears in an article if it is part of the title, the abstract or the list of keywords. You can easily see that cosine similarity is symmetric, that means that the similarity of x with y is the same as the similarity of y with x. We can't deduct the direction of a parent-child relationship between the keywords, for example.

A way of improvement here is to make sure that we calculate similarity using the one that is based on more literature sources, preferably online sources, and to provide access to those sources. The more the sources are, the less the probability to miss a term, which is new or rare.

After the similarity is measured, the results for each couple of terms create a distance matrix. This is a square matrix with all the terms on both dimensions. Each element represents the similarity between the term from the column and the term from the row. In case asymmetric similarity is calculated, the values are associated with column-row as parent-child relation between the terms. The distance matrix is a start situation for all generation algorithms.

Steven Camina [1] presents a detailed paper on automated taxonomy generation and describes in details some of the most popular algorithms:

- Dijsktra-Jarnik-Prim‟s (DJP) Algorithm,
- Kruskal‟s Algorithm,
- Edmond‟s Algorithm,
- Heymann Algorithm,
- The Genetic Algorithm.

The generation algorithms need a starting node to do their magic. That root node can be selected manually or by calculation. The measures for this calculation are Betweenness cenrality and Closeness centrality. Up to Camina they all, except for the Genetic algorithm, need almost the same time to finish and differ only in terms of implementation.

On this step, agreeing with Camina's conclusion, there will be probably not much to modify that will result in metataxonomy expansion. But these algorithms and some other successful researches should be reviewed and evaluated in terms of implementation and agility.

Figure 1 shows a summarized view of the process of automated taxonomy generation. The two main steps in the beginning (surrounded in the picture) should be subject of improvement for making the process appropriate for metataxonomy generation.

## 4.Conclusion

The review of the different algorithms for automated taxonomy generation clearly shows the possibilities to adopt one or more of the specific approaches and develop an algorithm for automated generation of metataxonomy. Most of the measures used for actually building a taxonomy can be influenced by the size of the sources. With applying these mechanism on larger foundation and doing the necessary modifications will result in creating the desired rich taxonomic basis, which seems to be a promising area for future research.

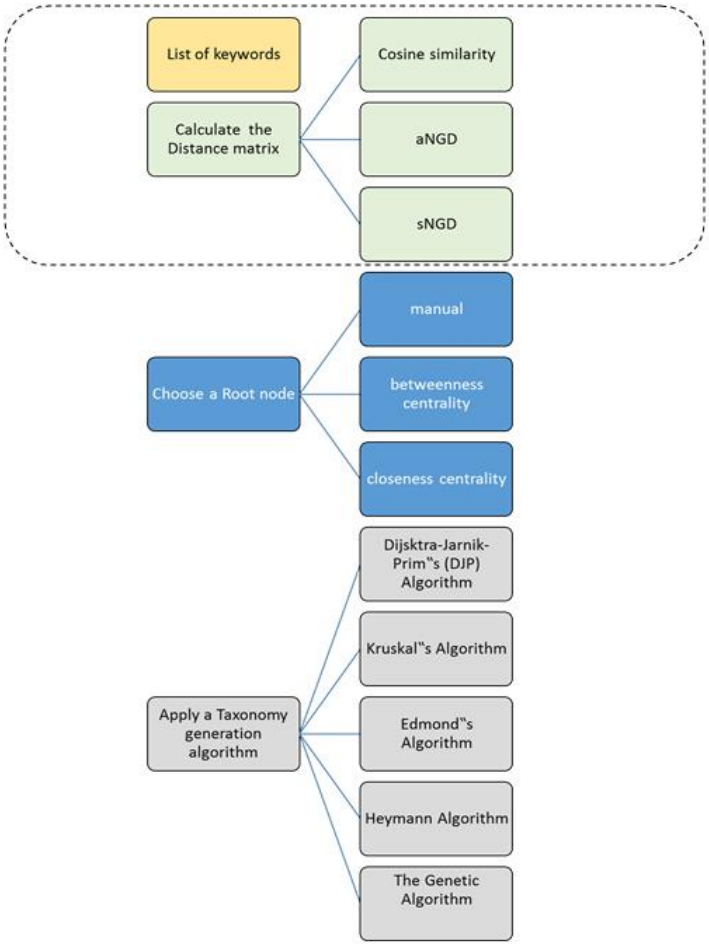



Figure 11. Summary of the process of automated taxonomy generation

## References

1. *A Comparison of Taxonomy Generation Techniques Using Bibliometric Methods: Applied to Research Strategy Formulation*2009
2. Blaschke, C., Valencia, A. *Automatic Ontology Construction from the Literature. Genome Informatics, Volume 13*, 2002
3. Cilibrasi R.L., Vitanyi P. M. B.2007*The google similarity distance. IEEE Trans. Knowledge and Data Engineering*
4. Firat, A., Woon W., Madnick S.2008*Technological Forecasting – A Review. Composite Information Systems Laboratory (CISL)*
5. Krishnapuram, R., Kimmamuru K2003*Automatic Taxonomy Generation: Issues and Possiblities. Lecture Notes in Computer Science*Berlin
6. Matsou, Y.2003*Keyword Extraction from a Single Document*
7. Mining the Structure of Tag Spaces for User Modeling.2007 *Data Mining for User Modeling, International Conference on User Modeling*
8. Sanchez, D., Moreno, A.2004*Automatic Generation of Taxonomies from the WWW. Practical Aspects of Knowledge Management*

# Building Multi-Touch User Interface

**Maria Marzovanova**

UNWE, 1700 Sofia, Bulgaria
mmarzovanova@unwe.bg

**Abstract.** Multi-Touch devices have set a number of new requirements for application development concerning both design and functionality. The focus of this paper is on how newly released tools and components for software development meet these requirements.

**Keywords.** User interface,multi- touch, components.

## 1.Introduction

Multi-touch is a technology that changed the traditional input style. Compared to using a mouse and a keyboard, interacting with a graphical interface just using your fingers is way faster and easier. It can be more powerful, but problems arise for designers and developers. Most frameworks provide only a limited set of standard gestures, and this is an issue that needs attention.

## 2.Multi-touch gestures

Let's begin with what the formal standard is and then continue with the created frameworks for programming for multi-touch. We can start with the multi-touch gestures, supported by the different OS. They cover almost the same basic functionality, such as pointing, rotating, scaling and moving objects, some may differ in the way to perform, but that differences are minimal.

The three tables below list the standard sets of gestures for Windows, Android and iOS.

**Table 3. Windows gestures**

| ***Gesture*** | ***How to perform*** |
|---|---|
| Pan | Touch and drag the page with either one or two fingers. |
| Zoom | To zoom out, touch two points on the item, and then move your fingers toward each other, as if you're pinching them together. To zoom in, touch two points on the item, and then move your fingers away from each other, as if you're stretching them apart. |
| Rotate | Touch two points on the item, and then move the item in the direction that you want to rotate it. |
| Press and tap | Press the item with one finger, then quickly tap with another finger, while continuing to press the item with the first finger. |

Table 4. Android gestures

| ***Gesture*** | ***Action*** | |
|---|---|---|
| Touch | Press, lift | Triggers the default functionality for a given item. |
| Long press | Press, wait, lift | Enters data selection mode. Allows you to select one or more items in a view and act upon the data using a contextual action bar. Avoid using long press for showing contextual menus. |
| Swipe or drag | Press, move, lift | Scrolls overflowing content, or navigates between views in the same hierarchy. Swipes are quick and affect the screen even after the finger is picked up. Drags are slower and more precise, and the screen stops responding when the finger is picked up. |
| Long press drag | Long press, move, lift | Rearranges data within a view, or moves data into a container |
| Double touch | Two touches in quick succession | Scales up a standard amount around the target with each repeated gesture until reaching maximum scale. For nested views, scales up the smallest targetable view, or returns it to its original scale. Also used as a secondary gesture for text selection. |
| Double touch drag | A single touch followed in quick succession by a drag up or down:<br>•Dragging up decreases content scale<br>•Dragging down increases content scale<br>•Reversing drag direction reverses scaling. | Scales content by pushing away or pulling closer, centered around gesture. |
| Pinch open | 2-finger press, move outwards, lift | Zooms into content |
| Pinch close | 2-finger press, move inwards, lift | Zooms out of content |

It's easy to notice that the few gestures supported by the OS are insufficient when you see the potential of multi-touch devices. As the authors of 'Towards a Formalization of Multi-touch Gestures' write, the standards lack a formal abstraction of high-level properties of the available gestures. This insufficiency leads to more complicated programming, because programmers have to work with the low level touch data. (Kammer, Wojdziak, & Keck, 2010).

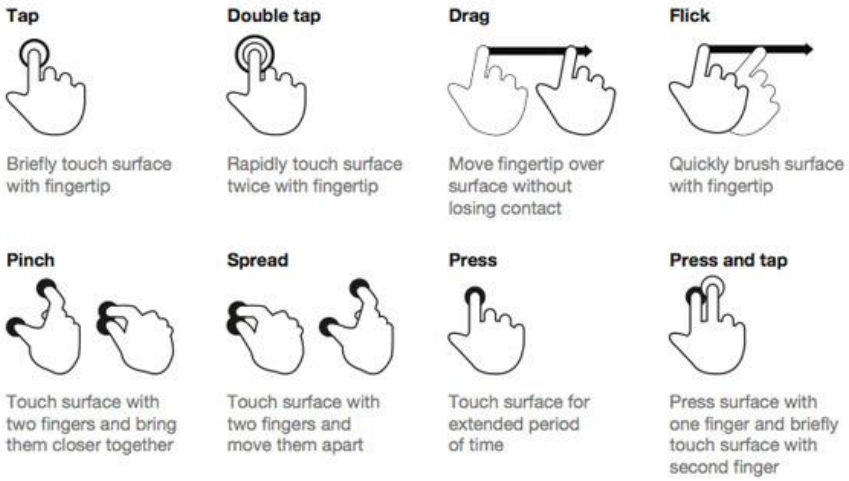



Figure 12. iOS gestures

Similar topic is involved in an entire chapter in 'Brave NUI World' (Wigdor & Wixon, 2011). They start the discussion with the fact that the action 'click' is still used in everyday speech even when it's performed on a multi-touch device and it's not a click at all. A 'Click' is a primitive, as they call it, which belongs to desktop programming. Primitives are basic action supported by the environment, or the building blocks of an interaction language. There is a picture provided for understanding their role.

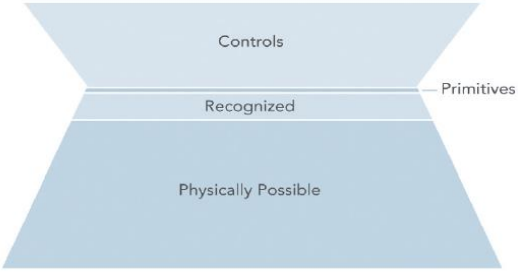



Figure 13. Primitives.

The main idea in this chapter of the book is to point that after refining all actions considering what is physically possible and what is recognized by the used hardware, the result is a list of primitives. The diverse combinations of primitives gives the controls used for application development.

## 3.Multi-touch Frameworks

A lot of works published around 2010 and 2011 on the topic, state that either more gestures are necessary, or a layer, that will distance the developer from the low level programming. One solution to that problem proposes Lode Hoste (Hoste, 2010), a driver which converts low level signals to TUIO events, a protocol that is supported by many multi-touch interfaces.

Another interesting work called Midas, is a declarative multi-touch interaction framework (Scholliers, Hoste, Signer, & Meuter, 2011). They divide a software engineering problem into two complexities – accidental and essential. As we can't reduce essential complexity, which due to the characteristics of the problem, they see a way to reduce accidental complexity, related to the choice of software engineering tools.

There is a taxonomy of some of the most popular multi-touch frameworks (Kammer & Mandy Keck, Taxonomy and Overview of Multi-Touch Frameworks: Architecture, Scope and Features, 2010). The authors create three categories – Features, Scope and Architecture and review the popular frameworks according to a list of criteria for each category. That list of criteria seems to meet the requirements of software developers, so the whole review is very useful and interesting.

The frameworks involved in the review are:

- MT4j
- SparshUI
- Surface SDK
- Breezemultitouch
- Miria
- Grafiti
- libTISCH
- PyMT
- GestureWorks.

In terms of architecture, the criteria list includes platform independence and event system specifics. The Scope group refers to the support for tangible objects and touch information. And the third list consists of features like standard gestures included in the frameworks, gesture extensibility and visualization support.

As we mentioned before, one of the main roles of the framework should be to extend the basic gesture set, which is proven to be insufficient. The framework should provide interfaces that can be extended for creating new gestures. The other level that should allow improvement is the set of visual components, either by extending available components, or by creating new ones.

## 4.Conclusion

This paper reveals the need of a layer distancing developers of multi-touch application of the low level programming and allowing them to focus on the actual gesture interaction and their interpretations for the applications. Multi-touch frameworks come to fill in this gap, providing diverse components and tools.

## References

1. Brave NUI World: Designing Natural User Interfaces for Touch and Gesture2011Morgan Kaufmann Publishers
2. Midas: A Declarative Multi-Touch Interaction Framework2011TEI 2011
3. Software Engineering Abstractions for the Multi-touch Revolution2010ICSE 2010
4. Taxonomy and Overview of Multi-Touch Frameworks: Architecture, Scope and Features2010Workshop on engineering patterns for multi-touch interfaceBerlinEICS
5. Towards a Formalization of Multi-touch Gestures2010ITS 201049-58Dresden, Germany

# Development of an Integrated Learning Experience Using an Educational Social Network as a Service

**Venko Andonov**

UNWE, 1700 Sofia, Bulgaria
vandonov@unwe.bg

**Abstract.** The benefits of using an Educational Social Network (ESN) in the higher education are well-known and an increasing number of institutions consider using such platforms. The design and implementation of an ESN for a given educational institution can follow three approaches: using existing public social networks; deployment of a standalone solution; using cloud services. A combination of these methods could be beneficial for both learners and educators and could simplify and accelerate the deployment lifecycle of a private solution. An architecture for an ESN as a service is proposed and implemented to investigate its efficiency.

**Keywords.** educational social network, services, learning.

## 1.Introduction

An Educational Social Network (EduSN, to be distinguished from enterprise social network – ESN) is a specialized online social network providing either an end-user software product or a platform and framework for enhancing the learning process for all of the participants in a given formal or informal educational environment [1, 2, 3]. In the higher education, such a system might be an engaging way to involve the students in teamwork; professional communication; sharing knowledge and experience; aggregation and evaluation of information sources; presentation of the individual skills; discussions; construction of knowledge. The most significant difference from a traditional e-learning system is the main workflow – it is not focused around the content itself but the connections between the participants or between the participants and the content, formed actively or passively during the interaction with the EduSN. Thus the purpose of an educational social network is not to replace the existing e-learning solutions with a better alternative, but to give both the students and the lecturers an additional environment, consisting of familiar tools and behavioral patterns.

Multiple studies [1, 4, 5] have shown that applying social networking concepts to an educational process increases the overall activity and satisfaction of the participants. However, multiple issues arise mostly around the privacy of the public social networks, the available toolset and lack of integration with the educational process itself. In order to implement an EduSN, an educational institution might follow one of these approaches: use an existing public social networking site; deploy its own standalone solution; use one of the available cloud services for social networks. Following the conclusions from the existing research, we have previously decided to use the second approach – a custom standalone solution, integrated with the internal information systems of the University of National and World Economy.

The challenges ahead of the Educational Social Networks are classified in three main categories: implementation, integration and adoption. While an implementation of an Educational Social Network might seem trivial, the low scalability of the social graph [6, 7]

might be critical in the case of a large user base (for example, University of National and World Economy has more than 20,000 active students and 550 full-time professors). To address this issue a specialized data model of EduSN has been designed, developed and tested and the results show that it seems to be an appropriate solution for the selected use cases. The integration with internal and external systems is a challenging task as well, because of the many different protocols and schemas of these services. This could be addressed by providing high level abstractions for the developers of the concrete EduSN. While the adoption of an EduSN seems like a non-technical issue, the other challenges (responsiveness, social graph traversal performance, feature-rich integrated environment, overall user experience) are critical for the motivation of the users to be active in the system.

The Software as a Service (SaaS) and Platform as a Service (PaaS) concepts represent "new ways of thinking about the delivery of computing capabilities in the context of cloud computing" [8]. While SaaS allows the "access to platform-enabled application by typically using a web browser" [9], PaaS consists of "specially crafted software development kits that allow developers to produce their own programs using all of the cloud-supported features" [9, 10]. The PaaS model has multiple benefits for the developers since it "shifts a significant portion of the traditional effort associated with developing, distributing and maintaining software towards the platform provider's end." [8]. The significant differences in subject areas, necessary toolsets, cultural specifics and accepted policies between the universities require the EduSN to be an open platform for customization and extension. At the same time, development of a product from scratch and managing the scalability issues on their own infrastructure would require a large and experienced IT staff. Development and deployment on top of several well-accepted abstractions would allow for a much better agility and focused, continuous customization efforts.

## 2.Educational Social Network as a Service

A dual-layer data model has been designed and implemented with the following goals: being a platform which integrates external services and gets integrated into such (both client or server-side) to serve as a behind-the-scenes social tier in the educational experience; natural, developer-friendly interface for educational social graph access, traversal and modification; abstractions, allowing the extensibility and evolution of the specific data model; easily swappable infrastructure components (persistent storage, caching, logging, authentication and authorization, web service stack). The logical layer is an abstract graph representation of the in-memory, persistent and remote storages and services. The educational social graph is a multimodal, multirelational, directed graph with key-value properties attached to each node and edge. The physical storage of this graph is a subject to multiple adapters, which process a given subgraph and its metadata in order to retrieve or modify the data. The functional layer models the specific educational context and workflow. It includes a set of basic data structures and operations, which are extensible.

From PaaS point of view an EduSN might provide a software development kit (SDK) allowing developers to extend and customize the functional layer as defined above. This allows for the development of complex software solutions on top of the educational social layer provided by the specified data model. A significant downside could be the vendor lock-in which is typical for the PaaS model [9].

The SDK allows for the following design and implementation steps:

- Analysis of the organizational structure, educational process specifics and the policies of the educational institution;

- Analysis of the specific requirements for the developed implementation of an EduSN – for example, only lecturers should be able to create groups; group topics can include the courses in the curriculum, etc.;
- Definition of the types of the nodes in the educational social graph following the principles of the object-oriented design and development; consists of the following components:
    - Unique identifier of the type – for example: Student, Article, Presentation, Exam;
    - Base type;
    - A list of attributes which put constraints and enforce policies;
    - An existing adapter for a physical data source;
- Definition of the value and relationship constraints – simple (like a number range) or complex (who can join a specific type of group);
- Execution of a migration procedure – translation of the current version of the definition to a specific physical data representation through adapters.

These steps are supposed to be executed in an iterative fashion, allowing for short product delivery cycles and frequent feedback from the end-users. The addition of unit and integration tests would allow for continuous delivery.

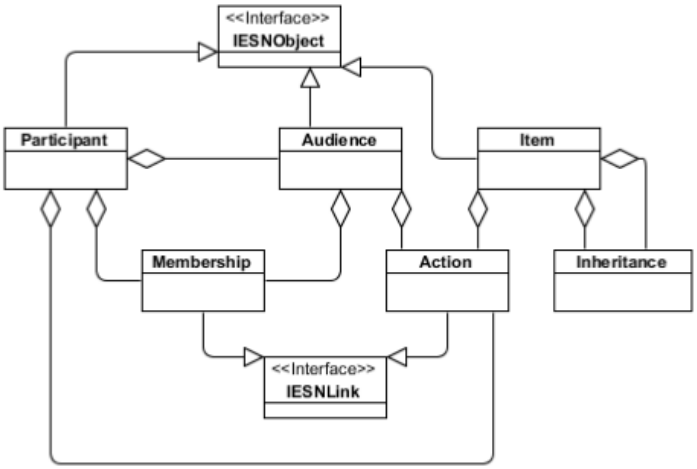



Figure 14: Base types available in the functional layer

The base node and edge types available for extension at the functional layer are shown on **Figure 14**. There are three node types (Participant, Audience and Item) and three edge types (Membership, Action and Inheritance). These are meant to cover the basic building blocks of a social graph. Each Participant can create zero or more Audiences and is simultaneously a part of one or more different Audiences. Each Participant creates zero or more content Items and allows the visibility of their Actions with the Items to a given Audience. Each Item can inherit another one creating a content tree.

## 3.Implementation

An implementation has been developed as a Microsoft .NET library. It depends on the OWIN standard interface and is therefore easy to test and deploy on any compatible host (IIS, Windows Azure, self-hosted process, etc.). An implementation of an Educational Social Network as a Service requires a modification of the current information architecture in order to allow for a multi-instance, multi-version solution. A multi-instance EduSN can be defined as a feature allowing any set of predefined users to create their own self-contained educational social platform. A multi-version EduSN would allow instances to be feature independent of each other and evolving on their own pace according to the specific requirements. From the provider's point of view there are two approaches to achieve this: multi-tenancy on a single deployment with appropriate software mechanisms to separate the instances; or separate deployments running in parallel.

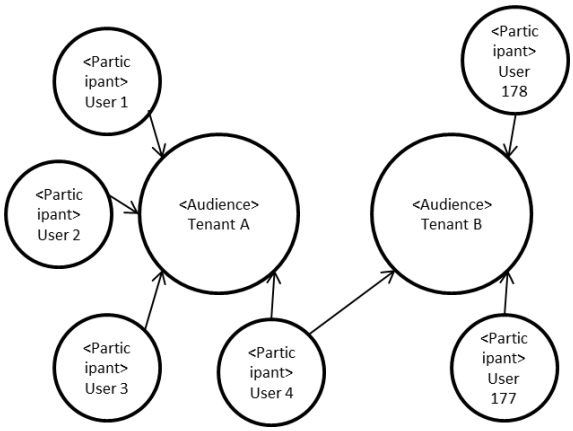



Figure 15: A multi-tenancy approach – a special type of Audiences play the role of instance selectors

In the multi-tenancy approach an Audience separates a given instance from another by including only the necessary participants, content and relationships. In a classical separate deployment scenario, each EduSN instance is independent at the cost of a significant overhead.

In order to select a preferred approach, both alternatives were tested for capacity and performance. The capacity is considered to be the maximum number of instances that can fit into 500 MB of memory, where an instance contains 14,000 students, 500 lecturers; 2,600 subject groups and 10,400 discussions, the relationships between them are defined based on the existing curriculum data. The read performance is evaluated as an execution of a newsfeed operation (100 sequential runs with random input of a participant UID) and the write performance contains an execution of a group join/leave operation (100 sequential runs with random input of a participant UID and group UID).

As expected, in the first scenario we are able to store more data at the cost of a slightly increased latency caused by the larger dataset. In the second scenario, we are able to hold fewer instances but with a slightly better performance.

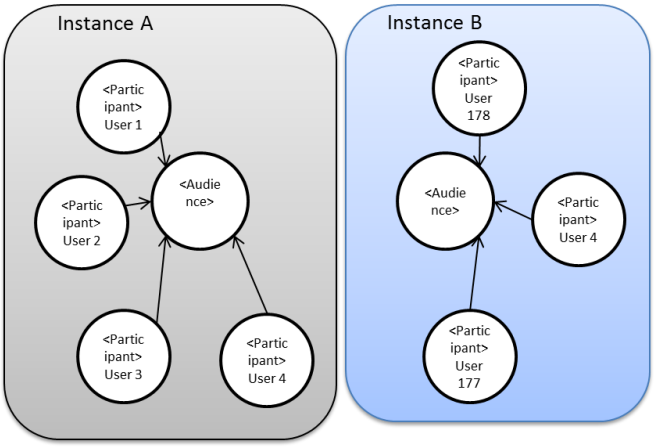



Figure 16: Separate deployments of EduSN – they might share hardware and application server but are logically separate from each other

Table 5: Comparison of the suggested alternatives

| *Scenario* | *Max. Instances on same hardware* | *Average Read performance (newsfeed)* | *Average Write performance (group join)* |
|---|---|---|---|
| Multi-tenancy | 12 | 0.101 s | 0.065 s |
| Separate deployments | 7 | 0.088 s | 0.60 |

## 4.Conclusions and Further Work

Since the possibility of alternative deployment methods of Educational Social Networks has not been studied before, this simplified implementation and test attempts to pick a better approach for further experiments in this direction. The results confirm the expected behavior – a trade-off between performance and capacity. However, the multi-tenant environment would be much easier for administration and it can be accepted as a preferred solution.

Further research is necessary in order to provide a better understanding of the benefits and the downsides of building an Educational Social Network as a Service – how to provide straightforward mechanisms to enable secure API access to internal information systems; evaluating the latency of the Internet connection compared to the use of a local network in an alternative on premise solution; concerns related to data security and privacy in the cloud; a developer-friendly toolkit for easy integration with external systems like MOOC (Massive Open Online Course) websites, Wikipedia, YouTube, Facebook and others; creating an open API and extensions (add-ins) ecosystem.

## References


1 Wodzicki K., E. Schwämmlein, J. Moskaliuk, *"'Actually, I Wanted to Learn': Study-related knowledge exchange on social networking sites,"* The Internet and Higher Education, vol. 15, no. 1, pp. 9-14, 2012.

2 Popescu E., D. Cioiu, *"eMUSE - Integrating Web 2.0 Tools in a Social Learning Environment"*.

3 Vassileva J., *"Toward Social Learning Environments,"* vol. 1, no. 4, pp. 199-214, 2009.

4 Bosch T.E., *"Using online social networking for teaching and learning: Facebook use at the University of Cape Town"*,Communicatio, 35, 2, pp. 185-200, 2009.

5 Roblyer M.D. et al., *"Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites"*, The Internet and Higher Education, 13, 3, pp. 134-140, 2010.

6 Adell J., *"Performance of Graph vs. Relational Databases"*, http://blog.everymansoftware.com/2011/07/performance-of-graph-vs-relational.html, 2011.

7 Hoff T., *"Why are Facebook, Digg and Twitter so hard to scale"*, http://highscalability.com/blog/2009/10/13/why-are-facebook-digg-and-twitter-so-hard-to-scale.html, 2009.

8 Kourtesis D., Kuttruff V., Paraskakis I., *"Optimising Development and Deployment of Enterprise Software Applications on PaaS: The CAST Project"*, Towards a Service-Based Internet, Lecture Notes in Computer Science, vol. 6569, pp. 14-25, 2011.

9 Teixeira C. et al., *"The Building Blocks of PaaS"*, Journal of Network and Systems Management, vol. 22, 1, pp. 75-99, 2014.

10 Walraven S. et al., *"Comparing PaaS offerings in light of SaaS development"*, Computing, vol. 96, Issue 8, pp. 669-724, 2014.

# Components of Cloud-Based Information Systems

**Vanya Lazarova[1], Ivan St. Ivanov[2]**

[1] UNWE, 1700 Sofia, Bulgaria
vlazarova@unwe.bg
[2] UNWE, 1700 Sofia, Bulgaria
ivan_st_ivanov@yahoo.com

**Abstract.** The paper reviews Cloud-Based Information Systems and Client-Server Systems in comparative way. Components of cloud-based information systems are shown. As the desktop or web-based client-server systems, cloud-based information systems have their components with specific functions. Three components are discussed – the client application, the net and the cloud data center.

**Keywords.** Cloud-based information systems, Components of the information systems.

## 1.Cloud-Based Information Systems and Client-Server Systems – comparative analysis

Cloud-based information systems are distributed systems as some client-server systems. Data processing in cloud computing systems however is not the same as the data processing in well-known client-server systems (see pic. 1).

In Cloud computing systems on the client computer are performed a few operations and the heavy work is done remotely in huge shared data centre.

Cloud computing systems are differed from the traditional client-server systems not only from the place where the main part of work is being done but from the technology of data processing.

Cloud computing systems are more powerful than distributed client-server systems, but in different way. While in recent client-server systems have become much faster by using parallel processors, the cloud computing systems rely on gathering together the large number of servers with average speed, huge memory and enormous disk space. In cloud-based information systems one big task – application and data - is divided on small pieces. These pieces, each consist of application and data, are spread between machines in the data centre, the data are processed and then one answer is formed from the pieces and the answer is returned to the client application. So applications and data are situated in the cloud, but they are coded in a way that they could be divided into pieces.

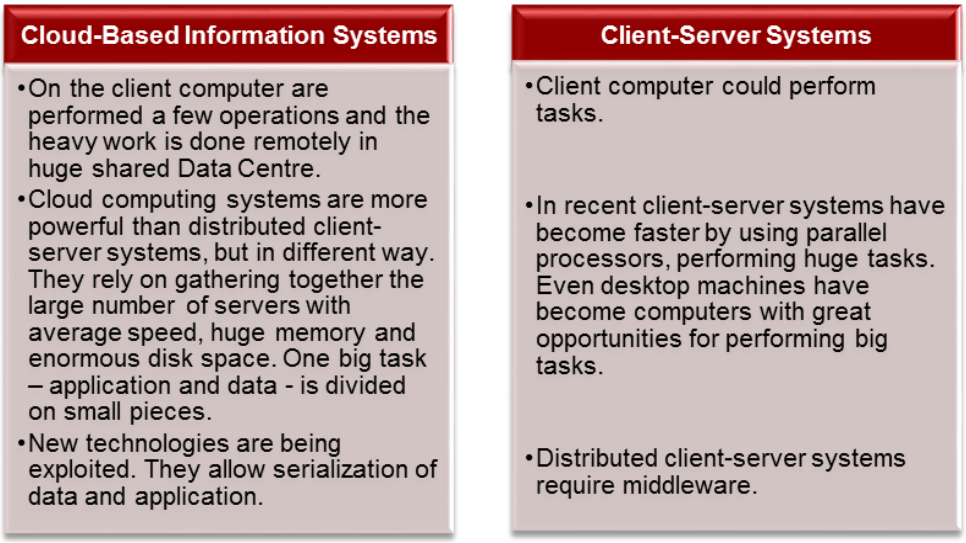
There are many new technologies that allow serialization of data and application. One of these technologies is transactional approach – serialized model of database. Typical for this approach is that it defines an isolation level of concurrent transactions to the database.

To summarize characteristics of cloud systems is used a methodological principle called BASE. BASE consist of three components Basic Availability, Soft State, Eventual Consistency:

• Basic Availability. The system is available in general (but not always!). There are actual data even if there are multiple failures. To be achieved this is used an approach of data replication and distribution between databases. Data is replicated and distributed over many systems for data storage. In the event of failure, the interruption affects access to a small segment of data, but does not lead to destruction of the entire database.

• Soft State. Last state of data could not be recorded, but in case of failure, previous data state can be restored and the work is not be interrupted.

• Eventual Consistency. At some time in the future database can be entirely consistent, but consistency is not it permanent condition. No guarantees when it will have happened. The transactional approach is used to ensure that simultaneous operations of the two client applications on the same data is possible without affecting consistency of the database.




Pic.1. Comparative analysis

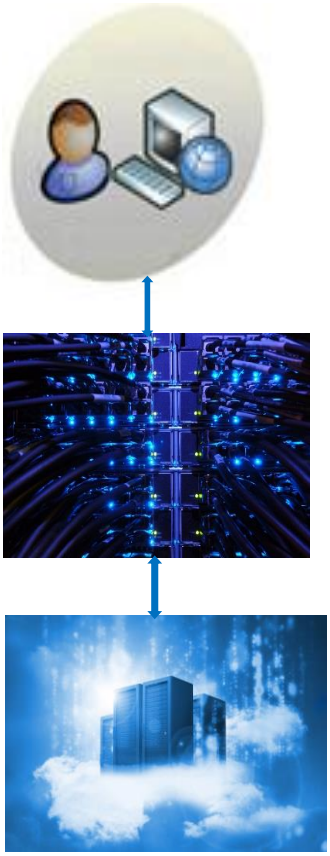
## 2.Components of Cloud-Based Information Systems

As the desktop or web-based client-server systems, cloud-based information systems have their components with specific functions. Three components are discussed here – the client application, the net and the cloud data centre (see pic. 2).

The first component is a client application. It could be a browser or a client program using same protocols as browser. The client application task is to send requests to cloud-based services and to receive responses from them. These requests and responses are usually carried out through a web page.

The second component is the network itself. The network is moving data physically from one point to another. The network represents a complex set of routers, DNS servers and servers of the different protocols.

The third component is the cloud data center and the informational system based on it. Data center aims to control the routing of data from the client computer to the cloud input point. After analyzing data, this entry point sends it to the specific cloud server. This server in communication with the other servers, builds the response which the client computer receives. So the application is a client to a particular server from the cloud, however, this particular server is a client of other servers and services.



Pic.2. Components

The communication between three components of the cloud-based information systems is complicated and includes a great number of hardware and software resources.

The cloud itself consist of some tiers and has its own architecture.

## 3.Conclusion

Stated in the report reveals the most common components of cloud-based information systems. Knowledge of the components and relationship between them helps to understand how cloud-based information system works from the user point of view. The client-server architecture is dominant until now, but it is quickly replaced with the architecture of cloud-based applications. Revealing the components of cloud-based information systems is only first step of understanding the architecture of cloud systems.

## References

1. Birman, K. P. Guide to Reliable Distributed Systems, Text in Computer Science. Springer-Verlag London Limited. 2012.
2. Lazarova.V. Classical web-based client-server systems and the ACID principles of the transactional approach, ICAICTSEE, 6-7 December 2013, Sofia, UNWE.

# Management of Cybersecurity Computer Assisted Exercises

**Irena Nikolova[1], Nikolay Tomov[2]**

[1] Space Research and Technologies Institute – BAS, Sofia, Bulgaria
irena.nikolova@mail.space.bas.bg

[2] Bulgarian Modeling and Simulation Association – BULSIM, Sofia, Bulgaria
nikolay.tomov@bulsim.org

**Abstract:** The paper presents an approach for managing a computer assisted exercises in cybersecurity – the environment development, execution and analysis of the exercise. The specific features of a typical cyber CAX are described and a unified organizational chart for managing a project for a cyber CAX based on the best project management practices is depicted.

**Keywords**: Computer assisted exercises, Cybersecurity, project management

## 1.Introduction

Nowadays nations are increasingly dependent on information and communication technologies. In light of this dependence the importance of the cybersecurity is growing with high rates and cyberspace elements become vulnerable to an expanding range of attacks by various groups of hackers, criminals and terrorists. The mostly spread cyber threats for the government agencies and private sector companies are thefts of sensitive information, cyber vandalism, phishing, POS and DDoS attacks. The nation's critical infrastructure, including the electric power grid, financial systems, social systems, air traffic and railways control systems, and communication networks, depends extensively on information technology for its operation.

In the recent years as a very powerful approach for verification of the preparedness of the responsible individuals and bodies in case of information security breaches in the organizations and the level of capabilities for mitigating the risks of cyber incidents occurring, the Computer Assisted eXercises (CAX) are used. In its nature CAX is a "synthetic" exercise, where the forces and resources are generated, operated and managed in simulation environment [1].

CAX is one of the valuable tools for enhancing the effectiveness of trainings and achieving interoperability at lower prices and decreased risks, in comparison to field exercises. CAX is a validated tool for testing doctrinal documents implementing the gained experience and lessons learned in military missions. CAX are specifically applicable in situations related to crises management, where the "recreation" of similar situations in real environment is too complex, economically ineffective and sometimes infeasible.

The experimentations in the cybersecurity domain in real operational environment are usually impossible. In case of any feasibility – the risks for interrupting the business processes are much higher, which can lead to considerable financial losses for the organizations.

Most effective from resources point of view, as well for achievement of desirable results in a short time frame by analyzing various alternatives, the means of the computer modeling, emulation and simulation are applied. In parallel, the human responses in case of

information security breaches could not be "programmed" entirely – a major factor is the decision making of the responsible authorities from different hierarchical organizational levels.

Considering the important role of CAX for enhancing the training capabilities in the security sector, a deficiency of unified CAX management for developing specific CAX environments has been identified. Such environment needs to form organizational system of concepts, approaches, methodologies, tools and means for effective realization of processes and its improvement. Design, development and utilization of unified CAX environment is a multidiscipline systematic approach for CAX management and afford opportunity for planning, execution and results assessment of CAX projects in the cybersecurity domain.

Below presented approach for managing the life cycle of a computer assisted exercise in cybersecurity has been used in two major exercises:

- „CYBERWINTER 2011" – first ever for the Bulgarian government administration computer assisted exercise in cybersecurity that was held in the Ministry of Transport, Information Technologies and Communications [2];
- „CYBERNRA 2014" – the first information security computer assisted exercise for the Bulgarian National Revenue Agency [3].

## 2.Cybersecurity Computer Assisted Exercise Concept

The Computer Assisted Exercise is an effective tool for individual and collective training for the achievement of a certain level of knowledge for an efficient response in case of information security breaches. The CAX reduces the risk level, enable time jumps and repeated scenarios for a short time interval, as well enable scenario simulation which practical implementation is very difficult with unexpected negative consequences.

The preparation and execution of CAX in cybersecurity (CyberCAX) differs from other domain specific computer assisted exercises as for crises management, disaster response, military operations etc., mainly in the scenarios, used software tools for modeling and simulation of network communications, devices emulation, virtualization of servers and applications and the targeted training audience. The phases of planning, execution and analyses are similar as in the other domain specific CAX.

Conducting a CyberCAX in any organization usually aims to validate the specific measures that should be undertaken or procedures that have to be followed during or after an information security incident. These measures include collaboration and coordination between the organizational units for improvement of them, as well some important interdependencies to be exposed in case of response to cyber threats and cyber incidents.

The main goals of a CyberCAX could be summarized as follows:

- Enhancing the capabilities of the organizations for protection of the information infrastructure against cyber-threats and information security breaches;
- Identification of organizational and technical vulnerabilities in the system for information security management, in the procedures and policies for responding in case of occurrence of cyber incidents;
- Analysing and foreseeing the new cybersecurity threats for the organization as a whole.

The advantages of CyberCAX find expression in the experimentation and validation of strategies, concepts, policies and procedures for information security before their practical implementation.

CyberCAX enable to verify the preparedness of the responsible expert staff to:

- Identify the occurrence of an incident;
- Do first response actions trying to mitigate data loss or other irreversible damages in the information systems, as well to close the system gap if that is the reason for the incident occurrence;
- Make damage assesment and to identify all the incident details;
- Prepare a recovery plan;
- Recover the hacked system.

CyberCAX are beneficial for the high level authorities responsible for making decisions when an incident occurs.

- To observe in real-time the operations for recovery after incidents;
- To determine the level of information security in the organization;
- To identify the weakness in the procedure for mitigation and response related to cybersecurity incidents;
- Reasons to assign tasks for improving the plans and procedures for mitigation and response;
- Evaluation of improvements.

CAX players are various depending on the type of the exercise, ranging from expert levels (system administrators, information security officers, experts and administration staff) to managing levels (head of department, director, CEO etc.).

## 3.CyberCAX Project Management

The achievement of the objectives of the CyberCAX project is ensured by good project management, which defines the exercise goals, organises the work packages and assigns the workload to highly motivated, qualified and experienced teams.

The qualitative project management rules and procedures ensure good coordination and mitigate or resolve potential conflicts that could arise within and between the working teams. Efficient communication and information flows guarantee the transparency and control of the work progress, its quality and associated risks, as well changes and the integration of specific activities.

The application of project management best practices in CyberCAX is a base for a unified, systematic, multidisciplinary approach for effective planning, execution and evaluation of CAX that provides tools and methods for an efficient management of the entire CyberCAX life cycle.

The CyberCAX management structure is laid out on the figure bellow. It shows the involved CAX units and their interactions. The structure is arranged in four levels:

- Strategic Level, where strategic decisions are taken;
- Management Level, where operative decisions are taken, task allocations is done, risk is estimated and reporting to the Strategic Level is provided;
- Operational Level, where the tasks are allocated, work packages are executed and the group efforts are joined;
- End Users Level that comprises EXCON Level and Training Level, where the CyberCAX players are involved and benefit from the results achieved by the upper levels.

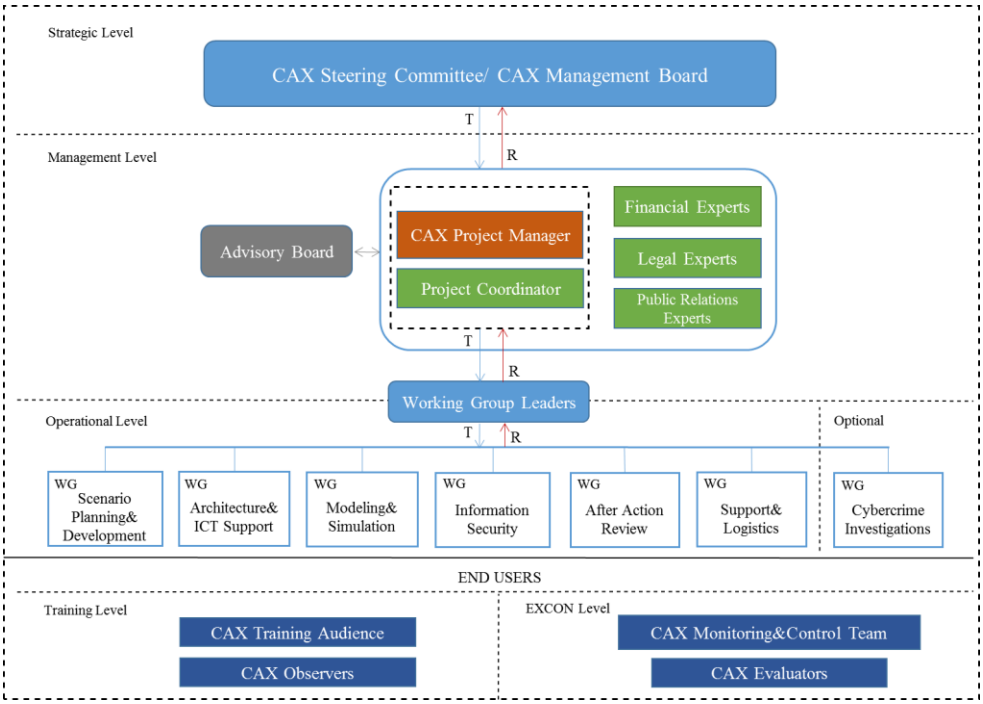



Fig. Organizational Chart for a CyberCAX Management.

The management structure could be tailored to the project work plan for each CyberCAX and to any specific requirements and team expertise, in order to provide successful CAX implementation. The presented approach, based on approved management procedures and techniques, has been successfully implemented in two completed CyberCAXes.

Each level of the organizational chart for a CyberCAX management comprises different units as follows:

- **CAX Steering Committee/CAX Management Board** – it is the strategic level decisions making body.

The CAX Steering Committee (CAXSC) or CAX Management Board (CAXMB) is composed of nominated representatives (or their designated deputies) from the Sponsors, Stakeholders and Principals, the Project Manager and other members, suggested by the Stakeholders and approved by the Committee. It forms the communication platform concerning the management issues of the work plan implementation and is in strong interaction with the Management Team.

The CAXSC take decisions on the following matters, but not limited only to them:

- Strategic orientation of the project;
- Approves or rejects the project's deliverables;
- Manages scope;
- Manages costs;
- Arranges funding;
- Manages project operational and political issues and risks;

- Coordinates with related projects and programs;
- The dissemination of the project and its results.

- **Advisory Board** - formed by experts and on demand representatives of the user organizations. Advisory Board should assist and guide the project activities towards a successful implementation.

Advisory Board is an external group of experts with specific areas of expertise related to the CyberCAX project, ranging from experienced cyber and IT experts to policy high-level experts. The utilization of such expertise provides additional quality assurance in the form of high-level reflections and guidance for the CyberCAX and its activities.

- **Management Team** is the operational management level.

The Management Team comprises a Project Manager and a Project Coordinator (in some cases) assisted by financial experts, legal experts and public relations experts. The Management team is responsible for the day-to-day administration, coordination, monitoring and control of the project.

**CAX Project Manager** is the individual who has the overall responsibility for the successful initiation, planning, design, execution, monitoring, controlling and closure of the project. In his/her key duties is the recognition of the risks that directly could impact the project success.

The Project Manager is responsible for:

- The successful realization of the entire project;
- Applying relevant project management processes;
- Leading the project team.

**Project Coordinator** is an integral member of the project team supporting the projects delivery of varying size and complexity. The Project Coordinator is responsible for organizing and controlling project activities under the guidance of the CAX Project Manager.

- **Work group (WG)** – work group team members execute the project tasks assigned by the CAX Project Manager, the Project Coordinator and the Work Group Leader.

**Work Group Leader** is responsible for the completion of the work package tasks, quality of the deliverables, and managing the dependencies by the other tasks.

The Work Group Leader's obligations include:

- Organization and conduct of regular workgroup meetings;
- Demonstration and reporting of the progress on workgroup tasks;
- Collaborate with the other Work Group Leaders;
- Provide feedback and reports to the CAX Project Manager;
- Maintaining of open communication with the CAX Project Manager;
- Participation in the review process.

- **End Users** are the main consumers of the project deliverables. They are not directly hierarchically dependent by the other levels.

**CAX Training Audience** comprises the exercise participants (individuals or teams) responsible for taking whatever actions are necessary to respond to simulated cyber incidents.

**CAX Evaluators** team consists of evaluators trained to observe and record participant actions. These individuals should be familiar with the exercising jurisdiction's plans,

policies, procedures, and agreements. This is individual who has no role to play in the exercise but witnesses events either to assess the preparations of the organization or individuals within it.

**CAX Observers** are exercise participants who play no role in the exercise, but audit the events either to assess the preparations of the organization or its individuals, or to learn lessons.

**CAX Monitoring&Control Team** comprises representatives from the Sponsors, Stakeholders, CAX Project Manager and additional experts who has to monitor and control the exercise to achieve its objectives.

The unified project management approach for CyberCAX management provides a framework for the managing and operational teams to make responsible decisions in alignment with the CAX project objectives, Meantime, it ensures a „preventive barrier" within a set of constraints and procedures, which guarantee that the CyberCAX project will not come out of the predefined plan and will achieve its goals [4].

## 4.Conclusions

The entire process of the environment development and conducting computer assisted exercise in cybersecurity is a large scale multidisciplinary task that requires adequate management based on the best practices, methodologies and management approaches. Ensuring high effectiveness, transparency and attaining the desired objectives in the planning and execution of the exercise could be achieved by applying unified approach for CAX management. The implementation of a CyberCAX in any organization is a key approach for having a "snapshot" of the relevance of the existing information security policies and procedures, thus building a confidence in the responsible and management staff that the organizational units are prepared to face the current cybersecurity threats.

The project management approach for CyberCAX management brings a high practical value and its implementation controls the quality of the verification and validation of concepts, operational procedures, current practices for information security assurance, as well develops capabilities for mitigating the cyber risk and increase the response efficiency in case of cyber incident occurrence.

Applying the project management methodologies and tools within the entire process for CyberCAX management is a key approach for enhancing the effectiveness, for directing the management to the results and providing transparency and control. A key factor for success in the CyberCAX management is the clear understanding of the final goals and the active and open communication for them between all stakeholders and participants involved.

## References

1. Cayirci, E.; Marincic D., Computer Assisted Exercise and Training, 2009
2. Bulgarian Ministry of Transport, Information Technologies and Communications, „CYBERWINTER 2011" Exercise - https://www.mtitc.government.bg/page.php?category=583
3. Bulgarian National Revenue Agency, "CYBERNRA 2014" Exercise - http://nap.bg/news?id=1887
4. Shalamanov V., Nikolova I., Tomov N., "Project Portfolio Management of Operational Analyses for the Transformation of the Security Sector", In Proceedings of the Annual Int. Conf. on Economic and Management of Security and Defense "Private Investment Initiatives for Armaments", Sofia, Bulgaria, UNWE, June 23-24 2005, 188-194.

# Modified Bayesian Method for Inference in Expert Systems

**Ilko Velikov**

UNWE, 1700 Sofia, Bulgaria
ivelikov@unwe.bg

**Abstract.** The paper reviews the basic principles of the Bayesian inference engine in expert systems. Pointed out are some of the shortcomings of the mechanism, regarding the quality of the used expert knowledge. Proposed is a based on interval computations modification of the Bayesian engine.

**Keywords.** Artificial intelligence, expert system, Bayesian inference

## 1.Bayesian Inference Mechanism

The inference mechanism we are using is based on a study by Thomas Bayes in the middle of the XVIII century. His approach is based on the assumption that practically for every hypothesis there is some prior beliefs (or probability) to be true. This probability can be very small, close to zero, but this doesn't influence over its usefulness for calculations. Therefore, we could assume that if there is some prior belief in a given hypothesis, then there would be data which shall be used in favor or against the given hypothesis. If there was not such data, the prior belief would stay unchanged under any circumstances and the hypothesis would be actually an axiom, thus not be interesting for any kind of analysis. On the other hand, by using the evidences (facts) related to the hypothesis, it is possible to modify the prior belief in order to get another, posterior probability, taking these evidences into account. According to the theory of the probabilities, for each hypothesis H there is a probability P(H). This is a prior probability of the hypothesis in case there are no proofs in favor of it or against it. Traditionally the interpretation of P(H) is frequency of the event H (materializing the hypothesis H) in series of random trials. In this case P(H) is called „objective probability" and its calculating is what statistics does. When P(H) is used to measure the degree of reliability, it is called „subjective probability" and most often it is gathered by interviewing experts. In both cases we used the same calculations for combining probabilities and the interpretation of the frequency coefficient we use for estimating the subjective probability and vice versa. This subjective probability (degree of belief) will change when receiving new information or by changing already existing information. Skipping al the intermediate transformations, the core Bayesian formulas we are using are:

(1) $$P(H|E) = \frac{P(H|E)\,P(H)}{P(E)}$$

and

(2) $$P(E) = P(E|H)\,P(H) + P(E|\overline{H})\,P(\overline{H})$$

The possible interpretation of formula (1) is that the likelihood of hypothesis H to be proven after providing the evidence E can be calculated by multiplication of its prior probability P(H|E) and the probability P(E|H) that E is going to be observed, considering that

H is also true. With this interpretation we can set aside from the denominator in formula (1) because it is simply a constant which can be easily calculated according to formula (2).

The core of the rule in formula (1) can be further clarified, using odds and likelihood ratios. If we divide formula (1) to $P(\bar{H})$, we receive:

(3) $$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(E|H)}{P(E|\bar{H})} \frac{P(H)}{P(\bar{H})}$$

Defining the odds of hypothesys H as

(4) $$O(H) = \frac{P(H)}{P(\bar{H})} = \frac{P(H)}{1-P(H)}$$

and the likelihood ratio as:

(5) $$L(E|H) = \frac{P(E|H)}{P(E|\bar{H})}$$

the posterior odds:

(6) $$O(H|E) = \frac{P(H|E)}{P(H|\bar{E})}$$

is derived from

(7) $$O(H|E) = L(E|H).O(H)$$

This way the Bayes formula says that the strength in the confidence in hypothesis H, based on the existing knowledge K and after providing the evidence E should be calculated as a result from the multiplication of two components: the prior odds O(H) and the likelihood ratio L(E|H). The first factor with the estimation of this confidence expresses future support for hypothesis H by the existing knowledge K. The second factor expresses retrospective support for H by already observed event E.

The new posterior probability can be used as a prior for new iteration of calculations, this time using a new proof in favor of the hypothesis until reaching a certain threshold for its admittance or its rejection.

The usage of the described inference mechanism requires appropriate structure of the knowledge base. There has to be a list in it with all hypothesis that are subjects to proving, their prior probabilities P(H), also sets of evidences, associated with each of the hypothesis, for which we know the likelihood ratios L(E|H). For the hypothesis from the last level it is necessary to store also thresholds of acceptance and rejection.

Although Bayes method is popular and proven, there are couple of notes we can make for it. Perhaps the most essential one is that with this method independence of evidences is supposed to exist. If two or more evidences in the knowledge base have some correlation in between, then it is evidently that at least one of them will affect secretly the result from the calculations which will lead to wrong posterior probabilities. At this stage the problem with the correlation of proofs has no theoretically grounded solution. This correlation could be reduced to acceptable levels by careful knowledge base design.

Another important problem with Bayes method is the one with formalizing knowledge which is used in it. There need to be evaluations for probabilities which is hardly understandable in some fields by the experts in those fields; moreover, the given opinions are not easy for quantification. In our previous work we suggested providing the experts with the opportunity to give interval evaluations. This increases the reliability of the expert opinions as compared to the precise value of the evaluation, where there is authenticity with complete concurrence, here when we work with intervals it is enough if the objective value is between

the upper and the bottom border. Of course, the given intervals evaluations of multitude of experts have to be generalized and only then they can be put in the knowledge base.

## 2.Expertons Theory

In this paper we will not review the usage in inference process of measurable facts – it is obvious that determined value of a parameter does not need extra treatment to be used. It is not the same however when using of undetermined, more often semantic data. Such data resists qualitative measuring and the most suitable way to turn it in knowledge of an expert system is through using expert evaluations.

If we accept that expert opinion is the main source of knowledge with undetermined data, then its formalization is mandatory precondition for making it usable in the inference. Usually the expert opinion is expressed in giving an evaluation according to nominal scale. Most often used scales consist of 3, 5, 7, 9 and 11 stages, where there are no any objective limitations – evaluating in a certain point of the scale depends only on the decisions made by the designer of the knowledge base.

Generally, evaluations of only one expert cannot be accepted as good enough for achieving authenticity of knowledge in the expert system when of course this does not mean that its reliability is in any kind of a doubt. There are cases where worldly practice shows that the view of one specialist in certain area is more trustworthy that the view of the majority of his/her colleagues. However, these are particular cases which are more of exceptions from the rule. And the rule is that assessment established by a multitude of experts is more reliable and more authentic than one established by only one expert. On the other hand, the aggregation of big amount of evaluations is obligatory in the formalization of knowledge and it has been issued as a problem for decades. Below we are going to review an approach for summarizing of expert evaluations, submitted in the 1980s by professor Arnold Kaufman from the University „Stendal“ in Grenoble. This approach is based on the theories of fuzzy sets and confidential intervals and it is called „Expertons Theory“ by its author.

Prof. Kaufman examined a case where based on the evaluations of number of experts, an interval of probabilities is found for each level $\alpha \in (0,1)$. It is well known that an interval of probabilities is not a probability. On the other hand, after generalization of experts‘ views there might be an aggregated interval of their assessments which later on might be used in calculations of probabilities.

We are examining the example with assessments given by ten different experts for the value of a chosen parameter, used in an inference process. For example, this could be the prior probability for the occurring of the event H. Using reliable values for this probability is significantly important for the effectiveness of the calculations with Bayesian method. The exact number of experts is chosen in a way to facilitate the calculations but at the same time it is valid for n-experts.

Each one of the ten experts offers confidential interval (a1, a2) in which it is the searched prior probability, using, for example scale between 1 and 11, where $(a1,a2) \in (0,1)$ and a1 might be equal to a2. The following evaluations are supposed to be given:

(8)

| Expert | Mark |
|---|---|
| 1 | (0.2, 0.5) |
| 2 | (0.5, 0.7) |
| 3 | (0.6, 0.8) |
| 4 | (0.3, 0.6) |
| 5 | 0.6 |
| 6 | (0.0, 1.0) |

| | |
|---|---|
| 7 | 0.8 |
| 8 | 0.3 |
| 9 | (0.0, 0.1) |
| 10 | (0.8, 0.9) |

The response (0.0, 1.0) is generally „no willing to answer“, unlike 0.5 which needs to be interpreted as „cannot evaluate“.

Next thing that needs to be done is taking count of the frequency of different evaluations separately for upper and lower limits of the intervals and by dividing them on the number of the experts we can calculate their frequencies.

(9)

| Low | | High | |
|---|---|---|---|
| Level | Frequency | Level | Frequency |
| 0.0 | 0.2 | 0.0 | 0.0 |
| 0.1 | 0.0 | 0.1 | 0.1 |
| 0.2 | 0.1 | 0.2 | 0.0 |
| 0.3 | 0.2 | 0.3 | 0.1 |
| 0.4 | 0.0 | 0.4 | 0.0 |
| 0.5 | 0.1 | 0.5 | 0.1 |
| 0.6 | 0.2 | 0.6 | 0.2 |
| 0.7 | 0.0 | 0.7 | 0.1 |
| 0.8 | 0.2 | 0.8 | 0.2 |
| 0.9 | 0.0 | 0.9 | 0.1 |
| 1.0 | 0.0 | 1.0 | 0.1 |

Using these frequencies, we could find the cumulative fequences

(10)

| Level | Cumulative Frequency |
|---|---|
| 0.0 | (1.0, 1.0) |
| 0.1 | (0.8, 1.0) |
| 0.2 | (0.8, 0.9) |
| 0.3 | (0.7, 0.9) |
| 0.4 | (0.5, 0.8) |
| 0.5 | (0.5, 0.8) |
| 0.6 | (0.4, 0.7) |
| 0.7 | (0.2, 0.5) |
| 0.8 | (0.2, 0.4) |
| 0.9 | (0.0, 0.2) |
| 1.0 | (0.0, 0.1) |

This set of intervals is called “experton”

The experton we established is used for calculating of mathematical expectation. It is received by addition of the lower and respectively the upper values of different levels in the experton except for level 0 and division of the result to 10. In this way from experton (10) we can receive the following mathematical expectation:

((0.8, 1.0) + (0.8, 0.9) + (0.7, 0.9) + (0.5, 0.8) + (0.5, 0.8) + (0.4, 0.7) + (0.2, 0.5) + (0.2, 0.4) + (0.0, 0.2) + (0.0, 0.1)) / 10 = (4.1, 6.3) / 10 = (0.41, 0.63)

This result is aggregated opinion of the experts for the prior probability P(H). In order to be clear that this presumptive probability is interval number, we will accept the notation $[P(H),PR(H)]$.

Similar survey has to be done also for the values of $(Hi)$ for every hypothesis in the system; also for the likelihood ratios $L(Ei,j|Hi)$ for each evidence. When we apply the expertons method to the results from the survey, respective substances that need to be stored in the knowledge base have to be the corresponding intervals $[P(Hi),PR(Hi)]$ and $[LL(Ei,j|Hi),LR(Ei,j|Hi)]$.

## 3.Modified Bayesian Method for Inference

These intervals can be used in modified formulas (4) and (5) in order to receive a new formula for calculation of the posterior probability P(H|E). This new formula will operates with intervals, not with numbers.

Arithmetic operations with intervals are defined like this:

Addition: [a, b] + [c, d] = [a + c, b + d]

Subtraction: [a, b] – [c, d] = [a – c, b – d]

Multiplication: [a, b] * [c, d] = [min(ac, ad, bc, bd), max(ac, ad, bc, bd)]

Division: [a, b] / [c, d] = [min(a/c, a/d, b/c, b/d), max(a/c, a/d, b/c, b/d)]

It needs to be acknowledged that the division is defined only when none of the values of the interval are null.

Modified formula for odds calculations as an interval looks like this:

$$(11)\qquad [O_L(H), O_R(H)] = \frac{[P_L(H),P_R(H)]}{1-[P_L(H),P_R(H)]} = \frac{[P_L(H),P_R(H)]}{[1-P_L(H),1-P_R(H)]} =$$

$$= \left[\begin{matrix}\min\left(\frac{P_L(H)}{1-P_L(H)}, \frac{P_L(H)}{1-P_R(H)}, \frac{P_R(H)}{1-P_L(H)}, \frac{P_R(H)}{1-P_R(H)}\right), \\ \max\left(\frac{P_L(H)}{1-P_L(H)}, \frac{P_L(H)}{1-P_R(H)}, \frac{P_R(H)}{1-P_L(H)}, \frac{P_R(H)}{1-P_R(H)}\right)\end{matrix}\right]$$

and the posteriori chance is transformed similarly and because of its complexity we will not be examining it in this paper.

Modified formula for likelihood ratio calculations as an interval looks like this:

$$(12)\qquad [L_L(E|H), L_R(E|H)] = \frac{[P_L(E|H),P_R(E|H)]}{[P_L(E|\bar{H}),P_R(E|\bar{H})]} =$$

$$= \left[\min\left(\frac{P_L(E|H)}{P_L(E|\bar{H})}, \frac{P_L(L|H)}{P_R(E|\bar{H})}, \frac{P_R(E|H)}{P_L(E|\bar{H})}, \frac{P_R(E|H)}{P_R(E|\bar{H})}\right), \max\left(\frac{P_L(E|H)}{P_L(E|\bar{H})}, \frac{P_L(L|H)}{P_R(E|\bar{H})}, \frac{P_R(E|H)}{P_L(E|\bar{H})}, \frac{P_R(E|H)}{P_R(E|\bar{H})}\right)\right]$$

And finally, the formula for calculation of the posterior probability of the hypothesis H, after providing the evidence E is:

(13) $$[O_L(H|E), O_R(H|E)] = [O_L(H), O_R(H)] * [L_L(E|H), L_R(H|E)] =$$

$$= [\min(O_L(H) * L_L(E|H), O_L(H) * L_R(E|H), O_R(H) * L_L(E|H), O_R(H) * L_R(E|H)),$$

$$\max(O_L(H) * L_L(E|H), O_L(H) * L_R(E|H), O_R(H) * L_L(E|H), O_R(H) * L_R(E|H))]$$

## 4.Conclusion

Bayesian inference mechanism is practically proven as up to standard instrument for processing knowledge in expert systems. However, in order for it to be used, it is necessary to register facts in the knowledge base that are related to probabilities of evidences in favor of hypotheses. If these probabilities are given by a big number of experts and moreover if they are taken as interval evaluations, then it is possible to achieve good reliability and authenticity of the registered knowledge. This way the suggested modified Bayesian inference mechanism will be generally more effective and it will allow giving more adequate evaluations of hypotheses that are checked with it. The proposed solution requires modifying of the way we calculate prior probabilities and this is realized by interval calculations, which is the core of this work. We came to the conclusion that this given composite method could make the Bayesian inference more reliable and useful.

## References

1. Savage, L. J., Elicitation of personal probabilities and expectations, Journal of American Statistical Association, pp. 783 – 801, 1971.
2. Buchnan, B.G., R. O. Duda, Advances in Computers, vol. 22, M. Jovits (ed.), London, Academic Press, pp. 163 – 216, 1983.
3. Pearl, Judea, “Bayesian Decision Methods” in Encyclopedia of AI, Wiley Interscience, 1987.The most significant consideration is that these quantitative models had to be built on the basis of information available to the organization. This is information that is collected on a daily basis in corporate databases as a result of current activities.

# Advances in Revision Control Systems

**Bogomil Traikov, Angelin Lalev**

D.A.Tsenov Academy of Economics, Em. Tsakarov Str. 2, Svishtov, Bulgaria
traikov@uni-svishtov.bg, lalev@uni-svishtov.bg

**Abstract.** The paper reviews the recent advances and paradigm shifts in the field of revision control systems. Although revision control systems are known for their relatively mature feature sets, they too, are affected by the ongoing changes in the way collaborative software development is evolving towards open, more loosely-coupled, Internet and cloud-facilitated models. This evolution, inspired and spear-headed by the open-source community, is not the only source of new ideas in the field, problems of even more theoretical appeal are also being solved, leading the way for new generation of revision control systems.

**Keywords:** Revision control systems, binary files, collaborative software development

Revision control is indispensable part of every software project and unsurprisingly, the software that addresses the problems of revision control has been envisioned and developed since 1970s. For nearly 45-year period, the progression of ideas and solutions in the area has been immense, but only recently, there has been explosion of different solutions triggered by major paradigm shifts in the area. We review three different aspects of the changes. They are related to architecture, models of use and theoretical understanding of these systems.

**The first** and most visible aspect is the one, related to the way developers use the revision control systems. Historically, revision control systems have evolved from ones with no network functionality whatsoever towards ones which implement strongly centralized client-server model over a global communications network like Internet. In recent years though, some new developments part with the client-server model giving rise to strongly distributed revision control systems. This change is not uniform in the sense that not all of the users of the older generation systems are migrating or will ever migrate to distributed ones. Even more – there are new projects like CVSNT which are built around client-server architecture.

The usefulness of these new generation revision control systems is only visible when they are considered in relation with the development models they facilitate. In the relatively distant past (the 1970s and 1980s), the typical software projects were relatively small, employing up to several dozen people, working in close cooperation together. Most often, they worked on one machine (mainframe or minicomputer) and were not separated by serious physical distances.

The first generation revision control systems like SCCS [1] (1972) and RCS [2] (1982) evolved in this environment. They were created out of the need to keep and organize all older versions of source code of a software project. In these early years the main requirement for the revision control systems was efficient use of limited resources of the computer in terms of space and processor time. When branches of the software were created, they were created as measure of last resort, usually because different versions of the software

were required for different clients. Most of the time branches were not created with intention to be merged back into the main version of the software. Such merging, when necessary, was done manually and required enormous logistical efforts.

In this environment, the revision control systems turned to the most obvious and easy method for coordination between developers, editing the main source tree – the lock model. Lock model dictates that when a developer works on file, this file should be made unalterable to other developers, that is, it would be "locked". This prevents situations, where different developers have inconsistent versions of the file. Of course, situations when several developers tried to edit single file of the source were not uncommon and were resolved manually and 'offline'.

The 'iconic' implementation of such systems was CVS [3], which in the beginning of 1990s combined the lock model with network functionality. Although CVS supported merging of different files, its commits were tracked still at per-file basis and branching required manual tagging of the branching points for subsequent merges. CVS and the centralized revision control systems, influenced by it, became ubiquitous and their feature set remained more or less frozen for the next decade.

By the turn of the century, CVS and centralized revision control systems as general were still very fit for the needs of software development. This was beginning to change for some projects. First to feel the change was the open-source movement. Open-source development, facilitated by Internet, suddenly increased the number of involved developers by several orders of magnitude. Big open-source projects soon 'dwarfed' even the biggest commercial developer communities, which still adhered by the closed development models. In the same time, the resources available to these projects, as well as geographical distances necessitating 'offline' style of communications, greatly increased the need for greater automation of all coordination activities, related to accessing the source trees.

By 2000s, big open-source projects like FreeBSD and X Window System had several hundreds of developers which contributed to the project regularly. The biggest of them by that time was undoubtedly Linux, which had by that time several thousand regularly contributing developers.

The model these projects adopted to cope with constantly flowing contributions was necessitated exactly by logistics of revision control. Most of these projects had 'core teams' of developers, whose responsibilities in great deal related to integrating other people's code into the project's source tree. Only core team members had write access to the CVS repositories, so to add a change to codebase, it had to be audited and committed by a member of the core team. In terms of organization, this made sense, because it prevented chaos in source trees. In many cases though, 'core teams' simply did not have the resources to integrate the changes fast enough. Entire projects like 386BSD and XFree86 were canceled because of perceived "slowness" and lack of flexibility related to integration of the proposed changes led to contributors abandoning them.

The Linux project had same problems, which were even more tangible, because of desire of Linux community to keep the core team small and centered on the project founder [4]. The solution they developed was, in essence, the distributed revision control systems. In sequence, the Linux community produced two of the most popular tools, used today - first BitKeeper and then Git. The main idea behind these tools is to automate branching and merging as much as possible, hopefully to the point that it allowed individual developers to check out a branch and work on it independently for long periods of time, allowing them and the core teams the much greater flexibility than older, centralized revision control systems did allow. Core features of these products included full duplication of the metainformation of the repository (eliminating the need for constantly referring to the one central repository),

checking out whole branches (versus the per-file approach of the centralized RCSes) and automated tagging.

Tracing the evolution of the distributed revision control systems and their features in relation to the development models they facilitate, makes clearly visible that their functionality offers less benefits to smaller projects. Projects that employ centralized development models and community of developers that communicate 'face to face' have much lesser benefits from employing distributed RCSes. Distributed revision control systems also make more difficult to control who has access to some particular part of source tree. On the other end of the spectrum, with ever increasing software complexity, big projects, including commercial ones, have much to win by switching to distributed RCS, because with the increase of the number of developers,

**The second** aspect, in which the evolution in RCS design makes progress, concerns the architecture and algorithms, used in such systems. Although the focus has shifted away from them, speed, as well as conservation of network bandwidth are still important features of RCS. Maximizing the speed and efficiency can be achieved by new algorithms and structures of data.

Indeed, a major difference between the newest revision control systems is the way they handle the data internally. Some of them, like Mercurial and BitKeeper are similar to the very first implementation of RCS in the way they handle the data. Files are stored in simple text file and the subsequent changes are inserted in the same file, denoted by special metainformation rows which act as nested logical braces, denoting the start of the change and the end of the change. This method is known as 'interleaved deltas'. The content, locked between the logical braces, may be interpreted as the additions, necessary to make older version of the file into newer one – so called 'deltas'.

Interleaved deltas have the advantage of simplicity and relatively fast processing. First historic implementations of RCS stored the data forwards in the sense, that the original file was stored with the deltas to make it any subsequent one. So checking out the latest version required application of all deltas to the original version. Great optimization of this process, done very early, was to switch processing 'backwards' – latest version of the file is stored with the deltas to make it any previous version up to the original one. Since the checkouts of the source trees more often concerned more recent versions of the file, this was great improvement, since only several deltas had to be applied to get the desired file.

More interest for the designers today represents the search for efficient way to compress the old deltas, so they take less space. Modern systems use constantly evolving heuristic algorithms for such purposes. Git for example, uses heuristics to seek several 'points' in the history of the file, which are similar to maximal count of ones that precede and succeed them in the history. This allows for efficient compression, using these points as base.

Distributed revision control systems that work on whole branches actually do provide good architectural synergy to delta compression techniques. They store each branch's files separately, but minimize duplication of files, by instead storing references to a file that is not modified between branches, so one version of the file is stored only one for all branches. This allows compression to achieve maximal elimination of duplicate information.

**Third aspect** of the changes involves advances in theoretical understanding of the algorithms, with relation to revison control systems. The most important in algorithmical perspective changes in RCS design concern the algorithms for discovering the differences between files and merging them automatically into one consolidated version of the file.

Around the end of the 1970s, the modern version of **diff** emerged from earlier heuristic designs. Diff treats lines of original document as atomic elements and given two

versions of the file, it outputs the differences between them. Doing this task in optimal fashion is quite complicated, because the changes may involve complex movement of some lines or group of lines upwards or downwards in the document. These movements may be interleaved by additions or deletions.

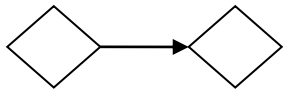
Diff solves the problem by finding longest common sequence of lines between the two files by the algorithm [5], described by the following recursively defined function.

$$S(X_i, Y_j) = \begin{cases} \emptyset \; if \; i = 0 \; or \; j = 0 \\ S(X_{i-1}, Y_{i-1}) \cap x_i \;\; if \; x_i = y_j \\ longer \; of \; S(X_i, Y_{j-1}) \; and \; S(X_{i-1}, Y_j) \; if \; x_i \neq y_j \end{cases}$$

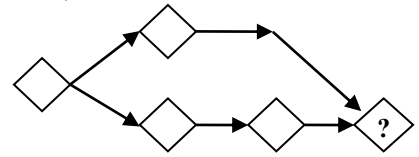
Where $X_i$ and $Y_j$ are the ordered lists of first $i$-th and $j$-th rows respectively from first and the second file and $x_i$ and $y_j$ are the $i$-th and $j$-th row of the first and the second file. In essence, the algorithm starts from the end of both files. If both ends contain the same line, it is in the longest common sequence. It is stored and the algorithms continue on both files with the last line removed. It the lines on the end of both files differ, the last line of the first file is removed and the algorithm is started recursively again. Same is done with the whole first file and the second file with the last line removed. Longer of both sequences, computed this way is the result of the algorithm.

Once found, the longest common sequence is ammended with the missing parts from the first and the second document. The parts of the first document are marked with 'meta-tags' to indicate deletion and the parts of the second document are marked with similar tags to indicate addition. In this fashion a 'patch' is created which exactly indicates what is the difference between the files and how to apply transformations to convert the first version into the second and vice versa.

This version of diff proved reliable contrasted to earlier heuristic designs of the same program. Without it, probably revision control systems would not become commonplace. By the beginning of 1990s a modification of diff was made available – so called **diff3**. Diff3 allows automated merging, which is pivotal for the development of modern distributed revison control systems. Given a file, edited independently by two developers, diif3 could create script that merges the changes into new file that integrates the changes in both files. For this purpose, diff3 needs additional information - the original version of the file (common ancestor of both files in the RCS tree).



a/ diff




b/ diff3

Diff3 applies classical diff twice and merges the results. First diff is applied on the pair, consisting of the original version of the file and the version of the first developer. Then diff is applied on the pair formed by the original version and the version of the second developer. The diff results are then integrated, using the lines that are fixed in both files to determine the boundaries of changed regions in both files. If the versions of the two developers differ in the same region, the conflict has to be resolved manually.

The diff3 algorithm actually relies on heuristics, based on some assumptions. The main one is the strong intuition that if two developers edit two different parts of the file,

separated by enough distance in terms of lines, their changes will not produce conflicts and would be obvious how to merge them, which diff3 will do.

In a relatively new result, dating from 2007 [6], this has been disproved trough formal analysis. This analysis proves that on certain conditions, diff3 will fail to merge two change sets, done in different parts of the file, no matter how big is the distance between them. The same analysis shows, that diff3 heuristic does not guarantee nor stability nor idempotence of the algorithm. That is, similar inputs may produce greatly different outputs on certain conditions. Even more, running the algorithm twice or more, using the previous run as input to the next, actually may fail to complete the resolution process, producing greatly different output. As a result, the search is on for non-heuristic algorithm to produce diff3 – a process that will be boosted by ever-increasing need for maximal automation of merging in revision control.

From the analysis, it is clear that there is a split in the needs of RCS software between smaller and bigger projects as well that there might be split between the projects that need to maintain greater secrecy of source code (some commercial products) and the ones, that are based on the open-source idea. It is yet to be seen if this will create split between the distributed revision control systems on one side and the centralized revision control systems on the other. It may be possible some of the distributed RCSes to be amended with features that allow they to be used in centralized fashion. In fact, such solutions need to address mainly the security implications of distributing the whole source tree of the project.

## References

1. https://en.wikipedia.org/wiki/Source_Code_Control_System
2. https://www.gnu.org/software/rcs/
3. https://en.wikipedia.org/wiki/Concurrent_Versions_System
4. Linux mailing list archives – http://lkml.iu.edu/hypermail/linux/kernel/9809.3/0957.html
5. Hunt, James W.;Mcllroy, M. Douglas (June 1976). An Algorithm for Differential File Comparsion, Computing Science Technical Report (Bell Labs)
6. Khanna, Sanjeev; Keshav, Kunal; Pierce, Benjamin C. (2007) A formal investigation of Diff3. Proceedings of 27th International Conference on Foundations of Software Technologies and Theoretical Computer Science.