UNIVERSITY OF NATIONAL AND WORLD ECONOMY

The Spirit Makes The Power

9TH INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION

ICAICTSEE – 2019

October 24 – 26th, 2019

University of National and World Economy

Sofia, Bulgaria

CONFERENCE PROCEEDINGS

UNDER THE AUSPICES OF:

ifip

the leading edge of information technology

PUBLICATIONS INDEXED BY:

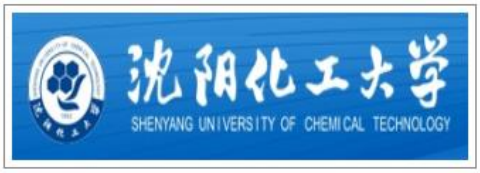
EBSCO HOST

AND

ProQuest

COORGANIZING UNIVERSITIES AND INSTITUTIONS

Shimane University, Japan



Shenyang University of Chemical Technology, China

University of Genoa, Italy

Ukrainian Federation of Informatics, Ukraine

National Changhua University of Education, Taiwan

University Danubius of Galati, Romania

Kharkiv National University of Radioelectronics, Ukraine

Vellore Institue of Technology, India

Tomar Politechnic Institute, Portugal

Volga State University of Technology, Russian Federation

COOPERATED BY

CREATIVITY AND HARMONY SOCIETY

PROCEEDINGS OF THE

9TH INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION

ICAICTSEE-2019

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: AUGUST 24TH, 2020

SIZE: JIS-B5

QUIRES: 21

PAGES: 336

VOLUME: 100

ISSN 2367-7635 (PRINT)

ISSN 2367-7643 (ONLINE)

PUBLISHING COMPLEX – UNWE, SOFIA, BULGARIA

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***
Gaetano **Cascini**	Politecnico di Milano	***Italy***
Olena **Chaikovska**	Kiev National University of Culture and Arts	***Ukraine***
Luisa **Chimenz**	University of Genoa	***Italy***
Sonia **Chipeva**	University of National and World Economy	***Bulgaria***
Sérgio Manuel Serra da **Cruz**	Universidade Federal Rural do Rio de Janeiro	***Brazil***
Tadeusz **Czachorski**	Institue of Theoretical and Applied Informatics, PAS	***Poland***
Linda **Daniela**	University of Latvia	***Latvia***
Vladimir **Dimitrov**	Sofia State University	***Bulgaria***
Marco **Ferrante**	University of Genoa	***Italy***
Marin **Galabov**	University of National and World Economy	***Bulgaria***
Carlos Arturo Torres **Gastelu**	Universidad Veracruzana	***Mexico***
Vasil **Georgiev**	Sofia State University	***Bulgaria***
Valentin **Goev**	University of National and World Economy	***Bulgaria***
Igor **Grebennik**	Kharkiv National University of Radio Electronics	***Ukraine***
Chih-Hsuan **Huang**	Hubei University of Economics	***China***
Aytekin **Isman**	Sakarya University	***Turkey***
Kalinka **Kaloyanova**	Sofia State University	***Bulgaria***
Madjid Malikovich **Karimov**	Tashkent University of Information Technologies	***Uzbekistan***
Sergii **Kavun**	Kharkiv University of Technology STEP	***Ukraine***
Hee Dong **Kim**	Hankuk University of Foreign Studies	***S.Korea***
Peter **Kokol**	University of Maribor	***Slovenia***
Özgen **Korkmaz**	Amasya University	***Turkey***
George **Kovacs**	Computer and Automation Institutre, NAS	***Hungary***
Oleksiy E. **Kovalenko**	Inst. of Mathematical Machines and Systems Problems, NASU	***Ukraine***
Patrick **Letouze**	Universidade Federal do Tocantins	***Brazil***
Violeta **Manevska**	St. Kliment Ohridski University - Bitola	***Macedonia***
Emanuel-Stefan **Marinescu**	Danubius University Galati	***Romania***
Liviu-Mihail **Marinescu**	Danubius University Galati	***Romania***
Célio Gonçalo Cardoso **Marques**	Polytechnic Institute of Tomar	***Portugal***

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***
Kaninda **Musumbu**	Universite Bordeaux	***France***
Tetsuo **Noda**	Shimane University	***Japan***
Florian **Nuta**	Danubius University Galati	***Romania***
Serghei **Ohrimenco**	Laboratory of Information Security, AESM	***Moldova***
Gustav **Olling**	IFIP TC5	***USA***
K.R. **Pardasani**	Maulana Azad National Institute of Technology	***India***
Monica **Pocora**	Danubius University Galati	***Romania***
David Nadler **Prata**	Universidade Federal do Tocantins	***Brazil***
Andy **Pusca**	Danubius University Galati	***Romania***
N.**Sharmila**	Vellore Institute of Technology	***India***
Yizhe **Shi**	Shenyang University of Chemical Technology	***China***
Yixun **Shi**	Bloomsburg University of Pennsylvania	***USA***
R.**Sivacoumar**	Vellore Institute of Technology	***India***
Rajagopal **Sivakumar**	Vellore Institute of Technology	***India***
Kamelia **Stefanova**	University of National and World Economy	***Bulgaria***
Tullio **Tanzi**	Institute Mines Telecom - Telecom ParisTech	***France***
A.F. **Verlan**	Pukhov Institute for Modelling in Energy Engineering, NASU	***Ukraine***
Hsin-Hung **Wu**	National Changhua University of Education	***Taiwan***
Parviz Ali **Zade**	OKAN University	***Turkey***
Plamena **Zlateva**	Institute of Robotics, Bulgarian Academy of Sciences	***Bulgaria***
Milan **Zorman**	University of Maribor	***Slovenia***
Zong **Xuejun**	Shenyang University of Chemical Technology	***China***

LOCAL ORGANIZING COMMITTEE

Emil **Denchev**	University of National and World Economy	***Bulgaria***
Rosen **Kirilov**	University of National and World Economy	***Bulgaria***
Katya **Kirilova**	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***
Veska **Mihova**	University of National and World Economy	***Bulgaria***
Alexandrina **Murgeva**	University of National and World Economy	***Bulgaria***
Mitko **Radoev**	University of National and World Economy	***Bulgaria***
Kostadin **Sheyretski**	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***

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 **9th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2019)**, http://icaictsee.unwe.bg/, which took place on October 24-26th, 2019 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 45 published research papers (12 from foreign participants), the explosion of fresh ideas and the establishment of new professional relations.

Dimiter G. Velev
ICAICTSEE-2019 Chair

CONTENTS

The 4th Industrial Revolution and Education in the Republic of Korea

James Kang

The University of Veliko Tarnovo St. Cyril and St. Methodius,
Veliko Tarnovo, Bulgaria
bulgariajames@gmail.com

Abstract. As the world entered the 21st century, the era of knowledge revolution 5.0 was reached. The speed of knowledge is rapidly progressing. Human life is changing at a fast pace with the Industrial Revolution 4.0, its 5G and AI, that symbolizes the knowledge revolution. At present times, the national competitiveness of each country depends on the level of knowledge accumulation. That's where the importance of education is greater than ever. The most crucial aspect in the era of knowledge revolution 5.0 is computer coding education and convergence education (STEAM). The Republic of Korea and other advanced countries have already introduced this education, have applied it to public education, and are conducting various research projects for innovative future education. In addition, in Korea a 4th Industrial Revolution Committee was newly established as an organization directly under the President to present a response and vision for the arrival of the Industrial Revolution 4.0, and to develop and utilize 5G and AI in various fields including education. The committee is presenting and preparing for the future. In particular, governments, corporations, and schools are jointly investing in the development of future schools for 5G and AI. In this paper, I am going to look into how Korean education has changed, the different changes that are pursued in the IR 4.0. Also, I will explore future possibilities by looking into the smart future education utilizing the development of computer advanced technologies.

Keywords: knowledge revolution, convergence education (STEAM), 5G, AI, 4th industrial revolution, computer coding, future schools.

1. Knowledge revolution 5.0.

Knowledge revolution 1.0 began with the human ability to think and communicate freely, as opposed to animals. Knowledge revolution 2.0 was manifested as the First Agricultural Revolution when humans relied on hunting and framing. Later on came the Scientific Revolution with the emergence of modern science and enlargement of the market, which is the 3.0 era. Knowledge Revolution 4.0 is the one leading to the Industrial Revolution where we find rapidly expanding mass production through new technologies and the steam engine. The Knowledge Revolution 5.0 refers to perspicacity and the changes in all fields where the shape of human life is drastically changed.

The Korean government plans to lead the fourth industrial revolution with the ICT Renaissance, a software powerhouse (2017). The main contents are (1) the establishment of the Presidential Committee on the Fourth Industrial Revolution (August 2017) and the establishment of a pan-ministerial plan to respond to the Fourth Industrial Revolution (Q317), (2) Investing heavily in R&D, human resources development of core intelligence information and implementing regulations to facilitate entry into ICT new technologies and services markets. (3) Establishing infrastructure for 5G and IoT networks, promoting data opening and

distribution, and discovering and spreading ICT convergence services such as smart home and precision medicine, Building IoT-only networks in 2017, commercializing 10GB Internet services in 2018, (4) commercializing 5G early in 2019, innovating software legal systems and public markets, strengthening human resources and technology capabilities, and implementing AI-based cyber security threats.

The Fourth Industrial Revolution Committee, in collaboration with 21 government ministries, established a "people-centered response plan to the Fourth Industrial Revolution for innovative growth". Under the plan is technology, with the vision of "implementing the people-centered Fourth Industrial Revolution". By closely linking industrial and social policies, the plan is to push forward strategic tasks in four areas: (1) intelligence; (2) innovation; (3) engagement; and (4) communication. It is called the policy brand "I-Korea 4.0."

"I-Korea 4.0" is not a simple technology development, but rather a package of "technology + infrastructure + diffusion + institutional improvement."The aim is to improve the quality of people's lives by simultaneously solving "structural tasks of economy and society" through the innovation of the overall intelligentization of society.

Since the announcement of the fourth industrial revolution response plan by the Ministry of Science and ICT, there have been many changes in Korean society [4 million AI speakers, 1 million 5G subscribers for the first time in the world, 7 times the number of AI programs (open API) use, 7,903 smart factories (period: 2017.11-2019.7)]. Artificial intelligence (AI) speakers (network), drones. The index was investigated by announcing indicators of the fourth industrial revolution in different areas such as autonomous vehicles (mobile), FinTech businesses, and smart factories (manufacturing). The results were surprising (2019.7.2).

Ultra-connected intelligent network	Artificial Intelligence (AI) speaker count (accumulation) more than doubled year-on-year to 4.12 million units.
	The number of Internet of Things (lot) service subscribers, which shows the status of entry into hyper-connected society, which is the basis of the fourth industrial revolution, increased 33.2% year-on-year to 8.65 million (as of January 2018).
	April 2019, 1 million 5G mobile telecommunication subscribers (based on 2019.6.10) – 10GB Internet network nationwide (national coverage) that will be serviced for the first time in 2019 to proactively respond to rapidly evolving network environments such as 5G is 6.86% June 2020, 6.34 million 5G mobile subscribers.
Intelligent technology	The number of use of artificial intelligence programs (open application program interface) developed and opened by the government for anyone to use when developing new services has increased by 7 times compared to the previous year to 12 million.
	The number of AI companies involved in intelligent technology, a key driver of the fourth industrial revolution, increased 25.7 percent year-on-year to 44 (as of 2018).

	The number of virtual reality/incremental reality (VR/AR) app uploads to the app market (app store, Google Play Store) was 7,065 (as of the end of April 2019).
Smart Mobile	Precision map for self-driving cars is 1,741km (28.9% increase compared to the previous year), and advanced road system (C-ITS) for autonomous vehicles is 323km (267% increase compared to the previous year).
	The number of vehicles that received temporary driving permits from the Ministry of Land, Infrastructure and Transport to operate self-driving cars increased 40.9 percent year-on-year to 62 units.
	The number of drone pilots qualified to operate the Smart Mobile System increased by 10,000 people from the previous year to 15,671 (end of January, 2018).
	Commercial (or over 12kg, subject to reporting) drones increased 60.0% year-on-year to 7,177 units (at the end of 2018.12).
Manufacturing	The number of smart factories (accumulated) being promoted to strengthen the competitiveness of small and medium-sized manufacturing companies is 7,903 (from January 2018 to February), and the government plans to build 30,000 by 2022.
	Manufacturer's robot production increased 13.1% to 3.18 trillion won (as of 2017).
FinTech	The daily average of 5.33 million cases and 230.6 billion won (2018), up 91% from the previous year and 123.4% from the previous year.
	The number of accounts and subscribers of Internet-specialized banks such as Kakao Bank and K-Bank increased 73.1% and 55.7% year-on-year, respectively, to 15.9 million accounts and 9.93 million subscribers.
	As innovative FinTech technologies spread and spread, the number of FinTech companies increased 4.9% year-on-year to 301.
Medical Industry	The number (accumulation) of hospitals participating in electronic exchange of medical information at medical institutions that have safely exchanged patient care information between medical institutions, not patients, increased 77.3% year-on-year to 2,316 hospitals.
Energy	The number of energy management systems (EMS) introduced through smart grid diffusion projects to expand the base of new energy industries increased 34.3% year-on-year to 564.

City	Local governments that participated in the smart city integrated platform infrastructure construction project to solve and manage urban problems such as transportation, crime prevention, and environment with the ICT innovation platform spread to 37 cities, up 15 from the previous year.
Agriculture and fisheries	The agricultural and fishery sectors also expanded smart farms incorporating ICT technology due to the influence of the Fourth Industrial Revolution, increasing facility horticulture by 22.2% to 4,900ha (hectare), and livestock production by 77.9%.
Resources: Ministry of Science and ICT. www.msit.go.kr, I-KOREA 4.0 (2.7.2019)	

2. Changing the educational paradigm

Currently, Korean education is in the era of Creator Education 3.0. The Korean Education 1.0 era was a state-led nationalist education, while the Korean Education 2.0 era was a liberal education.

2.1. Education 1.0 (Nationalism)

Nationalist education was a period of about 35 years from the 1960s to 1994, a stage of so-called catch-up development of education that learns and learns Western culture as soon as possible. At this stage, there was a strong tendency of nationalism in which the government was deeply involved in all curricula, and also the most effective memorization-oriented education for rapidly acquiring standardized knowledge of the West. Therefore, the assessment system was dominated by multiple-choice evaluations that were naturally focused on memorization.

Also, this was the time when semi-skilled or mid-skill personnel were needed the most because industrialization was progressing rapidly. So one of the key functions of education at the time was the mass production of semi-skilled or intermediate technical personnel, which are much needed in labor-intensive export industries (export light and heavy chemical industries). However, in the mass production of semi-skilled or intermediate technical personnel, memorization, repetition, and multiple-type evaluation were effective methods of education and training. Thus, during this period, the talent supplied by the education market and the talent demanded by the labor market were well matched.

2.2. Education 2.0 (Liberalism)

Liberal education is a period of about 20 years from 1995 to 2015. Specifically, it can be seen as since the educational reform of 5.31 in 1995. This was the time when the era of globalization began in earnest in the global community and the era of knowledge and informatization began. Globalization began in earnest as about 2 billion people were incorporated into the market economy due to the collapse of the socialist bloc since early 1990, and the era of industrialization (such as shipbuilding, oil refining, steel and automobiles) ended with the introduction of the IT economy-softening, and the era of knowledge and information service (IT Bio Information and Communication Semiconductor) was now opening. Now, even in education, the traditional uniform nationalistic educational paradigm has not been able to respond to the rise of the knowledge and information age, which requires a variety of skilled

workers and high-skill personnel. Thus, a new era of liberal educational paradigm could not be stopped. That was the historical significance of the 5.31 education reform.

5.31 Reforms typically had two liberal philosophies. One was to establish a consumer-centered education system. One of the basic directions of 5.31 was to expand the freedom and choice of education consumers and to make suppliers compete with each other for the provision of quality education, so that various educational programs could be introduced in response to changes in globalized knowledge and information.

And another liberal philosophy of 5.31 was to increase the autonomy of school governance. It was to change the centralized, hierarchical and regulatory nationalistic educational administration to liberal educational administration based on autonomy and responsibilities. In short, it was a great shift from uniformism to liberalism. Therefore, this era can be called the era of the Republic of Korea Education 2.0 when the liberal education paradigm was introduced.

2.3. Education 3.0 (Innovation and creation)

It can be said that the innovation-related education began around 2016. There are two main reasons why the evolution to Education 3.0 has become inevitable. One is the fact that the era of the fourth industrial revolution, dominated by artificial intelligence robots, the Internet of Things (IoT), mobile devices, 3D printers, unmanned vehicles, and nanotechnology, is beginning in earnest. And the other is the full-scale evolution of the so-called Network Society. The former demands highly creative talent. Ability beyond existing standardized knowledge is needed. And the latter demands highly cooperative talent. The ability to cooperate with others is the basis for new creation and innovation.

The combined talent of these two, or super-skill, is the talent of the super-skill required by the fourth industrial revolution. By educating these talents (in fact, schools, businesses and research institutes must create a successful economy and a nation in the era of the fourth industrial revolution).

3. Future Education

South Korea's education level is one of the leading groups in the world. Entering the year 2000, the Korean government has developed smart education programs and welcomed the Korea Education 3.0, continuing to challenge future education such as computer coding education and STEAM education.

Pearson's Global Index of Cognitive Skills and Educational Attainment		
Country	**Z-Score**	**Rank**
Finland	1.26	1
South Korea	1.23	2
Hong Kong	0.90	3
Japan	0.89	4
Singapore	0.84	5
United Kingdom	0.60	6

Netherlands	0.59	7
New Zealand	0.56	8
Switzerland	0.55	9
Canada	0.54	10
Ireland	0.53	11
Denmark	0.50	12
Australia	0.46	13
Poland	0.43	14
Germany	0.41	15
Belgium	0.35	16
United States	0.35	17
Hungary	0.33	18
Slovakia	0.32	19
Russia	0.26	20
Resource: https://bit.ly/2BLBbEp		

3.1. Smart Training

Korea's full-fledged smart education began with the "Smart Education Promotion Strategy" (2011). Basic research on smart education has been conducted since 1997. Prototypes of digital textbooks were developed in 2006, and commercialization of digital textbooks has been discussed since 2007. At that time, the Ministry of Education developed subjects such as Korean, English, and math into digital textbooks. Textbooks developed in the past are not currently used. This is because the curriculum was revised in the middle and some pointed out that the effectiveness of digital textbooks was insufficient. For this reason, smart education is being carried out again with a new policy launched in 2011.

Smart education consists of two main things. They are "smart classrooms" and "digital textbooks." In addition to digital devices such as tablets, smart classrooms are equipped with electronic blackboards, TVs, and wireless Internet devices to study new education methods. Teachers who run smart classrooms use various programs to conduct classes.

3.2. Computer Coding Training

At the 20th World Knowledge Forum in Korea, Hardy Patobi (Founder and CEO of Code.org), who is called the father of online coding education, called coding education a future education. Technology is changing our jobs significantly, and more than half of the current jobs disappear within 30 years. Coding education should be taught without distinction between students and teachers (The current Pope is also a student at Code.org. Currently, 40 million students are attending the world, of which 500,000 are Koreans.).

South Korea has made software coding education mandatory for elementary, middle and high school students since 2018. Coding education helps develop logic, creativity and problem-solving skills by learning the principles of computer programming language. Under the 2015 revised education program, middle schools will receive SW education for more than 34 hours in information subjects and elementary schools for more than 17 hours in practical courses in fifth and sixth graders starting from 2019. The teachers' group released "REDPEN" coding in November 2019. Woongjin Thinking introduced 'STEAM Textbook' in 2018. It is a program that combines coding principle with maker education. Talent education also released the second series of "Scottie Go!" in December 2019 in line with the mandatory coding education. Genius education included coding education contents in 'Milk Tea Elementary'. Most of these products are intended for infants and elementary school students.

3.3. Convergence Talent Training STEAM

The Ministry of Education, Science and Technology of Korea presented "Future STEAM Education Enhancement" (2011) in its "2nd Basic Plan for the Promotion and Support of Science and Technology Talent" (2011) to enhance understanding, interest and potential of science and technology. "Science and Technology-Art Convergence (STEAM)", including Arts (Art and Humanities) in STEM, was announced as the main policy. STEAM aims to create new values based on convergence knowledge in various fields, including science, technology, engineering, art and mathematics, and to foster talent with a comprehensive problem-solving power and STEAM literacy of the personality living together (Baek Yoon-soo et al, 2012). STEAM education worked. STEAM-trained students showed that their creative and fusion thinking (problem awareness, problem solving) improved, and after STEAM classes, self-directed learning skills, scientific preferences, math and science interests, self-efficacy, and consideration skills increased [“Effects of STEAM Education” Kang Nam-hwa et al., 2017; Sungmin Lim et al., 2014].

Through STEAM education, both students and teachers were transformed into pleasant classes, STEAM classes were improved into student-centered classes such as design, discussion and presentation of students, and participation and satisfaction in the classes were enhanced. In particular, the analysis of STEAM class video showed that the ratio of teacher explanation time in general classes was more than 90%, while the ratio of teacher explanation time in STEAM classes decreased to 8%-48% [“Effects of STEAM Education” Kang Nam-hwa et al., 2017; Hyunjoo Park et al., 2015].

3.4. Innovation Future School

Innovative Future School is a future school model that aims to cultivate creative democratic citizens based on the technology-integrated educational environment.

Korea organized the "Innovation Future School Planning Committee" focusing on the Seoul Future School [The Seoul Future School (Changdeok Girls' Middle School) was designated in 2014 and offers a leading model of future education, including ICT-based educational activities, integrating curriculum, classes, and evaluation, and establishing a school environment that matches the future learning system.], the Seoul Innovation School [Starting with the designation in 2011, the Seoul-type Innovation School has been contributing to the spread of student-participating classes, teacher learning communities, democratic and communicative school culture, and school models with local communities.] and Innovation School teachers, the Frontier Teachers' Group, and officials from the Office of Education to present the curriculum, school environment, teacher competency, and direction of school culture for the newly created innovative future school (2019.3):

1. Curriculum: Expands the autonomy of organizing and operating curriculum to operate a student-centered individualized curriculum, and design and operate a convergence curriculum in which schools, parents, and communities participate. It also leads individualized learning and student-centered assessment systems through an online learning platform.

2. School environment: Establish a technology-integrated school environment in which all educational activities such as class-evaluation, school (class) operation, and school administration are combined organically with school space.

3. Teacher Competence: Restructuring of curriculum for future competency development, technology utilization, student participation class design, integration of lessons and evaluations, teacher training to build competences, such as democratic communication for the school community that is being built together, individual and joint Research, research environment creation.

4. School Culture: Create a'discussed faculty meeting' culture of communication and participation for schools that participate and cooperate, and schools with parents and villages, strengthen the culture of cooperation and sharing through activation of the teacher learning community, expand student autonomy and democracy Strengthen citizens' capacity, expand parental involvement programs, and engage in community-related education

4. Conclusion

There are two key future-oriented education contents in the innovation school. One is to overcome the existing top-down education administration system and maximize the autonomy of schools and teachers to revive the bottom-up activism and dynamism of schools and teachers, while the other is to take care of each student according to their various dreams and talents and secure more creative classes. There are four features here: (1) the pursuit of a 'classroom revolution' beyond innovation; (2) the uninterrupted learning, the education that opens everyone's possibilities; (3) the fact that it is a creative educational space where learning, rest, and play come together (4). Education in the era of the Fourth Industrial Revolution calls for a fundamental change in the paradigm. Future education should focus not on knowledge transfer but on giving learning methods and enjoyment of learning. It should be an education that goes beyond the literacy in dealing with ICT and artificial intelligence to cultivate insight that penetrates the wisdom and knowledge of life, and to foster cooperative spirit, communication and empathy as a social being.

In terms of intelligence, machines will be ahead of people, but emotional areas will remain human territory forever. In this very sense, we have no doubt that the Korean people will develop into a leading nation in education in the era of the fourth industrial revolution.

References

1. Kim Jong-min, Park Bong-won and Park Sang-yong (2012). "High-tech Education in the Digital Age (Q2L) and Gangwon-do." Gangwon Development Institute. Number 178, 2012.8.14.
2. Report on the Results of the "2019 STEAM Teachers' Association", Hyeonhwa High School, 2019.11.15
3. Kim Young-chul, Im Chun-soon, Ban Sang-jin, Oh Heon-sik and Shin Joon-sup (2004), "Planning for the diversification of different academic systems in the Information Knowledge Society. Seoul: Maeil Business Newspaper.
4. Park Jong-hyun, Bang Hyo-chan, Kim Se-han, Kim Mal-hee, Lee In-hwan, Choi Byung-chul, Lee Kang-bok, Kang Sung-soo and Kim Ho-won, "The Future of the Internet of Things." Seoul: Electronic Times, 2014.
5. Lee Jong-jae, Kim Sung-yeol and Don Adams, "60 Years of Korean Education." Seoul: Seoul National University Press and Culture Center, 2010.

6. Jeong Jeong-yeong, Kang Tae-hoon, Kim Gap-seong, Ryu Sung-chang and Yoon Hong-ju, "A Study on the Implications of Teacher Policy in the Future of Educational Environment." Seoul: Korea Educational Development Institute, 2013
7. Kang Tae-joong, Kang Tae-hoon, Ryu Sung-chang and Chung Yoo-young, "Planning for Educational Development Strategies for Intelligent Information Society." Korea institute of education and development institute, 2016.
8. Gye Bo Kyung, "The Policy and Change of Future Schools in the World." Seoul Education, No. 224 (2016.10.5).
9. Kim Dong-hwan, "System accident. Competition: Prehistory.", 2011.
10. Kim Sung-il, Yoon Mi-sun and So Yeon-hee, "Korean Students' Interest in Study – Actual Condition, Diagnosis and Prescription." Journal of the Korean Psychological Society, 14(1), 187-221, 2018.
11. Ha Won-kyu and Choi Nam-hee, "The Fourth Industrial Revolution", 2015.
12. Heo Kyung-chul, "Operation of Future Society and School Curriculum." Curriculum Research, 19(1), 67-93, 2001.
13. PR Taek-geun, Yoo Ji-won, Im Jong-woo, Ahn Moon-hwan and Lee Ji-young (2016). "A Research for Strategies of Activating ICT convergence education." Ministry of Science, ICT and Future Planning, 2016
14. Kim Young-ae, "The Present Situation and Development Direction of Smart Education in Our Classroom Revolution." Seoul: Korea Educational Development Institute, 2011.
15. Park Hyun-joo, "Final report on basic research for the establishment of a comprehensive plan for convergence education." Chosun University. 2019.
16. Park Chul-woo, "The Fourth Industrial Revolution, Convergence Talent and Education Innovation", 2017.
17. Park Chul-woo, Monthly Science and Technology, p. 302, 2017
18. The Economist of Central City Magazine, 1361, 2016.
19. ISSUE PAPER Issue Paper CP 2017-02-06, Prospects and Tasks of School Education in the Fourth Industrial Revolution Era, Korea Educational Development Institute, 2017.
20. "The Prospects and Challenges of School Education in the Age of the Fourth Industrial Revolution", Issue Paper No. 6 (2017-02-06), Korea Educational Development Institute's Education Policy Network Research Center, 2017.
21. "Fourth Industrial Revolution Response Plan I-KOREA 4.0 ", Fourth Industrial Revolution Committee, 2017.
22. "Future Education Framework and Future School Studies suitable for the era of the Fourth Industrial Revolution." the Fourth Industrial Revolution Committee, 2018.
23. "The Five-Year Reorganization of Government Administration of Government Administration", the State Affairs Planning Advisory Committee, 2017.

The 4th Industrial Revolution and Education in North Korea

James Kang

The University of Veliko Tarnovo St. Cyril and St. Methodius,
Veliko Tarnovo, Bulgaria
bulgariajames@gmail.com

Abstract. Just three years after the term of the Fourth Industrial Revolution came into effect, a next-generation industrial revolution led by artificial intelligence, robot technology and life sciences, appeared at the World Economic Forum in 2016. The products of the revolution thereof are already penetrating our daily lives. North Korea is a well-known isolated country around the world. They have developed IT technology in their own way and have been aware of the 4IR. Therefore, North Korea can be seen as an industrial society before the third industrial revolution was completed. However, the country is facing challenges from the 4th Industrial Revolution in real life and education. Especially after the Go Match between Lee Se-dol, a former South Korean professional Go player of 9 dan rank, and Google's AI Go program AlphaGo in 2016, North Korea became aware of the changes caused by new technologies such as AI, big data, and blockchain (NK Economy, http://www.nkeconomy.com). In this paper, I would like to examine the changes in the perception of North Korea's fourth industrial revolution and its application in the field of education, its unification with South Korea, and the present the direction of development.

Keywords: Fourth Industrial Revolution, AI, STEM, Future education

1. The 4th Industrial Revolution

In the 1980s, North Korea began developing artificial intelligence using the theory of fuzzy theory and expanded it to include pattern recognition, speech analysis, and natural language processing. Full-fledged AI research began in the late 1980s when the AI Go program was developed. The "Seunbyeol" artificial intelligence Go program and the "Ryongnamsan" voice recognition software are representative products of development. Kim Il Sung University has been continuing to research and publish papers on artificial intelligence.

North Korea was shocked by the fourth industrial revolution that had already begun at the World Economic Foundation in Switzerland in January 2016. ["The concept of the Fourth Industrial Revolution came out for the first time at the World Economic Foundation in Switzerland in January 2016." Based on the analysis of its impact on the economy, culture, education and society as a whole on the global scale, many countries are currently seeking ways to prepare for it, and comprehensive reviews of the education system and curriculum to foster future talent are under way." (North Korea's Rodong Newspaper, 2019.9.10)].

AI field

With leader Kim Jong-un entering the regime, North Korea launched its fourth industrial revolution through science and technology policies represented by the "Computerized Numerical Control (CNC) policy of the whole country." Among them, artificial intelligence is what North Korea is most interested in these days.

North Korea has already developed Go artificial intelligence software "Eunbyul" since 1997, led by the Chosun Computator Center (KCC), a software development organization established around the 1990s, and the Pyongyang Information Center (PIC). "Separate" combines Monte Carlo's method of calculating and selecting the best number based on the results of the existing mock-up, with its own algorithms to run the program. The Monte Carlo method was also used for Google's "AlphaGo" and Nurigrim's "Dolbaram," one of the best Go artificial intelligence in Korea. "Eun Byul" won her first World Computer Go Championship title in 1998, and achieved four consecutive wins from 2003 to 2006.

Since then, North Korea has continued to release upgraded versions such as "Seunbyeol 2006" and "Seunbyeol 2010." In particular, Eunbyul 2010 won the 2009 World Computer Go Championship in Japan, leaving behind an undefeated legend. Eun-byul also won the competition against other artificial intelligence using up to 1,000 CPUs, using only 16 CPUs.

In 2017, Kim Il Sung University introduced its own language translation program, "Ryongnamsan 5.1." Ryongnamsan Mountain is an AI program that translates social science documents in English into Korean, and supports the translation of knowledge in more than 30 specialized fields such as physics, biology, chemistry, mathematics, information technology, earth science, and medicine. According to North Korean researchers, a large-scale circular neural network language model has been applied to the program and the accuracy of speech recognition is 98 percent. It is also characterized by a limited number of recognizable words, with a vocabulary recognition rate of 6 characters per second.

At the 29th National Information Technology Performance Exhibition in November 2019, North Korea presented the "Intelligence House," a concept of smart home, by the Information Technology Research Institute of Kim Il-sung University's Institute of Advanced Science. It is the AI speaker 'intelligent high voice' that acts as a control tower for this intelligence house ["It is a device that can recognize human voice commands and realize automatic control of fans, air conditioners, televisions and lights," he said. "It was developed in North Korea's own way." (North Korean Rodong Newspaper, 2019.11)].

North Korea has already been interested in developing face recognition, voice recognition and iris recognition technologies since the mid-2000s, and has actually applied them to security programs at military plants ["Although the level of related technology in North Korea is not comparable to that of South Korea, it is noteworthy that they are developing their own methods by reading international trends and interpreting them in their own way," said Dr. Kang Young-sil (a researcher at the North Korean Science and Technology Research Center)].

In addition, North Korean researchers are working hard to develop AI technologies such as Artificial Neural Network (Artificial Neural Network) and Deep Belf Network (DBN) to apply them to various fields such as text recognition, voice recognition, process efficiency, and games.

North Korea's Artificial Intelligence Study [A Study on the Fourth Industrial Revolution and Inter-Korean IT Economic Cooperation, 37p]

Source of dissertation	Subject of a thesis	Content of a thesis
Journal of Kim Il-sung's University Studies, Vol. 62 No. 8 of 2016	“One method of sound model using deep trust network in sound recognition.”	Use 'Deep Brief Network' (DBN), one of the following machine learning techniques, for voice recognition

Journal of Kim Il-sung's University Studies, Vol. 63 No. 4 in 2017	"One Way to Build a Massive Recursive New Gyeongmang Language Model for the Adult Ceremony of Joseon Eoryeon Sound"	By applying a large-scale circular neural network frozen language model to Ryongnam Mountain, a Korean language recognition program, an AI program that translates English-language science and technology texts into Korean, supports more than 30 professional translations including math, physics, chemistry, biology, information technology, old environment, and medicine.
Journal of Kim Il-sung's University Studies, Vol. 63 No. 10 of 2017	"Planning the construction of Ray Low of Aluminium Thin Plates Using Artificial Neural Network"	AI technology is applied while studying the use of re-easer on aluminium thin plates to write down the Artificial Neural Network (Artificial Neural Network) technology to increase the efficiency and precision of the process.

2. The Education Field

North Korea is well aware that the existing education system is focused on fostering memorized talent and that most of its education methods rely on memorization, so it cannot respond to technological development and industrial changes in the era of the fourth industrial revolution [North Korea's Rodong Newspaper. 2019.9.10]. Accordingly, the government is seeking policies to foster creative talent that creates new knowledge.

2.1. STEM Training [NK Economy (http://www.nkeconomy.com)]

In 2019, Kim Jong-un emphasized science and education as a national policy. STEM is an abbreviation of science, technology, engineering, and math, and is also trying to reform the education system in North Korea as it aims to foster creative convergence talent [NK Economy (http://www.nkeconomy.com)]. The first middle school in Pyongyang has achieved the result of creating a new method of education (stem education) in the process of carrying out projects to nurture students into creative talents. North Korea is seeking to apply stem education from kindergarten to universities [Experts of North Korea's Arirang Echo. 2019.8.25].

In 2019, at Mirae Small School in Pyeongcheon, Pyongyang, teachers directly participate in the development of artificial intelligence education programs and use them for students' classes. It is a program that evaluates the pronunciation of English words on the microphone and displays the score on the screen. In addition, we are maximizing the understanding and learning effects of students by using new technology such as showing moving 3D images in math and science classes. In addition, AI education programs are produced in various formats according to students' age and psychological characteristics.

2.2. Future education

North Korea is making innovative plans to seriously analyze and review existing problems in school education and boldly accept teaching methods and methods that fit new aspects of future school education.

With the introduction of information and communication (IT) technology at the educational site, lectures are being rewritten using intelligent information technology, and the educational environment is being designed based on digitally, and electronic books are replacing paper books. North Korea is pushing for the development of new future schools in harmony with the changes in intelligence and information society.

North Korea's Science and Technology Center (Source = KBS NEWS)

3. Conclusion

North Korea has stressed the importance of science and technology and the development of high-tech industries over the years, but has not used the term "fourth industrial revolution." Then, he began mentioning the fourth industrial revolution in 2019. And the term "fourth industrial revolution" appeared in the introductory materials displayed by the Pyongyang Information Technology Bureau at the National Informatization Exhibition-2019 held in November 2019. Now, lectures on the fourth industrial revolution are being given to residents.

There is the only divided nation in the world of the Korean Peninsula. South and North Korea have been divided for 75 years. Efforts to reunify Korea have been made even more in full swing since the inter-Korean summit in June 200. As an isolated country, North Korea has many limitations to its own challenges of the Fourth Industrial Revolution. South Korea has led cooperation, including the IT sector, in the wake of the 2018.4 inter-Korean summit in Panmunjom.

The South Korean government is pushing for joint research and cooperation between the two Koreas on future science and technology, selecting "South and North Korea's joint response to the fourth industrial revolution" as a research task in December 2019. Cooperation with the South is the only way for North Korea to develop.

The following is about IT economic cooperation between South Korea and North Korea.

IT Economic Cooperation between South and North Korea by Steps [A Study on the Fourth Industrial Revolution and Inter-Korean IT Economic Cooperation, 39p]

Step of Cooperation	Goal	Contents of IT Economic Cooperation
U.N.-North Korea sanctions lifting (Stage 1)	• Mitigate tension in inter-Korean relations • System maintenance and foundation. • Focused on people's livelihood humanitarian dimension cooperation	• Organization of IT Economic Cooperation Promotion Committee • Exploring the industrial status and detailed development of joint economic cooperation projects through the comparative advantage of the two Koreas in relation to the Fourth Industrial Revolution • Developing and resolving obstacles to economic exchanges and cooperation related to the Fourth Industrial Revolution between the two Koreas. • Establish a joint IT research center and conduct human exchanges and technical exchanges while operating the facility.
After the UN sanctions against North Korea were lifted,(Stage 2)	• Inter-Korean Exchange Active • Active IT joint business development	• Development of joint technology to promote cooperation in inter-Korean fourth industrial revolution and advance into new markets in the future • Establish and operate an IT-specialized complex in the DMZ to promote business cooperation and cooperation between the government and the government.
inter-Korean mutual aid Integration Phase (3 Step)	• Inter-Korean Social Integration • Integration	• Interchange and cooperation related to the Fourth Industrial Revolution between the two Koreas and explore new markets under the New Economic Initiative

References

1. Kim Heung-kwang, Moon Hyung-nam, Kwak In-ok, Fourth Industrial Revolution and North Korea, Su-in, 2017.
2. Im Jong-in, Kwon Yu-ju, Jang Kyu-heon, Baek Seung-jo, North Korea's current status of cyber warfare and South Korea's national response strategy, National Defense Policy Research Volume 29, Volume 4, 2013.
3. Cho Sung-sun, Quantum Computer Development Trends and Implications, Information and Communication Technology Alcohol Promotion Center, 2018

4. Hyundai Economic Research Institute, the possible link between North Korea's economic development zone and the idea of a "special economic zone for unification," 2018,
5. Kim Sang-bae. Inter-Korean relations in the era of the Fourth Industrial Revolution. Seoul National University. 2019
6. Lee Chun-geun, Kim Jong-sun, and Nam Dal-ri. 2014 “Planning for Inter-Korean ICT Cooperation” Policy Research 2014-28. Science and Technology Policy Institute.
7. NK Economy (http://www.nkeconomy.com)
8. Unification Ministry UNITV, "There is no stone without meaning, Baduk"
9. KBS NEWS, "[These days, North Korea] emphasizes 'science and technology achievements' ahead of the 9.9 Independence Movement."
10. KBS NEWS, "[Yose North Korea] Early Science and Technology Education... Other than emphasizing 'multi-faceted experience'."
11. KBS NEWS, "[Closed-up North Korea] and "From..." "Competition for entrance exams is fierce."
12. MBC Unification Observatory, "[North Korea now] transformed the North Korean police into English with artificial intelligence technology, "kindness instead of pressure?"
13. ZDNet Korea, "Baduk Game AI.South Korea's 'Dolbaram' vs. North Korea's 'Seunbyeol' http://www.zdnet.co.kr/view/?no=20160316105549
14. NK Economy, "North Korea's No. 1 Artificial Intelligence Technology" http://www.nkeconomy.com/news/articleView.html?idxno=701
15. Seoul Pyongyang News, "Status and Prospects of Artificial Intelligence Development in North Korea" http://www.spnews.co.kr/news/articleView.html?idxno=4306
16. TECH m, "North Korean version of the Fourth Industrial Revolution Fever, including Artificial Intelligence and Augmented Reality," http://techm.kr/bbs/board.php?bo_table=article&wr_id=4913
17. Development of Remote Education Program for NK Economy, North Korea and the U.S. Electronic Learning Standards, http://nktoday.kr/?p=13988
18. BBC News, "AI SPEAKER: North Korea has launched an intelligent home-household that appears to be a smart home concept." https://www.bbc.com/korean/news-46206854
19. South Korea's Unification Ministry said, "North Korea, too, should find out about the North's artificial intelligence (AI) in the era of the fourth industrial revolution."
20. Schwab, Klaus. 2016. The Fourth Industrial Revolution. World Economic Forum.
21. Schwab, Klaus et al., The Fourth Industrial Revolutio n, Portfolio Penguin, 2016.
22. http://news.mk.co.kr/newsRead.php?no=554708&y ear=2017)
23. http://techm.kr/bbs/board.php?bo_table=article&w r_id=4913.
24. http://www.dihur.co.kr/1518.

Creative cues to Identify and Design Opportunities Driven by Artificial Intelligence Technologies

Ma-Lorena Escandon-Quintanilla[1], Alejandro Gutierrez-Lopez[2]

[1] Ryerson University, Toronto, Canada
[2] McGill University, Montreal, Canada
lorena.escandon@ryerson.ca, alejandro.gutierrezlopez@mcgill.ca

Abstract. Access to new information and communications technologies allow companies and organizations to provide their clients with better products or services, or improve their internal processes. There is an opportunity to integrate artificial intelligence into new processes, products and services. The main challenge is to generate ideas that are technologically feasible, economically viable, and useful for the users. Idea generation can present challenges such as: fixation on dominant design, settling on "good enough" ideas, and not understanding how a technology works. By using creative cues, it is possible to elicit more ideas, with the potential to maximize the benefits of a new technology. Solution designers need to understand where and how artificial intelligence can be integrated into existing products and services, and most importantly, how to design novel products and services that could not have been possible without artificial intelligence. This article presents a modified customer journey map and a card set designed based on TRIZ inventive principles to elicit creative solutions driven by artificial intelligence technologies, and the validation of the design with three domain experts.

Keywords. Artificial intelligence, New product development, New service development, Creativity tools.

1. Introduction

In recent years, there have been significant advancements in the field of artificial intelligence (AI). The availability of large datasets, the increase of computing power, and improvements in different analytical techniques have created a new golden era for artificial intelligence. This renewed interest in artificial intelligence has started to permeate into companies and organizations of all sizes. Nonetheless, most managers and engineers in charge of designing new solutions are not usually trained in the technologies and techniques used in artificial intelligence. There is a need for business users to understand where and how artificial intelligence can be integrated into existing products and services, and most importantly, how to design novel products and services that could not have been possible without artificial intelligence.

Designing new products and services using old conventions limits the possibilities of what new technologies can help the organization achieve. Some examples of applied artificial intelligence can be seen in recommendation systems, speech and image recognition and automated translations, among many others.

The introduction of artificial intelligence represents a change in the paradigms and mental models of decision-makers about what can be done with these new approaches. There are multiple technologies, techniques, and tools in artificial intelligence, all of which serve

different purposes. Learning where and how to apply them requires a considerable amount of time and resources.

The main challenge is to generate ideas that are technologically feasible, viable for the business, and useful for the users. This requires collaboration between multiple areas of the organization; technologies must be aligned with the business objectives, considering legal requirements and the ethical aspects of the technology and the data that will be used.

The purpose of this paper is to introduce a proposed process and creative cue cards that are meant to help participants in their design efforts during the idea generation stage. By providing creative cues from TRIZ, data types and digital transformation opportunities, this process and card set can support the idea generation process for innovation with artificial intelligence and machine learning.

After the process and cards were designed, they were validated by three domain experts in different areas: Human-Computer Interaction (HCI), software development and deployment (SW), and User Experience design (UXd). This paper will cover the validation process, and the lessons learned from the expert validation iterations. The next section introduces the different elements covered by the cards. Later, the design and validation of the cards and process are presented before outlining the conclusions.

2. Creativity and idea generation for new artificial intelligence based products and services

2.1. Idea generation

Creativity and idea generation are paramount to the creation of new business processes, products and services. However, idea generation is no easy task. Participants attempting to generate an idea with no support usually face common obstacles, they tend to fixate on known or popular solutions, or to stop generating ideas once they have found one that is "good enough" [1]. By using creativity tools, like card-based games, it is possible to provide participants with creativity tools to elicit more ideas, with the potential to maximize the benefits of a new technology [2]. Participants are also limited by their understanding of how something works, their previous knowledge and experience, and their capacity to recall information.

Original ideas stem from the combination of unrelated knowledge [3][4], but participants have to be motivated to make these connections [5]. The purpose of idea generation sessions or workshops is to set an environment and implement creativity tools and techniques that will help participants produce, combine and express ideas. The ideas of others can promote the creation of related ideas or new ideas [6], while creativity tools and techniques are useful to bring participants to explore ideas "outside their normal frame of reference" [7].

Creativity tools, like card-based games and canvases, have proven to be an effective aid for idea generation tasks, helping overcome fixation issues, and serving as cues for new solutions and combinations.

2.2. Artificial intelligence and machine learning

Artificial intelligence is a field of computer science that aims to provide computers with the ability to learn and solve problems [8]. There are different domains, techniques and tools covered under the umbrella of artificial intelligence.

Machine learning is a domain of artificial intelligence that studies the development of algorithms and mathematical models to find patterns in data sets. It has garnered relevance in the last years, mainly due to the increased availability of data, the increase in processing capacity, and cloud-based architectures that make processing more accessible. Technological

resources like data storage and data processing can be acquired on an as-needed basis, making it more affordable.

2.3. TRIZ

TRIZ is a problem solving and creativity theory that proposes a series of methods to imagine the future, and solve technical problems [9]. Within the proposed methods in TRIZ, the Matrix of contradictions provides users with 39 generic engineering parameters, whose contradictions can be solved with 40 inventive principles.

One way to help participants generate ideas is to create playing cards with the inventive principles that can be materialized by integrating AI to a business process, product or service. TRIZ principles have already been used in other fields, for example, [10] distilled the principles that were directly applicable to problems of sustainable design. Some examples of inventive principles that can be realized through the integration of artificial intelligence are: Segmentation, Merging, Consolidation or combining, and Feedback.

2.4. Customer Journey Map

Customer Journey Maps (CJM) have been widely adopted by design thinking practitioners as a way to identify "pain points" for users in a myriad of industries. The purpose of the CJM is to first model the current situation for a user engaging with a solution, to identify areas of opportunity, and then to reimagine the solution to enhance the experience of the user.

Depending on the facilitator, the time available for the design session, and the purpose of the session, CJMs can be adjusted for depth and complexity. A CJM is usually comprised of three main "zones" (work areas in a template): Actions/Steps, Feelings/Emotions and Thoughts. The actions/steps are the observable actions taken by the user, whereas the feelings, emotions and thoughts are unobservable, and are the way in which designers must use their empathy to understand the user's experience.

In examples where the CJM includes several forms of interaction between the user and the company, designers have added a "zone" called touchpoints, where they can identify the different technologies (mobile app, website, phone call, in-store kiosk, etc.).

3. Design of the artifacts - TRIZ-based cards and process

As a solution for the issues on how to generate innovative ideas for products and services with based on artificial intelligence, or with an artificial intelligence component as outlined above, we designed a process and a set of cards to be used as cues to support creative elicitation for the generation of ideas for new processes, products and services with an artificial intelligence base or component.

3.1. Research method

Following the design science research artifact-design cycle, we first created the cards based on TRIZ inventive principles, a survey of data types, a survey of emotions, and the identification of digital transformation opportunities using artificial intelligence and machine learning algorithms. Then, based on documented issues with idea generation sessions, we designed a process to use the cards coupled with a Customer Journey Map, to identify the areas of opportunity where a "pain point" can be resolved by integrating innovative applications of artificial intelligence and/or machine learning.

Once the process and cards were designed, three domain experts were consulted to make improvements on the process and the concepts included in the card set.

3.2. Cards and CJM process design

We created cards with "creative cues" that include:

- TRIZ innovation principles - The TRIZ innovation principles will serve as the main creative cue, by providing designers with ideas on how to innovate, to break the fixation with the dominant design.
- Types of data - The types of data cards will allow designers who are not familiar with the different types of data to understand what kind of information can be exchanged.
- User emotions - User emotions will expand the possibilities to imagine other feelings that can be evoked, to go beyond "happy" and "sad" or "frustrated", we propose emotions like "empowered", "confident", or "relaxed" to prompt more creative solutions.
- Digital transformation opportunities - Digital transformation opportunities cards were designed based on the fact that TRIZ is based on patents for physical inventions, and the digital business models have other opportunities that can also be prompted in the Customer Journey Map.

A sample of the initial card set can be seen in Figure 1:

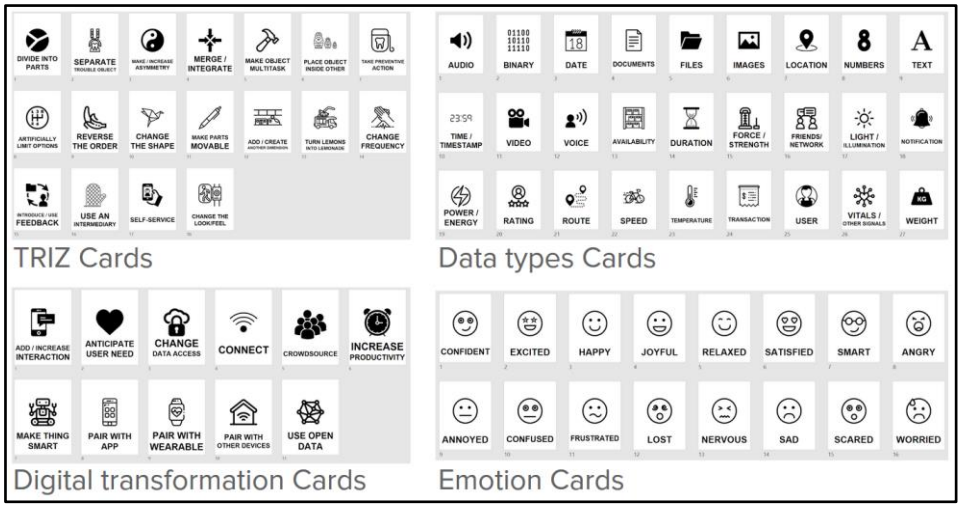


Fig. 1. Initial cards design (including TRIZ, Data Types, Digital Transformation and Emotions)

The CJM was adapted to include "Data exchanges" as a complement to the technology touchpoints. Here, designers can illustrate the information exchanges that occur between the user and the service or product provider. The template for the CJM can be seen in Figure 2.

The process to use the cards was initially designed to have the users create a map of the current situation (as-is scenario). They would then be provided with the four sets of cards, and they would generate new ideas in a future scenario (to-be scenario) to include the new available technologies. This process is depicted in Figure 3.

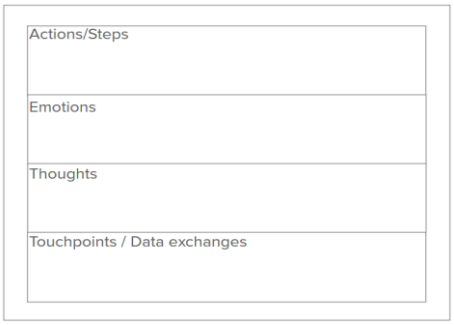


Fig. 2. Adapted CJM to include Data exchanges

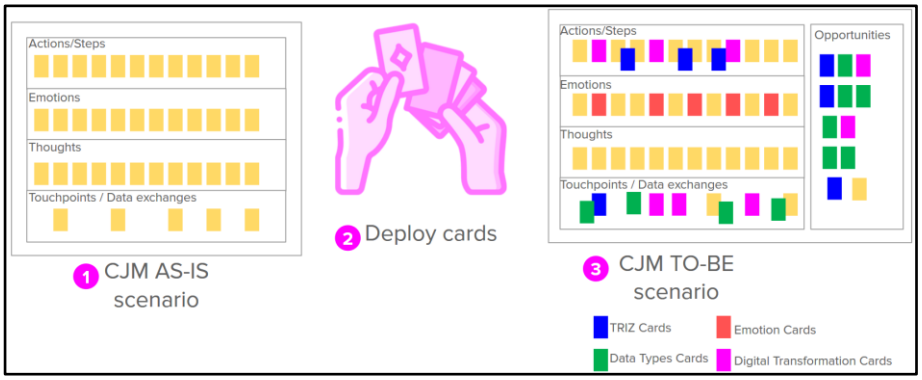


Fig. 3. Initial CJM process design (including As-Is scenario, deploy Data Types, and To-Be scenario using cards)

3.3. First iteration - HCI expert validation

In the first iteration of validation, we explained the process and the cards to an HCI expert. She has experience designing creative collaboration sessions with technical and non-technical users to attempt to include artificial intelligence in the journalism industry. In her experience, it is important to first create trust between participants, and leave the technology to a second moment of the collaboration, as non-technical users may withdraw from the conversation, and technical users will immediately try to implement the technologies they are familiar with.

From our proposal, she appreciated the usefulness of a process and tool that can promote dialogue and creative thinking during the idea generation stage, before "solutions" have been defined. The abstraction level of the concepts are abstract enough so that both technical and non-technical users can understand them and discuss big ideas, before diving into the details of how the solution will work.

Key suggestions

- Reduce number of cards - can be overwhelming and confusing for participants, they will try to add all cards even if it's not necessary.
- Scaffold the deployment of cards, provide TRIZ and Emotion cards first (focused on creativity), and data types and digital transformation cards later (focused on technology).
- Provide additional information in cards - cue questions, definitions, examples, etc.

- Add a blank card to afford creativity (shown in Figure 4).

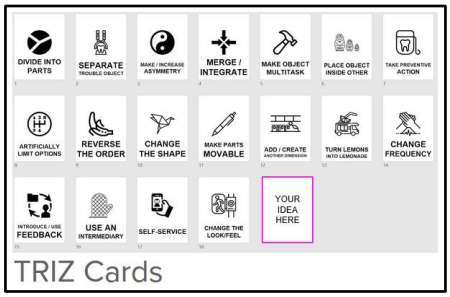


Fig. 4. TRIZ Cards with new "blank" card for flexibility

3.4. Second iteration - Software expert validation

During the second expert validation, we presented the process and tools to a software developer in the banking industry. He has had several experiences where collaborations with final users are needed to define digital solutions. During the conversation, he expressed how it can be difficult for final users to imagine a different way to do things; final users expect to follow the same process but with a "digital" tool (basically copying a paper form to a web form). Another difficulty is that organizations will most of the time already have a supplier whose solution has to be retrofitted to the process.

The proposed process and cards appear to be useful to have a conversation between business users and the technical team, to take a step back and define the actual problem that's being solved, instead of just translating processes from one technology to the next.

Key suggestions:

- Digital transformation cards could be complemented with principles of data solutions like data cubes, to facilitate the conversation between tech experts and users.
- Blank cards can be filled by tech experts in the company, to leverage existing technologies within the company.
- Creative cards can include "creative constraints" that align with the company requirements.

3.5. Third iteration - UX designer

The third expert is a UX designer who has experience implementing user research, he is familiar with tools and methods to elicit requirements and observe user interactions with systems to design more inclusive and accessible systems.

In his experience, to get creative ideas it is better to let participants in a design session go bold with their first "to-be" scenario (what if you had a magic wand that could make everything you imagine?), and then introduce the organizational constraints and all the information gathered about the resources available for the solution (data available, technologies available, etc.).

Key suggestions:

- Instead of an "as-is"→ "to-be" scenario where the to-be already includes constraints, add a third iteration to have three CJMs: "as-is"→ "ideal to-be"→ "implementable to-be" (minimum viable product) - Pictured in Figure 5.
- After the "as-is", ask participants to define the specific problem that needs to be solved, and what is the benefit, applying the 5-whys technique.

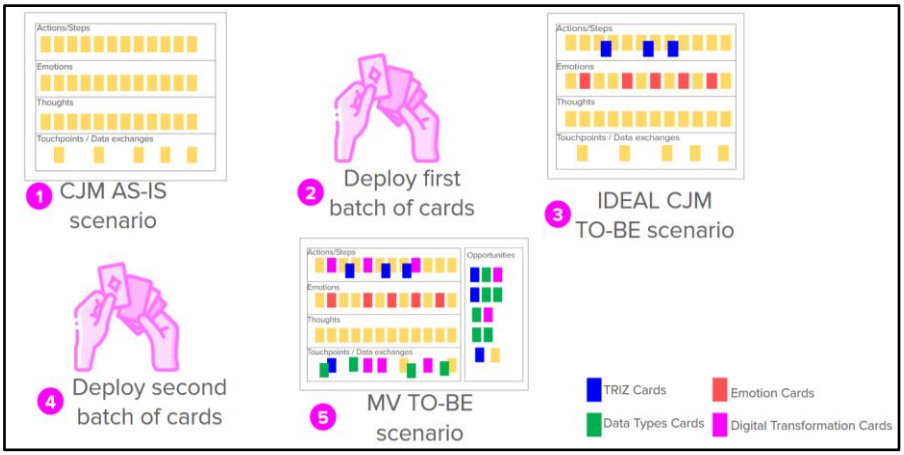


Fig. 5. The final process includes an Ideal to-be scenario and an implementable to-be scenario

4. Conclusion

After completing three cycles of design, validation and improvement of the cards and the process to use them, some recommendations and ideas were found to be recurring from the three experts' opinions. The main feedback received for the process and the cards is that there is a need to use tools to facilitate the elicitation and communication of ideas during the design of solutions within organizations. This need is more palpable when the teams working to design the solution have different roles in the company and varying levels of understanding of technologies like artificial intelligence and machine learning.

A second idea that came to light in the three reviews was the importance of being able to customize and expand in the number of designed cards to include technologies or solutions developed or used in the company, as well as the resources already available.

To avoid overwhelming participants with too many cards, the ideal order in which different types of cards should appear should be clearly defined. The two human-side experts - the expert on Human-Computer Interaction and the User Experience Designer - both agreed that the teams should begin with emotions and the TRIZ inventive principles in the first iteration of the to-be scenario, and the data types and digital transformation cards in the second iteration to create minimum viable solutions after the team has created new ideas without constraining the potential solutions to pre-defined technologies and resources.

Future work includes the implementation of the process and cards with users to validate the design.

References

1. M. Agogué, K. Levillain and S. Hooge, "Gamification of Creativity: Exploring the Usefulness of Serious Games for Ideation," *Creativity and Innovation Management*, vol. 24, no. 3, pp. 415-429, 2015.
2. M.-L. Escandón-Quintanilla, M. Gardoni, and P. Cohendet, "Improving concept development with data exploration in the context of an innovation and technological design course," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 12, no. 1, pp. 161–172, Feb. 2018.
3. A. Koestler, *The act of creation*, 6th ed. London: Hutchinson, 1964.
4. M. Nielsen, *Reinventing discovery*, Princeton, N.J.: Princeton University Press, 2012.

5. L. Fleming and M. Szigety, "Exploring the tail of creativity: An evolutionary model of breakthrough invention," *Advances in Strategic Management*, vol. 23, pp. 335-359, 2016.
6. J. Munemori, T. Yoshino, and K. Yunokuchi, "A spiral-type idea generation method support system for sharing and reusing ideas among a group," *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace* (Cat.No.01CH37236), 2001, vol. 3, pp. 1898–1903 vol.3.
7. G. Dove, and S. Jones, "Using Information Visualization to Support Creativity in Service Design Workshops." *Proceedings Service Design*, 2014.
8. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach,* Upper Saddle River, New Jersey: Pearson College, 2010.
9. G. Altshuller, "The innovation algorithm. TRIZ, systematic innovation and technical creativity/translated by L. Shulyak and S. Rodman," *Worcester: Technical innovation center,* 1999.
10. A. Cherifi, M. Dubois, M. Gardoni and A. Tairi, "Methodology for innovative eco-design based on TRIZ," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 9, no. 3, pp. 167-175, 2015.

Digital Marketing Support System for Agriculture

Iryna Borodkina[1], Heorhii Borodkin[2], Kateryna Kotsuibivska[1]

[1] Kiev National University of Culture and Arts, Kyiv, Ukraine
[2] National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
borir@ukr.net, katysivak@gmail.com, george.borodkin@gmail.com

Abstract. The article analyzes issues related to the development of an information system which was designed to support digital marketing in the agricultural market. Topical issues of creating and using marketing information systems are identified. The main areas of digital marketing and the information flows that emerge are explored. The tasks to be solved by the digital marketing support information system are determined. The general concept of building a marketing information system is considered. The main components of the marketing information system are distinguished. Requirements are set for system functionality that will allow effective use of all available sources of information. Here proposed the example of the implementation of a marketing information system. The stages of creating a model of an information marketing system based on the application of SADT methodology are described. The use of the DFD approach made it possible to reflect the data flows for the developed system and to visualize the processes of their processing. In this case data flows were considered both as material and information flows. The results of the modeling of the system by means of UML are provided.

Keywords. Marketing information system, Agricultural products, Information flows, Software system's design by means of UML, Digital marketing.

1. Introduction

The current stage of development and reform of the Ukrainian economy is characterized by a constant increase in the activity of small and large enterprises in the domestic market, as well as tendencies of its active integration into the world community, development intensification of new business forms. In these circumstances, the interest of Ukrainian manufacturers to study the problem of developing a product promotion strategy is growing.

Marketing activity in Ukraine is becoming increasingly important today. To maintain it, you need to have constantly updated online and accurate marketing information. This can be achieved through the use of information systems capable of servicing multimillion-dollar transactions daily. That is why issues related to the use of the latest information technologies to provide convenient information support for digital marketing in the domestic and foreign markets for Ukraine are necessary and relevant. One of the possible ways to solve this problem is to create appropriate marketing information systems.

2. Basic problems of digital marketing development

The use of modern information technology in the context of intensive development of market relations is today a significant factor in effective management and marketing. [1] Increasingly, modern information technologies and systems are being used to track and analyze information flows. This has given rise to a new direction of marketing - digital marketing. Digital marketing is a state-of-the-art information technology to attract and retain customers using digital technologies. Digital marketing includes such digital information

channels as the Internet, radio, television, mobile technologies. Digital strategies allow you to attract virtually all possible online and offline resources.

If the information system solves the problems of gathering, analyzing and processing information that is necessary for the preparation and decision making of marketing, converts it into the data that is needed for marketers and managers of marketing services, then such a system belongs to the class of marketing information systems (MIS). Much attention is being paid to the study of different aspects of the creation and use of MISs today. Topical issues include:

- determining the role of information in decision making;
- substantiation of the needs of enterprises in marketing information;
- study sources of marketing information;
- development of classification of marketing information sources;
- identifying possible forms of marketing information organization;
- substantiation of methodological toolkit for creation and functioning of MIS;
- identifying the main benefits of implementing MIS;
- methodology development for using information systems in the case of implementation at a digital marketing enterprise.

Despite the diversity of marketing systems, we can distinguish the following main components (subsystems): subsystem processing of internal marketing information; subsystem collecting external current marketing information; the marketing research subsystem; marketing information analysis subsystem. It should be noted that the assignment of subsystems corresponds to the functions that the marketing information system must perform. Among the information sources for such systems are information that already exists in other enterprise systems, information that is purchased and information that is collected independently. The processing of this information can effectively solve the problems of estimating and modeling the level of implementation of plans and implementation of marketing strategies based on statistical analysis methods; evaluating the results of implementing digital marketing strategies and activities; identification of favorable conditions, potential opportunities, difficulties and problems [2, 3]. In this way, the marketing information system transforms the information into data that is absolutely necessary for the managers of the structural divisions of the company and specialists of marketing services. The ultimate purpose of processing such data is to identify trends in market development, to ensure appropriate and timely response of the company management to any fluctuations in the market, to increase profits by reducing the risks in the company [4, 5].

3. The general concept of building an information-analytical system to support digital marketing in agriculture

The need to improve the quality of marketing services in the field of agricultural products requires the development of an appropriate information system, which will meet the information needs of managers in marketing information and will allow to plan, forecast, solve problems and control the activity of the enterprise. This system should have a functional that will allow to use effectively all available sources of information, to analyze the information received, to study the needs of the market, to analyze its sales, to forecast business trends, to analyze the range and quality of competitors' goods, to provide advisory information to improve the product - marketing activities of the enterprise, to carry out short-term, medium-term and long-term forecasting, to study market reactions to a new product, to form and pricing and distribution policies.

More recently, to successfully enter the market and successfully promote products, it was enough to use the methods and technologies of Internet marketing, which provided a sufficient level of profit. Today, Internet technology is no longer enough for a successful business. New digital communication channels appear with a wide range of opportunities for promoting goods. Internet marketing has been replaced by Digital Marketing.

Digital marketing is a comprehensive approach to promoting a company, its products and services in a digital environment, covering both online and offline consumers using games, mobile phones and other digital communications.

Digital marketing integrates a large number of different technologies (social, mobile, web, CRM systems, etc.) with sales and customer service; provides constant high-quality two-way communication between the advertiser and the end user of the product / service. It combines digital technologies and human resources, maintaining the right balance, based on the needs of the target audience and the properties of the proposed product. Digital marketing creates optimal conditions under which it is possible to be relevant to the market, evaluate and analyze the results of promotion, flexibly respond to the needs of leads and adjust product or service. Digital promotions is successfully used in online stores, online training, selling services and products with a long transaction cycle, to promote the brand of large companies and startups [6].

The table 1 analyzes the main differences between Internet marketing and digital marketing [7, 8].

Table 1. Internet marketing and Digital marketing comparisons

	Internet marketing	Digital marketing
Scope of application	Online sphere	Online + Offline
Distribution channels	Internet feeds	All types of digital channels (Internet, mobile applications, digital advertising, tablets and game consoles, digital TV, etc.)
Target Audience	Everyone who has access to the Internet	Everyone who has access to the Internet + attracting an offline audience to the online market
The way to communicate with the audience	E-mail newsletters, landing pages, sites, advertising (search, banner, targeted, contextual), etc.	Everything used in Internet marketing, + digital television, advertising in online games and mobile applications, instant messengers, interactive screens, POS-terminals, local networks of large cities

Currently, digital marketing uses the following resources:

- digital TV and online radio;
- advertising in applications, instant messengers, online games;
- SMS and MMS;
- advertising on interactive and street LED screens, self-service terminals;
- SEO and SEM - search engine marketing on the web;
- contextual advertising, banners and teaser advertising;
- SMM - social media marketing;
- affiliate marketing, in which the webmaster receives payment for attracting each visitor or buyer;
- E-mail marketing.

All tools that are used in digital marketing can be divided into 3 main groups:

Group 1 - paid resources: all digital resources for which the company needs to be paid;

Group 2 - own resources: sites, pages in social networks that the business owns and can manage;

Group 3 - communication resources: paid free or acquired information channels.

Studies have shown that today the most popular such digital marketing strategy options:

- Website development through an audit, as a result of which recommendations and a plan of further actions, budget are formulated.

- SEO strategy and search engine promotion.

- Development of the online sales systems.

- Internet PR, implemented through articles, banners, social networks, blogs.

- Email marketing with its development strategy and email distribution plan.

- Examination of industry-specific business. Here it is necessary to take into account two characteristics of the company, namely its macro industry and micro industry (specialization).

Digital marketing has such advantages:

- Allows you to more deeply introduce a fairly limited target audience with your product. This can be done using contextual advertising by creating joint projects with the most popular portals for the target audience.

- Allows you to reach a susceptible group of people. It uses social media, mobile and Internet applications.

- Allows you to talk about the product to as many potential consumers as possible. It uses Internet PR and non-standard large-scale promotions.

- Constant communication with the consumer ensures the promotion of goods. Today the most common way to ensure constant contact with the audience is the social media space.

The disadvantages of digital marketing are:

- It is impossible to solve the problem in a short time.

- Digital marketing cannot rely on 2-3 tools. It is necessary to constantly expand this list.

- For the successful use of digital marketing, it is necessary to constantly improve the tools, check their effectiveness, analyze and make adjustments.

Support Information System may meet needs of digital marketing.

4. Modeling and designing of the digital marketing support system for agriculture

Structured Analysis and Design Technique (SADT) was applied to the analysis and design of the digital marketing support information system in the agricultural market.

In accordance with this methodology, the SADT model reflects the functional structure of the object, that is, the actions and relationships between these actions. As a result, the model of information marketing system was built according to the SADT methodology. At Fig. 1 the SADT context-level diagram presents.

The basic processes of the system are more fully reflected in the zero-level SADT diagram (see Fig. 2).

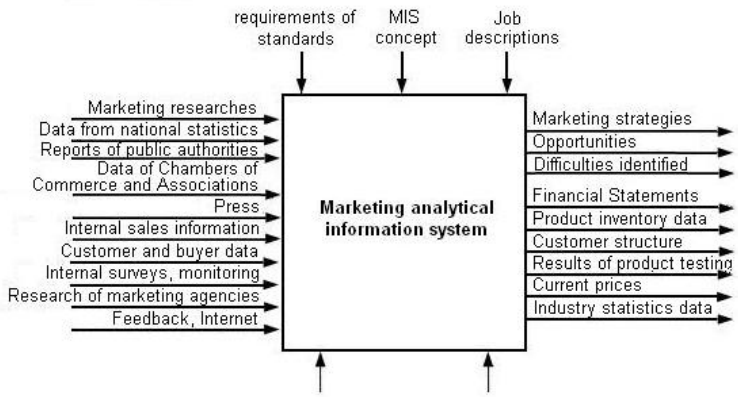


Fig. 1. Context-level SADT diagram.

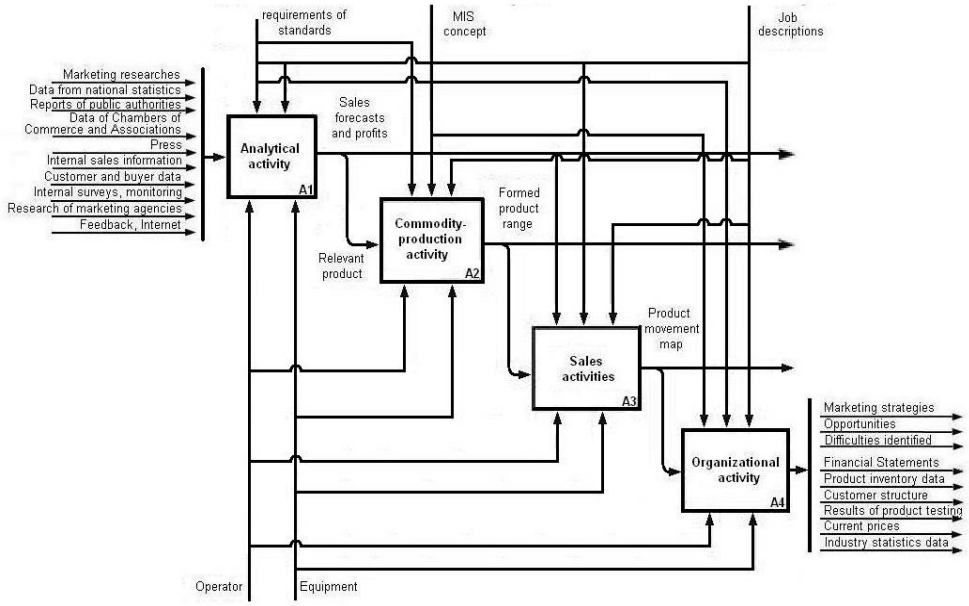


Fig. 2. Zero level SADT-chart

The next stage for system construction is the data flow diagrams (DFD) design (see Fig. 3). Such diagrams make it possible to display the data flows for the system being developed and to visualize the system working processes. Data flows in this case were considered both material and information flows. Material flows are raw materials, auxiliary materials, production products, equipment, vehicles and more. Information flows are order data, market status, availability of raw materials, stock sizes in warehouses, work performed, bottlenecks in manufacturing, manufactured products, cost data and expected level of prices, etc.

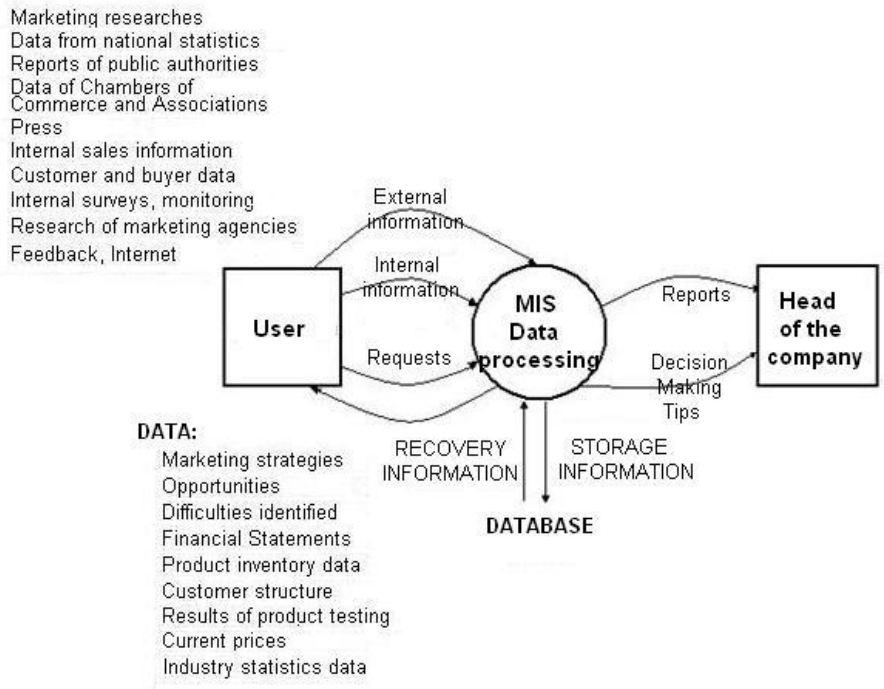


Fig. 3. Data flow diagrams (DFD diagram)

For system designing, the authors use standard of software graphic description - Unified Modeling Language (UML). The list of basic operations (functions) the system can perform at the Use Case Diagram (see Fig. 4) presents.

The system provides the following functions [9]: generalization of the learned information, the client and partner involvement, market analysis, internal marketing audit.

To build the internal architecture of the system, common standards for system modeling and documentation were defined. It were formed the formation of the main abstractions of the subject area and initial formation of architectural levels. Identification of the main abstractions based on the description of the subject area and the specifications of the requirements for the system allowed develop the scheme of data systems.

The following entities are implemented at this scheme:

- Customer - the person who wants to order agricultural products must have such attributes as surname, first name, gender, age, address and contact telephone number.
- Goods - agricultural products that have a category, price, item.
- Supplier - a company that supplies agricultural products.
- Warehouse - a place where agricultural produce is stored.
- Marketer - manager who sells and controls the database.
- Ordering is a requirement to purchase agricultural products.
- Delivery - providing the buyer with the goods at the specified address.

The Fig. 5 displays the corresponding ER-diagram for the developed system conceptual data scheme [10].

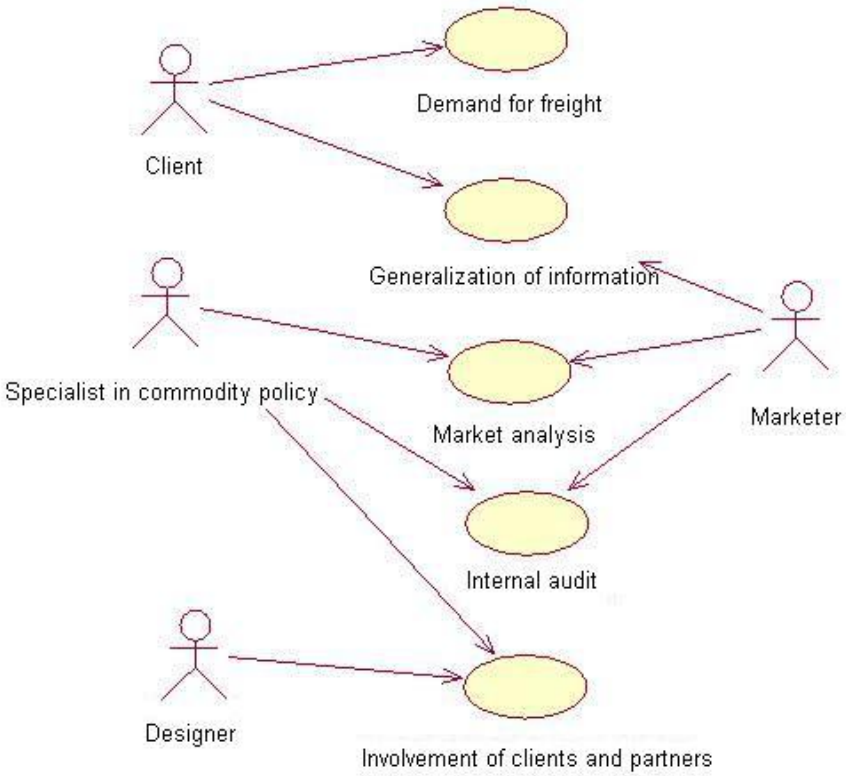


Fig. 4. User case diagram

We used Sequence diagrams to model the temporal aspects of the internal and external system protocols. Fig. 6 shows the ordered by the manifestation time sequence diagram of the object interaction for the developed system. This diagram describes the following sequence of actions: a marketer analyzes the market and deals with the design of advertising; the customer orders the products; the marketer processes the order; the contract is executed and the marketing plan is developed; orders are delivered to the customer.

At Fig. 7 presents the developed activity diagram to illustrate the actions per-formed by the system. According to it, the marketer analyzes the market, encourages potential customers. The customer chooses products, orders and pays for them. The marketing department prepares the contract, prepares the order, coordinates the order with the client, transfer the order to the commercial department.

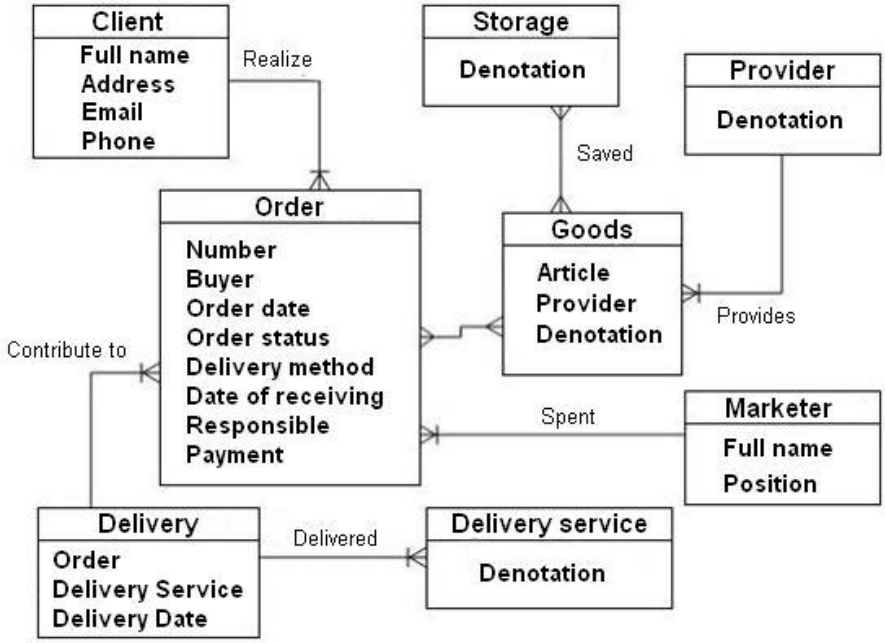


Fig. 5. ER-diagram of the digital marketing support system

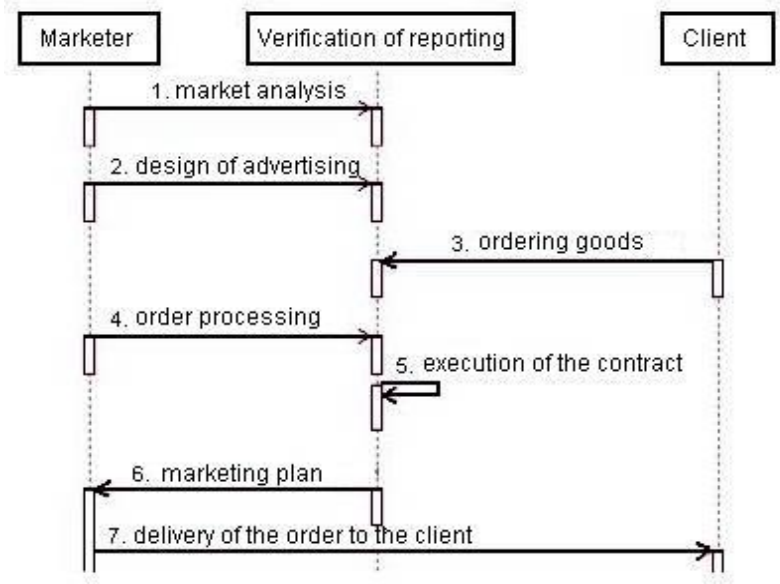


Fig. 6. Sequence diagram for order maintenance on the system

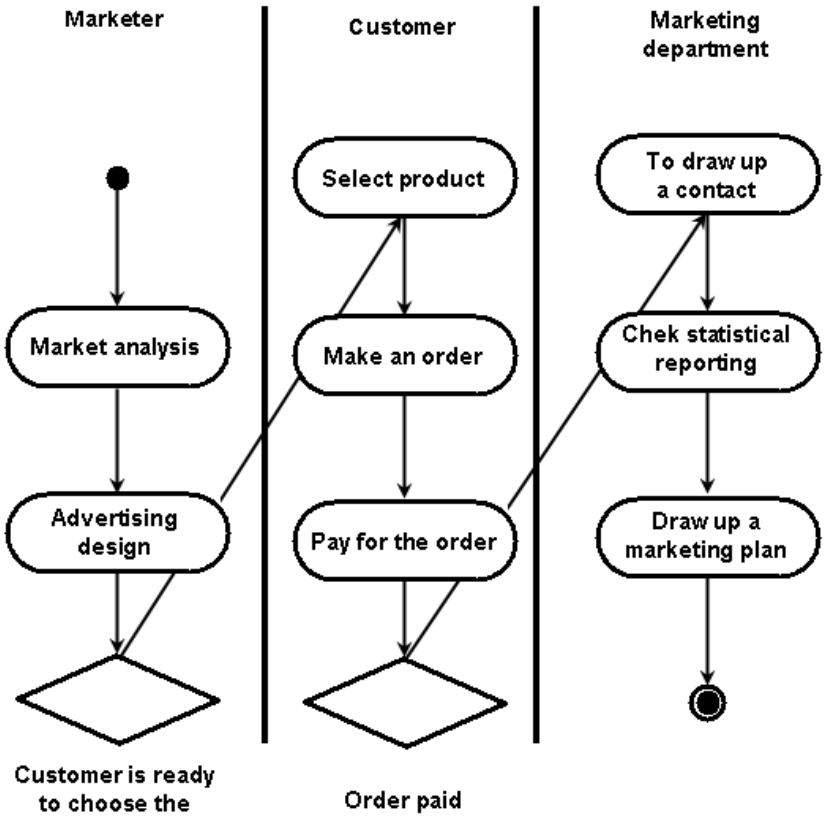


Fig. 7. Activity diagram.

5. Conclusion

The development of an information system to support digital marketing of agricultural enterprises is a relatively new and promising area for the development of information and telecommunications technologies. It provides prompt access to the marketing information of the enterprise in order to further analyze it. The purpose of such analysis is

- determining the needs and characteristics of the market,
- analysis of sales of goods,
- studying business trends,
- studying competitors' products,
- providing advisory information,
- automation of processes of commodity-marketing activity of the enterprise,
- implementation of short-term, medium-term and long-term forecasting,
- market response to a new product,
- policy formulation and distribution of goods.

The developed marketing information system is intended to support digital marketing in the agricultural market. It fully complies with the requirements and provides marketers with assistance in making commercially sound decisions, reducing costs and generating additional revenue.

References

1. Talvinen J.M. Information systems in marketing: Identifying opportunities for new applications European Journal of Marketing, 1(29), 10-14 (1995).
2. Chandra S. A., Subhashish S., Siddhartha D. Intelligent marketing information systems: computerized intelligence for marketing decision making. Marketing Intelligence & Planning, 2(13), 4-6 (1995).
3. Giaglis G.M., Mylonopoulos N.G., DoukidisI. The ISSUE methodology for quantifying benefits from information systems. Logistics Information Management, 1-2(12), 50–62 (1999).
4. Chaffey D. (2018) What is Digital Marketing? A visual summary. https://www.smartinsights.com/digital-marketing-strategy/what-is-digital-marketing/
5. DeVault G. (2019) Using Marketing Information Systems (MIS) https://www.thebalancesmb.com/what-is-a-marketing-information-system-mis-2296892
6. Crawford I.M. Food and Agriculture Organization of the United Nations. Rome, 1997
7. Gilorkar R. (2019) Marketing Information System: Introduction, Importance, Components and Advantages http://www.economicsdiscussion.net/marketing-2/marketing-information-system-introduction-importance-components-and-advantages/31600
8. Harmon R. Marketing information systems - Encyclopedia of Information Systems, Vol. 3 (2003) Elsevier Science (USA), 137-151.
9. Kotni V.V. D. P. A Study on Internal Mechanism of Marketing Information Systems - International Journal of Computer Science and Technology, 2(3), 64-67 (2011).
10. Berry, M. J. A., Linoff, G. Data Mining Techniques for Marketing, Sales and Customer Support, 3rd Edition. New York: Wiley, 1997, 888 p.

Influence of HRM Practices on Total Quality Management

Ivana Marinović Matović[1], Pedja Milosavljević[2]

[1] Addiko Bank AD, Niš, Serbia
[2] Faculty of Mechanical Engineering, Niš, Serbia
ivana.m.matovic@gmail.com, pedja@masfak.ni.ac.rs

Abstract. Total quality management is an imperative for successful and profitable business, whose importance is increased due to a rise in market competition. It is necessary to implement and nurture the total quality management process permanently, by following the development of business organization. Total quality management in global business environment has been recognized by customer orientation, continuous quality improvement, teamwork, focus on employees and stakeholders. The concept of total quality management directs its efforts to produce and deliver quality products and services, that is, fully meet consumer expectations. It achieves the greatest effectiveness when interacting with human resources management. Therefore, the purpose of this study was to point out that nowadays the relationship between human resource management and total quality management is necessary and significant. Primary link of this relationship is related to the attitude that employees are the only real connection between organization and end-users of products and services. Relationship between organization and employees is directly reflected in employee-customer relationship quality. To achieve business excellence, high importance should be attributed both to corporate culture of treating employees as the most significant asset, and to investment in quality trainings and education. Only organizations that nurture and develop a positive working environment, and manage relationship with their employees and stakeholders, can expect high quality products and services, thus ensuring customer satisfaction and a long-lasting business success. This study presented human resources management as a process of introducing changes in organization, with the aim of continual business efficiency improvement. Continual improvement is a basic prerequisite for implementation and development of total quality management concept. In this regard, various processes in human resource management at organizational level were analyzed. The findings of this study suggest that there is no organizational growth and development without successful implementation of total quality management concept, which further depends on the development of human resources management, as dominant organizational competitiveness factor.

Keywords. Total quality management, Human resources management, Quality, Human resources, Employees, Business organization.

1. Introduction

This paper starts from the basic premise that building quality human resources is halfway to achieving the full quality of products and services. Total quality management involves managing human resources in a way that enables the most efficient identification, development, maintenance and use of human resources in organization. It is the employees who create the quality, which explains the importance of human resources management for achieving the goal of contemporary business, achieving business excellence and world-class products and services (Bowen & Lawler, 1992).

Contemporary approach to human resource management (HRM), in its orientation towards complete quality, offers efficient methods and techniques especially for generic and development functions (selection, training and development, performance appraisal, rewards), as well as practically applicable solutions in the function of total quality. The implementation of these solutions directly affects the achievement of quality, its further generation and increase through employees. This means that total quality management (TQM) in organization, and thus its developmental and competitive capacity, depends primarily on human resources management.

Given that human resources, because of their specificity in relation to other organizational resources, represent a critical point in establishing and improving a system of total quality, the purpose of the paper is to answer the following: what is the interaction between human resource management and total quality management and how generic HRM processes support a system of complete quality. Since human resource management implies the way to achieve and improve the total quality in organization, the HRM principles and TQM should be complementary. These challenges must be adequately addressed by HRM theory and practice and provide managers with professional and practical support in responding to them. It is a challenge for contemporary management and requires the development of specific capabilities, skills and methods of managing human resources in organization. In line with such requirements, we can identify securing, developing, retaining and motivating, aligning HRM policies and organizational strategies, and driving change as a critical task for each manager.

2. The Importance of HRM in Total Quality System

Increasing the speed of change and volatility in international economic environment has led to changes in HRM value for achieving business excellence. Considering human resources management, the following elements were singled out as a natural consequence of such a situation: ensuring and retaining high quality human resources, quality and quantity of human factor, supporting and improving the execution of tasks at all levels, increasing knowledge and talents in organization, providing conditions for total quality management.

The process of globalization, technology development, hyper competition, increasing consumer demands, has led to a new model of human resources, based on the principles of total quality. In this context, the basic HRM goals are defined: competitiveness, profitability, efficiency, flexibility and human resources quality. Analyzing the hard and soft elements of TQM, Rees highlighted its connection to human resource management (Rees, 1994). Hard elements relate more to quantitative methods, techniques and tools such as total quality control, just-in-time production, six-sigma, total defect elimination. On the other hand, the soft elements of TQM program include HRM processes and activities that generate employee commitment to quality, continuous improvement, and customer-focused values. (Rees, 1994). Miller and Cardy (2000) propose that the processes of securing, developing and motivating employees should provide the quality human resources necessary to establish and improve a system of total quality (Miller & Cardy, 2000). Also, Dale and Cooper (1994) highlight the importance of human resources in achieving TQM goals. Processes such as training, development, team building, motivation, are key to enhancing TQM activities (Dale & Cooper, 1994).

Reward system for a total quality management also highlight the importance of human resource management for TQM. TQM awards provide a framework for achieving a high level of quality, based on TQM principles. They play an important role in promoting and rewarding excellence in organizational performance. They raise the quality standards of organizations. It is not certain that implementing a TQM in organization will lead to business excellence, but it

is very likely that organization has achieved business excellence if it is the award-winner (Table 1).

Table 1. Comparative analysis of purpose, core values and concepts and eligibility criteria of three world models of business excellence

Award	Purpose	Core values and concepts	Eligibility criteria
Deming Prize DP	Developing a unique TQM brand for organization fully customized to its specificities	• Top management • Process control • Kaizen promotion • Future planning	1. Policies 2. Organization 3. Information 4. Standardization 5. Human resources 6. Quality assurance 7. Maintenance 8. Improvement 9. Effects 10. Future plans
Malcom Baldrige National Quality Award MBNQA	Serves as a working tool for understanding and managing performance. The purpose is to assist in the effectiveness of organizational procedures, capabilities and results by facilitating communication by creating preconditions for benchmarking and sharing best practices among all types of organizations in the US	• Visionary guidance • Customer-centric excellence • Organizational and personal learning • Valuation of employees and partners • Mobility • Focus on the future • Innovation-driven management • Fact management • Social responsibility • Focus on results and value creation • System perspective	1. Leadership 2. Strategic planning 3. Customer and market orientation 4. Measurement, analysis and know-how management 5. Focus on human resources 6. Process management 7. Business results
European Foundation for Quality Management Award (EFQMA)	Recognizing the european or global role of models in their approaches and outputs, as well as providing independent feedback to organizations to help them continue the journey towards excellence	• Results orientation • Focus on clients • Leadership and constancy of intent • Process and fact management • Human development and involvement • Continuous learning and innovation and improvement • Partnership development • Social Responsibility	1. Leadership 2. Policy and strategy 3. People 4. Partnerships and Resources 5. Processes 6. Customer results 7. People results 8. Society results 9. Key Performance Results

Source: Samardžija, J., Kolak, G. (2009). Komparativna analiza tri najveća svetska modela poslovne izvrsnosti, *Ekonomski vjesnik: Review of Contemporary Entrepreneurship, Business, and Economic Issues*, 22(1): 211- 220.

Although each quality award has its own unique categories and quality criteria, there are some common areas, including human resource management practices. By looking at the ten primary elements in Deming Award (http://www.juse.or.jp/deming_en/) that serve as a checklist for quality assessment, we can conclude that employees are the most important resource for improving quality in organization. Investment in employee education and development, as well as the organization itself and management modes, occupy a special place in Deming Award, which confirms the importance of human resource management for achieving total quality.

The Malcom Baldrige National Quality Award criteria (https://www.nist.gov/baldrige) include key values and concepts, such as a focus on meeting customer needs, quality-based leadership, continuous improvement and learning, employee evaluation, information and analysis-based management, strategic quality planning. The importance of HRM for achieving quality in this model is particularly emphasized through the triad leadership. The triad leadership consists of strategic planning, consumer focus and human resource management.

The European Foundation for Quality Management (EFQM) award also emphasizes the importance of human resource management in achieving business excellence (https://www.efqm.org/). The essence of this award model is the assumption that excellent results are achieved through a partnership of resources and processes. The philosophy of this model is based on the superior performance achieved by involving employees in the organizational processes. The award model is divided into two parts: activators and results. Activators are leadership, employee management, strategy and policy, resources and processes. These five aspects make the business system and support the transformation of inputs into outputs. The results are human satisfaction, consumer satisfaction, impact on society and business results (Samardzija & Kolak, 2009).

A strong focus on human resources and HRM system can be found in different theories of total quality management. A comparative review of the most significant theoretical approaches to quality, the most prominent of which are Deming (2000), Juran (1964), Crosby (1980), Feigenbaum (1945), and Ishikawa (1968), can be concluded that all theorists, although in different ways, highlighted the importance of HRM issues for quality.

3. Human Resources Management in Total Quality Function

The quality importance, and overall quality issues in contemporary businesses, significantly influences organizational philosophy and management. New management practices, with quality in the center of interest, require significant changes in employee attitude and motivation, together with shifts in workplace behavior.

The primary link between total quality management and human resources management is reflected in the belief that employees are the basic link between organization and end-users of products and services. The organization's relationship with employees is directly reflected in relationship of employees to customers. This is especially evident in service organizations, which are growing in number in developed world, and occupy up to 90% of the total economy. It starts with the fact that human resources are not like other organizational resources. People have their own interests, goals, character and observation angle of reality. Employees are individuals with their own egos and are willing to invest their knowledge, skills and energy for organizational goals, in order to achieve their own goals, that is, to meet their needs. The organization and employees are in relationship of interdependence, from which both parties can profit under certain conditions (Šiber, 1999) The contemporary approach to human resource management offers a framework for successfully meeting the needs and goals of employees and aligning them with organizational goals.

Managerial quality challenges are management's ability to recognize the talents of employees, to compose team roles, to set challenging goals, to support the team in crisis, and to successfully manage not only formal organization, but also temporary and informal organizational forms.

Quality management implies a high level of professional competencies of employees, but even more their commitment (to quality, work, customers and organization). This involves good practices in decision-making process of human resources selection (recruitment, selection, orientation) and their further professional development (training, coaching, career management, performance appraisal), and even more decision-making process of identification of employees with organizational goals (job type, development opportunities, promotion, rewards, job security, etc.). Therefore, motivation of employees and HRM is an important segment of the TQM system.

All this suggests that connection with TQM, which results in orientation of HRM to total quality, requires its strategic dimension. Bowen and Lawler (1992) point out that the importance of employees in achieving and improving the TQM system offers the HRM a golden opportunity. This opportunity is primarily reflected in the recognition of strategic importance of HRM for TQM system in organization. By enhancing TQM, HRM directly influences organizational performances, just like other organizational functions which add value, such as marketing, production, finance, research and development (Bowen & Lawler, 1992).

The importance of quality HRM in organization lies in the fact that human capital is often a critical factor for effective total quality management. Since in contemporary organizations quality as strategic resource becomes more important than the strategy itself, its development can be stimulated only by adequate management, that is, quality management practice of human resources.

The need for strong and strategically significant cohesion of total quality management and human resource management is necessary to achieve organizational competitive advantage. Successful implementation of TQM concept requires professional procedures of human resources management, which ensure the required quality of human resources. Successful implementation of TQM concept requires the redefinition of basic methods and major tools in human resource management. Solution should be sought in the application of contemporary methods and techniques, which entails primarily intensive management education (education of decision makers) in this field, as well as those who professionally address the issues of employees in organization. This, at the same time, requires the adaptation of formal and non-formal educational programs, which will give more theory and practice to HRM.

4. The Role of Generic HRM in Total Quality Management

Human capabilities, motivation and engagement are not self-explanatory, so managers must discover and utilize this valuable potential. In this respect, the generic HRM functions are of importance for ensuring total quality. Therefore, it is necessary to consider what are the main functions and how they affect the quality development? The organization and processes of HRM essentially determine the attitude of employees towards the business, the organization, and its goals, and thus the possibility of quality development. HRM processes should provide an adequate number of employees who have the necessary capabilities and qualities, knowledge and motivation. Basically, all HRM processes have a strong motivational effect and thus affect better performance, positive impact on quality, and direct employees' behavior in a desired way.

The following functions are identified as generic HRM functions that directly affect quality: planning, attracting, developing and retaining of employees (Figure 1). For HRM to be a factor of quality development, it is necessary to model these functions based on strategic criteria, have motivational and development effect, and be focused on generating business performances. In designing HRM generic processes, the following principles should be followed (DeCenzo & Robbins, 1999): All processes should be interconnected and form a unique HRM concept; All processes should be based on unique strategic criteria; HRM processes should express the core values of organization and send specific messages to employees; HRM processes should be manager-managed with the professional assistance of HRM professionals.

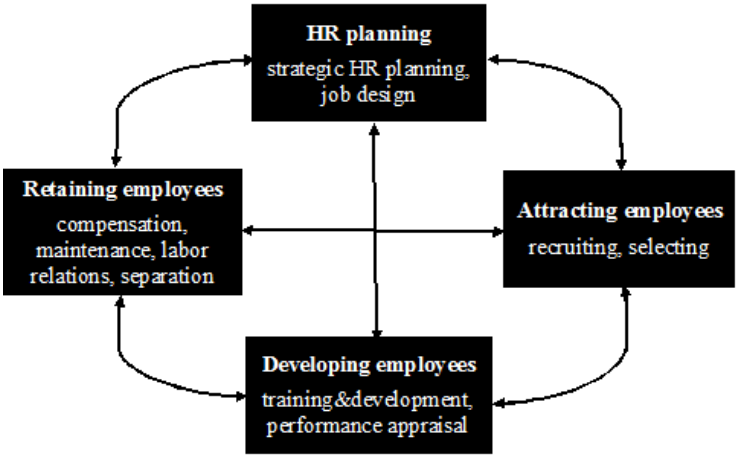


Fig. 1. Generic HRM functions
Source: https://www.slideshare.net/JubinJohny1/human-resource-70122137

HR planning is a very important HRM function that contributes to the achievement of TQM goals. Quality planning is basically employee planning. For a quality strategy to be implemented, organization must have the right number of employees with the right knowledge, skills and capabilities. Integrative linkage of quality strategy and human resources strategy, which is a key difference between HRM and personnel function, requires that HR managers look ahead, in the future, anticipating future needs, not only in number but also in quality of employees, i.e. knowledge, skills and capabilities that will carry the organizational development based on quality. As human resources have become critical resources for achieving total quality, their strategic planning is increasingly becoming a means of gaining competitive advantage. Little or no correlation between quality strategy and human resources strategy is a basic problem of low competitiveness.

Clear job designing, along with the profile of requirements, is one of the basic HRM processes known as job design, which enables risk reduction in decision-making regarding recruitment, selection, development, performance appraisal and reward. This process enables clarity of requirements and expectations with respect to employees, as a starting point in ensuring the desired quality-oriented work. This reduces the risk of misunderstanding of the employee role in terms of quality.

The recruitment and selection process, based on job requirements and strategic priorities of the organization, reduces the risk of wrong selection decisions and provides satisfactory inputs of knowledge, skills and overall human resources in organization, needed

to establish and develop the quality of products and services. Systematic activities on orientation and socialization of new employees reduce the risks of long-term and inadequate integration of new employees into the organization. The way in which organization attracts employees, the criteria it sets when recruiting, the way it introduces them to the work processes, depends on whether the quality employee, on which the business quality depends, will enter and remain in organization.

In order to achieve a competitive advantage based on high quality products and services, it is not enough to ensure that quality enters the organization, through the competencies of employees. Establishing and developing total quality requires the active role of managers in identifying, developing and directing knowledge, capabilities and skills, that is, the quality of employees' potential. Today's organizations operate in dynamic environment and total human knowledge is increased by geometric progression. Organizations risk obsolescence of existing competencies if they do not invest into ongoing development of employees. This process provides flexibility, multifunctionality and professional mobility within the organization. Training, rotation, enrichment, job expansion, career management are significant processes of employee development and thus primarily influence the quality development in organization. This reduces the risks of adapting to new job requirements and providing replacement in the event of employee fluctuation or absence.

Total quality management assumes an appropriate performance appraisal, i.e. evaluation of work quality of employees in organization. The process of performance appraisal plays a special role in HRM, which assumes continuous monitoring, evaluation and directing of work results and work behavior, based on appropriate assessment criteria, methods and systems. Evaluating work performance, as a process with a primary development dimension, allows timely understanding of deficiencies in the quality of existing knowledge, skills and capabilities. By systematic and appropriate criteria based on employee performance appraisal, managers could identify corrective measures in a timely manner and thus reduce the risks of poor performance.

By creating a competitive and motivational compensation system, organization demonstrates how much it values quality. If organization values quality, then it also needs to confirm that with the compensation system. To be able to manage total quality, the compensation system must provide a link between performance and reward, provide feedback to the employee on work and commitment, ensure the attractiveness of reward, transparency and fairness of reward criteria.

The extent and manner of employment termination can significantly affect the organizational stability, both in terms of costs and in terms of preserving key competences as a factor of quality. The organization therefore needs to keep this process under control and align with its own interests. Maintaining core competencies in organization involves monitoring job satisfaction and identifying factors of dissatisfaction, which should be a guideline in HRM activities. The belief that employees can be quickly and easily replaced in organization is a big misconception and such practices usually end in poor business results. Today, many organizations are very active in attracting talent and potential, so it is often the case that entire teams cross over into competition, which leaves a huge gap in business processes. It is one of the biggest risks to quality development in organization. On the one hand, it takes time to find the right people and fill important jobs, and on the other, the competition gets knowledge and information. Finally, activities that improve the safety and health of employees reduce the risks of absenteeism, replacement costs and reduced work potential, seen as risks in quality development.

Conceptual model representing the right mix of HRM practices to TQM effectiveness is given in Figure 2. The recruitment and selection process, teamwork and employee

empowerment, training and development, performance appraisal and compensation appear to be an indicator for organization to apply a designed approach system and methods, and it provides impact on the organization performance (Usrof & Elmorsey, 2016).

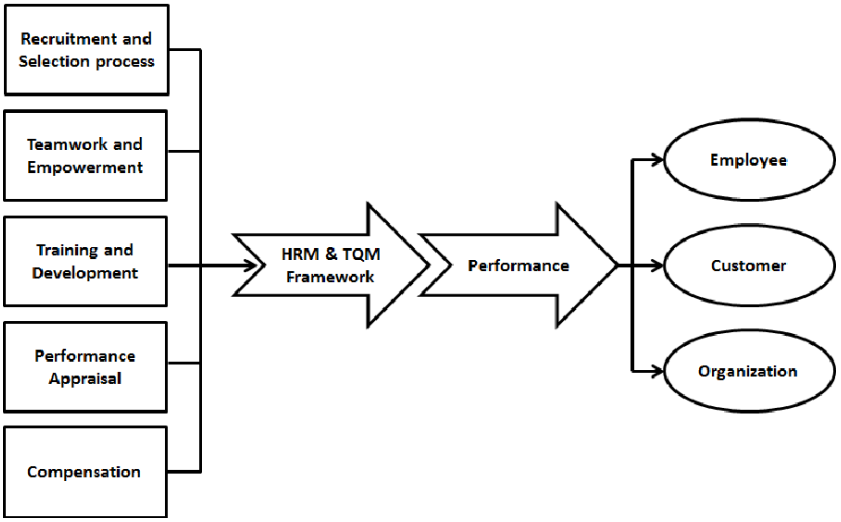


Fig. 2. Conceptual model of HRM impact on TQM

Source: Usrof, H.J.H., Elmorsey, R.M. (2016). Relationship between HRM and TQM and its Influence on Organizational Sustainability, International Journal of Academic Research in Accounting, Finance and Management Sciences, 6(2): 21–33.

It can be clearly observed that every HRM process has a clear impact on total quality management in organization. Employee is a key organizational resource and total quality management is strongly linked to the way human resources are managed. The contemporary HRM concept seeks to standardize generic management processes in this field, which are based on behavioral theories.

5. Conclusion

The main problem of contemporary organizations is how to be different or have something that makes the consumers decide to use their products and services, not some other. This diversity, which certainly ensures market success, is based on the quality of products and services, which becomes the basis of organizational competitive advantage.

The question of human resources importance for TQM system stems from the very nature of these resources. Much more complex, however, is the question of how this remarkable and complex resource, which significantly different from all other resources, set in the function of total quality. The answer to this question can be found in the analysis of factors that influence employee behavior in organization and contemporary management methods, which form the content of HRM concept as theory and practice. Therefore, the intention of this paper was to highlight human resource management as a very important factor for achieving total quality. The basic point made in this paper is that there is no growth and development of organization without successful implementation of TQM concept, which further depends on the development of human resources in organization, and that failure to comply with this results in low competitiveness. The main purpose of this paper was to

emphasize the importance of HRM orientation on quality for achieving the organizational success, and considering the system of total quality from a human resources perspective.

Based on all of the above, it can be concluded that the total quality system has become the strongest means of achieving a competitive advantage of organization, and human resources management is a path that ensures the support and development of quality, primarily through ensuring the excellence of employees' potential.

The importance of quality HRM in organization lies in the fact that human capital is often a critical factor for effective total quality management. Since in contemporary organizations quality as a strategic resource becomes more important than the strategy itself, its development can be stimulated only by adequate management, that is, quality management practice of human resources.

In conclusion, it should be said that total quality management, as a key resource for the competitive advantage of organization, must become a daily practice, that is, become part of organizational policy. The world's most successful organizations attribute their success, innovation and ability to adapt to changes, to successful management of their employees' potential. These potentials, that is, the quality of knowledge of employees, and especially that hidden, tacit knowledge, is considered the most valuable resource, and the proper management of it, is an activity that makes the difference between successful and unsuccessful.

References

1. Bowen, E.D., Lawler, E.E. (1992). *Total Quality – Oriented Human Resource Management*, Publication, Center for Effective Organizations, School of Business Administration, University of Southern California, Los Angeles.
2. Crosby, P.B. (1980). *Quality is Free*. New American Library, New York, USA.
3. Dale, B., Cooper, C. (1994). Introducing TQM: The Role of Senior Management, *Management Decision*, 32(1): 20-26.
4. DeCenzo D.A., Robbins, S.P. (1999). *Human Resources Management*. John Wiley & Sons.
5. Deming, W.E. (2000). *Out of the crisis* (1. MIT Press ed.). Cambridge, Mass.: MIT Press.
6. Feigenbaum, A.V. (1945). *Quality control: principles, practice and administration; an industrial management tool for improving product quality and design and for reducing operating costs and losses*, New York: McGraw-Hill.
7. https://www.efqm.org/
8. http://www.juse.or.jp/deming_en/
9. https://www.nist.gov/baldrige
10. https://www.slideshare.net/JubinJohny1/human-resource-70122137
11. Ishikawa, K. (1968). *Guide to Quality Control*. Tokyo: Asian Productivity Organization.
12. Juran, J.M. (1964). *Managerial Breakthrough: A new concept of the manager's job*. New York: McGraw-Hill.
13. Miller, J.S., Cardy, R.L. (2000). Technology and managing people: Keeping the "human" in human resources. *Journal of Labor Research*, 21(3): 447-461.
14. Rees, C. (1994). Teamworking and services quality: the limits of employee involvement. *Personnel Review*, 28 (5): 455-473.
15. Samardžija, J., Kolak, G. (2009). Komparativna analiza tri najveća svetska modela poslovne izvrsnosti, *Ekonomski vjesnik: Review of Contemporary Entrepreneurship, Business, and Economic Issues*, 22(1): 211- 220.
16. Šiber, F.B. (1999). *Management ljudskih potencijala*, Zagreb: Golden Marketing.
17. Usrof, H.J.H., Elmorsey, R.M. (2016). Relationship between HRM and TQM and its Influence on Organizational Sustainability, *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 6(2): 21–33.

Promoting Green Consumption Behavior: An Investigation of Environmental Awareness of College Students in China

Chih-Hsuan Huang[1,2], Hsin-Hung Wu[2,4], Li Li[5], Yan Song[6], Ying Gao[6], Zishuang Wang[6]

[1]School of Business Administration, Hubei University of Economics, Wuhan City, China
[2]Institute for Development of Cross-Strait Small and Medium Enterprise, Wuhan City, China
[3]Department of BA, National Changhua University of Education, Changhua City, Taiwan
[4]Department M-Commerce and Multimedia Applications, Asia University, Taichung, Taiwan
[5]Law & Business College of Hubei University of Economics, Wuhan City, China
[6]School of Business Administration, Hubei University of Economics, Wuhan City, China
tititacer@hbue.edu.cn, hhwu@cc.ncue.edu.tw, lisat23@163.com
songyan9908@163.com, 1563473902@qq.com, 286113428@qq.com

Abstract. Green consumption behavior in daily life has become an important research topic in China. Particularly, several green courses and programs are added to strengthen the environmental awareness of college students at universities in recently years. In fact, college students can be the high-efficiency actors who are good at learning ability, creativity, and flexibility and adaptability. The aims of current study is to understand the green consumption in practice from the views of college students in China by extending the Theory of Planned Behavior. The results indicated that green consumption attitude, subjective norms, and perceived behavioral control had statistically significant impacts on intentions of green consumption, respectively. Moreover, green consumption atmosphere, green promotion, and green product were to found to have a significant and positive relationship with green consumption intentions. Promising directions for future research are outlined.

Keywords. Green consumption behavior, College students, Theory of planned behavior.

1. Introduction

In order to pursue sustainable development and management, the concept of green development has been emphasized on the 18th and 19th CPC National Congress in China. How to promote the green consumer behavior has become one of essential issues for the development of sustainability. Generally, green consumer behavior reflects any consumers' actions directed toward the remediation of environmental issues/problems [1,2]. The concept of green consumer behavior involves three aspects: first, advocate consumers to use eco-friendly products or improve public health through green products. Second, at the end of the consumption process, ensure that garbage disposal does not cause environmental pollution [3, 4]. Third, guide consumers to change their perception of consumption by inculcating respect for nature, pursuit of health, simultaneously, pursuit of quality of life, attention to environmental protection, and conservation of resources and energy to achieve sustainable consumption [4].

Many studies on behavioral research have demonstrated that the traditional continuum model of behavioral change (i.e. the Theory of Planned Behavior, TPB) have good explanatory power across a wide variety of decision-making contexts [5,6,7], but for the discussions of

college students in green consumption are still ambiguous. Therefore, the green consumption behavior of college students in China were considered as main subjects in this research. College students are suggested to be high-efficiency actors who are good at learning ability, creativity, and flexibility and adaptability. Additionally, college students might be important peer groups to encourage the members to be green in the family. It is assumed that if we know more about green consumption behavior of college students, educational programs, marketing strategies and promotional messages could be designed to promote the behavior more effectively. Following this introduction, the second section of this paper describes a review of the literature on theory of planned behavior. The third section then presents the methods used in the empirical study. Finally, the results are discussed.

2. Relevant Literature

Behavioral research is presented as a way to increase our knowledge about consumer behavior and explain factors that predict behavior in daily life. Many studies on behavioral research have demonstrated that the traditional continuum model of behavioral change (i.e. the Theory of Planned Behavior, TPB) have good explanatory power across a wide variety of decision-making contexts [8,9,10]. According to the TPB, behavioral intention is the best predictor of a given behavior and that intention is predicted by three conceptually independent determinants; attitudes towards the behavior, subjective norms and perceived behavioral control. Attitudes refer to the individual's overall positive or negative evaluations of performing a given behavior [11]. Studies in the environmental field have demonstrated that a positive relationship between attitude and intention exists [6,12]. A subjective norm is conceptualized as the social pressure that consumers feel about whether to perform a behavior or not [11,13]. Several studies have introduced subjective norms as an essential predictor of behavioural intention to perform environmental behaviours, including public transport [6,14], organic products [15]. Additionally, perceived behavioral control represents a consumer's beliefs about the presence or absence of factors that facilitate or impede the performance of a behavior. Consumers will have strong behavioral intentions to perform a specific behavior if they perceive that they can easily act on the behavior [13]. Prior studies have supported the notion that perceived behavioral control is a significant predictor of behavioral intention [16].

In the recent years, the TPB has been used in different areas of green consumer behavior analysis. There are many scholars in the green consumption areas such as green consumer market [17]; green hotel [18]; green organic products [19]; green travel [20]. However, there is little discussions of the promotion of college students' green consumption behavior in China. However, the factors suggested in the TPB appear not to be completely appropriate in the analysis and prediction of consumer environmental behavior. Studies argued that in order to enhance the college students' perceptions of green consumption, creating a green consumption atmosphere is critical [21]. Similarly, the development of green products and appropriate promotion are important factors for encouraging green consumption behavior [22,23]. In addition to three original TPB predictors, green atmosphere, green promotion, and green product were integrated into our proposed model. As shown in Figure 1, the hypotheses are proposed as follows:

H1: Green consumption attitude has a positive impact on green consumption intention
H2: Green consumption subjective norm has a positive impact on green consumption intention
H3: Green consumption perceived behavior control has a positive impact on green consumption intention
H4: Green consumption atmosphere has a positive impact on green consumption intention
H5: Green consumption promotion has a positive impact on green consumption intention
H6: Green consumption product has a positive impact on green consumption intention

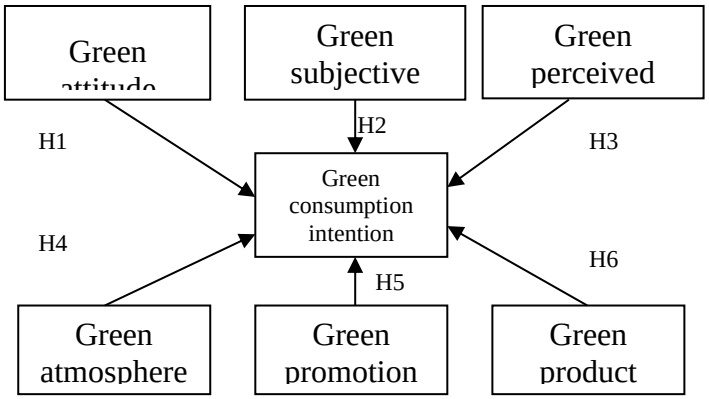


Fig. 1. Proposed Research Framework

3. Research Methods

Respondents are college students in Hubei Province in China. A total of valid 286 questionnaires were collected using convenience sampling. The questionnaires were designed according to the relevant literature in this study. Five items were used to measure attitude toward the behavior, subjective norms and perceived behavioral control, respectively. Green consumption atmosphere and green consumption promotion were both assessed using four items. Six items were adapted to evaluate green consumption product. Green consumption intention was adapted using four items. Five-point Likert scales anchored by 1 (strongly disagree) and 5 (strongly agree) were used throughout the questionnaire. After data screening, person correlation and regression analysis were employed using SPSS 20.0 to identify the proposed paths in the framework.

4. Results

As presented in Table 1, the examination of the sample demonstrated that the highest proportion of students is female (74.2 percent), majority of the students were juniors (45.4 percent) and business major (84.0 percent).

Table 1 Demographic information of the respondents (n = 286)

Demographic Variable	Frequency	Percentage
Gender		
Male	74	25.80
Female	212	74.20
Grade		
Freshmen	2	0.60
Sophomores	105	36.50
Juniors	130	45.40
Seniors	49	17.50
Major		
Business	240	84.00
Law	25	8.50
Science	10	3.50
Language	6	2.10
Art	5	1.90

Table 2. The Results of Mean, SD, Cronbach's Alpha

Research constructs	Mean	SD	Cronbach's α
Attitude	17.934	2.826	0.734
Subjective norm	19.626	3.409	0.951
Perceived behavior control	18.576	3.039	0.794
Atmosphere	15.549	2.293	0.766
Promotion	15.276	2.246	0.852
Product	15.762	2.216	0.847
Intention	18.305	2.153	0.931

As shown in Table 3, we conducted a Pearson's Correlation Analysis to realize the relationship between five constructs [25]. The results of Pearson's Correlation Analysis demonstrated that behavioral intentions were highly significant to green product and green promotion, respectively. Subjective norm was highly significantly related to green promotion. Additionally, green product is highly positive to green promotion [26].

Table 3 The Results of Pearson's Correlation Analysis

Construct	1	2	3	4	5	6	7
1. Attitude							
2. Subjective norm	0.587**						
3. Perceived behavioral control	0.547**	0.476*					
4. Atmosphere	0.365**	0.424**	0.459**				
5. Promotion	0.536**	0.423**	0.620**	0.539**			
6. Product	0.466**	0.397**	0.548**	0.512**	0.656**		
7. Behavioral intention	0.478**	0.466**	0.531**	0.488**	0.623**	0.637**	

Note: *: $p < 0.05$; **: $p < 0.01$

The results of regression analysis showed that subjective norm ($\beta = 0.151$, $p = 0.006$) had a positive impact on behavioral intentions, which supports H2. Moreover, both green promotion ($\beta = 0.253$, $p = 0.000$) and green product ($\beta = 0.331$, $p = 0.000$) had a significant and positive relationship with behavioral intentions, providing support for H5 and H6 [27].

Table 4 The Results of Linear Regression

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Standard Error	Beta		
1	(Constant)	3.075	0.813		3.781	0.000
	Att	0.035	0.046	0.044	0.769	0.443
	SN	0.100	0.036	0.151	2.788	0.006
	PBC	0.064	0.043	0.086	1.478	0.140
	GAtm	0.022	0.052	0.023	0.428	0.669
	GPm	0.254	0.064	0.253	3.964	0.000

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Standard Error	Beta		
	GPd	0.337	0.069	0.331	5.662	0.000
	R^2=0.516					

Note: Att: green attitude; SN: green subjective norm; PBC: green perceived behavioral control; GAtm: green atmosphere; GPm: green promotion; Gpd: green product

The summary of the hypotheses results is shown in Table 5. Overall, green attitude and green perceived behavioral control do not have significant impact on behavioral intention. Four factors (i.e. subjective norm, green atmosphere, green promotion, and green product) are to have positive impact on behavioral intention in this study.

Table 5 The Summary of Hypotheses Results

Hypothetical Path	Expected sign	Research results
H1: Green attitude towards the behavior → Behavioral intention	+	Not Support
H2: Green subjective norm →Behavioral intention	+	Support
H3: Green perceived behavioral control→ Behavioral intention	+	Not Support
H4: Green atmosphere→ Behavioral intention	+	Support
H5: Green promotion→ Behavioral intention	+	Support
H6: Green product→ Behavioral intention	+	Support

4. Discussion

This paper provides a holistic view on green consumer behavior theory in terms of the TPB. The results show that consumers' behavioral intentions are determined by their subjective norms, green atmosphere, green promotion, and green product. Generally, Students have a basic understanding of green consumption, but the specific content of the green consumption is insufficient. On the basis of the results of this study, we make the following suggestions.

For policy makers, green consumption policy is not only the guideline to guide people's rational consumption, but also an important guarantee to promote the standardization of green consumption. Several issues should be further considered. First, systemic and standardized green product certification systems should be established. Second, the policy makers expand the promotion of green consumption concepts through media channels such as television, the website, and other social media.

For enterprise, green consumption is a kind of high-level consumption concept, which is highly related to environmental awareness and social responsibility. Several issues should be concerned. First, most college students find it difficult to identify markers such as the green product logo, which to a certain extent inhibit the green consumption of college students. Enterprise needs to pay attention to the development of green products, waste reduction, and product quality assurance to enhance corporate identity to encourage the college students. Second, the business model of green marketing should be developed to promote consumers' brand identification and competitive advantages.

For educators, the results indicated that the levels of green consumption of college students are still needed to enhance. Several issues are suggested. First, arranging more green consumption programs/courses in the universities to change the attitude and cognition of college students on green consumption, thereby promoting college students' green consumption behavior. Second, to build a green campus, several green activities are suggested to conduct on campus. For example, green consumption knowledge quiz activities or garbage classification knowledge competitions to improve college students' green consumption awareness

References

1. Kotler, P., Armstrong, G. *Principles of marketing*, Pearson, Upper Saddle River, Sydney, 2010.
2. Kazmi, S.H.H. Consumer behaviour & marketing communication, Excel Books, New Delhi, 2010.
3. Peattie, K. Green consumption: behavior and norms. *Annual Review of Environment and Resources*, 35, 195-228, 2010.
4. Connoly, J., Prothero, A. Green consumption: life-politics, risk and contradictions. *Journal of consumer culture*, 8, 117-145, 2008.
5. Huang, C.H., Lings, I., Beatson, A., Chou, C.Y.H. Promoting consumer environmental friendly purchase behaviour: a synthesized model from three short-term longitudinal studies in Australia. Journal of Environmental Planning and Management, 61(12), 2067-2093.
6. Huang, C.H., Hsu, W.C., Huang, K.I., Hsu, S.M., Huang, Y.C. The extension of the theory of planned behavior to predict the use of public transport. *Asia Journal of Business and Management*, 3(5), 418-429, 2015.
7. De Leeuw, A., Valois, P., Ajzen, I., Schmidt, P. Using the theory of planned behavior to identify key beliefs underlying pro-environmental behavior in high-school students: Implications for educational interventions. Journal of Environmental Psychology, 42, 128–138, 2015.
8. Jun, J., Arendt, S.W. Understanding healthy eating behaviors at casual dining restaurants using the extended theory of planned behavior. *International Journal of Hospitality Management*, 53, 106–115, 2016.
9. Han, H., Hsu, L.T.J., Sheu, C. Application of the theory of planned behavior to green hotel choice: Testing the effect of environmental friendly activities. *Tourism Management*, 31 (3), 325–334, 2010.
10. Pakpour, A.H., Zeidi, I.M., Chatzisarantis, N., Molsted, S., Harrison, A.P., Plotnikoff. R.C. Effects of action planning and coping planning within the theory of planned behaviour: A physical activity study of patients undergoing Haemodialysis. *Psychology of Sport and Exercise*, 12 (6), 609–614, 2011.
11. Ajzen, I. *Attitudes, personality, and behavior*. Open University Press, N.Y., 2005.
12. Paul, J., Modi, A., Patel, J. Predicting green product consumption using theory of planned behavior and reasoned action. *Journal of Retailing and Consumer Services*, 29, 123–134, 2016.
13. Ajzen, I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211, 1991.
14. Chen, C.F., Chao, W.H. Habitual or reasoned? Using the theory of planned behavior, technology acceptance model, and habit to examine switching intentions toward public transit. *Transportation Research Part F*, 14, 128–137, 2011.
15. Tseng, W.C., Chang, C.H. 2 A study of consumers' organic products buying behavior in Taiwan-ecologically conscious consumer behavior as a segmentation variable. *International Proceedings of Economics Development and Research*, 84, 43-48, 2015.
16. Abrahamse, W., Steg, L., Gifford, R. Vlek, C. Factors influencing car use for commuting and the intention to reduce it: A question of self-interest or morality? *Transportation Research Part F*, 12, 317-324, 2009.
17. Dahlstrom, R.H. Green marketing management. Mason, South-Western Cengage, O.H, 2011.
18. Moreo, A. *Green consumption in the hotel industry: An examination of consumer attitudes*. University of Delaware, Newark, 2008.

19. Carrus, G., Passafaro, P., Bonnes, M. Emotions, habits and rational choices in ecological behaviours: The case of recycling and use of public transportation. *Journal of Environmental Psychology*, 28, 51-62, 2008
20. Vega-Zamora, M., Torres-Ruiz, F.J., Murgado-Armenteros, E.M., Parras-Rosa, M. Organic as a heuristic cue: What Spanish consumers mean by organic foods. *Psychology & Marketing*, 31(5), 349-359, 2014.
21. Zhang, L.Y. An analysis of the guiding strategy of green life consumption culture in China, *Western Finance and Accounting*, 422(5), 76-77.
22. Wu, B., Li, D., Xie, Z. Research on the influencing factors of consumers' preference for green products. *Soft Science*, 28(12), 89-94, 2014.
23. Xu, N.N. Research on the cultivation path of college students' green consumption concept. Shanxi University of Finance and Economics, Master Dissertation, 2018.
24. Nunnally, J.C. *Psychometric theory (2ed.)*. McGraw-Hill, N.Y., 1978.
25. Zikmund, W.G., Babin, B.J. *Essentials of Marketing Research*. 3rd ed. South-Western, Mason, O.H., 2007.
26. Cohen, J., Cohen, P., West, S.G., Aiken, L.S. *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum Associates, London, 2013.
27. Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A. *Applied regression analysis and other multivariable methods*. Duxbury Press, Belmont, CA, 1988.

The Role of Perceptions of Management Plays in Promoting the Development of Patient Safety Culture in the Healthcare Organizations

Chih-Hsuan Huang[1, 2], Hsin-Hung Wu[3, 4], Ruiqing Lv[5], Siqi Wang[5], Xuan Ding[5], Zixian Guo[5]

[1]School of Business Administration, Hubei University of Economics, Wuhan City, China
[2]Institute for Development of Cross-Strait Small and Medium Enterprise, Wuhan, China
[3]Department of BA, National Changhua University of Education, Changhua City, Taiwan
[4]Department M-Commerce and Multimedia Applications, Asia University, Taichung, Taiwan
[5]School of Business Administration, Hubei University of Economics, Wuhan City, China
tititacer@hbue.edu.cn, hhwu@cc.ncue.edu.tw, 1213934990@qq.com
1021664735@qq.com, 1241565882@qq.com, 1392197644@qq.com

Abstract. Patient safety culture is a fundamental factor to influence the system safety, and, more importantly, a positive patient safety culture allows medical staff to reduce medical errors, learn from failures, and commit to safer care. While there is a stream of research on the factors that affect patient safety culture, limited studies so far have concentrated on investigating the impact of management leadership on patient safety culture. The current study aims to recognize the influence of management leadership affect patient safety culture from the perceptions of medical staff by using Chinese version of Safety Attitude Questionnaire (CSAQ). Our results revealed that teamwork climate, stress recognition, and working conditions have direct effect on safety climate. Moreover, significant indict effects of the perceptions of management on safety climate was mediated through teamwork climate, stress recognition, and working conditions, respectively.

Keywords. Chinese version of safety attitude questionnaire, Patient safety culture, healthcare, medical staff.

1. Introduction

According to the annual report of the World Health Organization, 1 in 10 hospitalized patients experience medication-related harm due to adverse events. Several safety improvements have been conducted to bring the importance of patient safety culture in healthcare organizations globally [1]. The European Network for Patient Safety (EUNetPaS) in 2006 defined patient safety culture as "An integrated pattern of individual and organizational behavior, based upon shared beliefs and values that continuously seeks to minimize patient harm, which may result from the processes of care delivery" [2]. In other words, patient safety culture in healthcare organizations exhibits the extents of hospital management's commitments to patient safety as well as the care of medical quality.

One of the instruments that have been broadly used to assess the perceptions of patient safety culture from medical staff is the Safety Attitudes Questionnaire (SAQ) [3,4]. The original SAQ developed by Sexton et al. [5] has a six-dimension of patient safety culture, namely, teamwork climate, safety climate, perceptions of management, job satisfaction, stress recognition, and working conditions. Based on the original Chinese version of SAQ (CSAQ)

developed in 2007, the Joint Commission of Taiwan (JCT) integrated emotional exhaustion and work–life balance, to expand the existing CSAQ in 2014 [6,7] to more comprehensively measure patient safety culture. These two new dimensions concentrate the perceptions toward fatigue feelings and work-life conditions from the viewpoints of medical staff [7].

There is a stream of research on the factors that affect patient safety culture in the discipline of healthcare management. For example, Li [8] and Lee [9] both stated that teamwork climate is the most significant determinant to affect medical staff's attitude toward patient safety. Kim and Weng [10] incorporated the work experience of physicians and nurses into several patient safety-related constructs and further pointed out that management implement are able to improve safety climate by concentrating more on job satisfaction, teamwork climate for the more experienced staff, and working conditions for the less experienced. However, limited studies so far have concentrated on investigating the impact of management leadership on patient safety culture [11,12]. This study aims to echo the influence of management leadership affect patient safety culture in the healthcare organizations by using the instrument of CSAQ. Additionally, a mediation effect of the instrument of CSAQ is further investigated to draw insights on the mechanisms for the interaction between management leadership and patient safety culture.

2. Review of Patient Safety Culture

Patient safety culture is a foundation to achieve safety and quality of health service deliver. Previous studies argue that hospital staff with a positive attitude towards patient safety would produce positive health-related outcomes, such as solid physician-patient relationship [13,14], outstanding patient-oriented services [15,16], and harmonious safety atmosphere across hospital units [6,7]. In addition, hospital management are able to monitor the changes and trends of medical quality in a hospital by assessing the patient safety culture [17].

The Safety Attitude Questionnaire (SAQ), developed by Sexton et al. [5] has been broadly used to assess the perceptions of patient safety culture from the viewpoints of medical staffs [13,17]. The SAQ has 30 questions, divided into six dimensions, including teamwork climate, safety climate, perceptions of management, job satisfaction, stress recognition, and working conditions (As shown in Table 1). Teamwork climate refers to relationships and the degree of cooperation among staffs; safety climate refers to organizational commitment to patient safety; job satisfaction is about positive about work experience; stress recognition reflects stress factors that link to work performance; perceptions of management means administrator approval, and working conditions is based on the perceived work-environment quality [5,13,18].

In 2014, the JCT reconstructed the questionnaire to explicitly consider the emotional state of the medical staff during medial encounters. Three hospital-level aspects of safety culture (i.e. hospital management support for patient safety, teamwork across hospital units, and hospital handoffs and transitions) were removed whereas two new constructs (i.e. emotional exhaustion and work–life balance) were added to the questionnaire. EE describes a loss of passion for work, emotional exhaustion and feelings of depersonalization [19] whereas WB refers to balancing both work and other aspects of life [20]. As shown in Table 1, the latest Chinese version of the SAQ (CSAQ) consists 46 questions into eight constructs.

Table 1 The latest Chinese version of the SAQ

Teamwork climate	(1) Nurse input is well received in this clinical area
	(2) In this clinical area, it is difficult to speak up if I perceive a problem with patient care*
	(3) Disagreements in this clinical area are resolved appropriately (i.e. not who is right, but what is best for the patient)
	(4) I have the support I need from other personnel to care for patients
	(5) It is easy for personnel here to ask questions when there is something that they do not understand
	(6) The physicians and nurses here work together as a well-coordinated team
Safety climate	(7) I would feel safe being treated here as a patient
	(8) Medical errors are handled appropriately in this clinical area
	(9) I know the proper channels to direct questions regarding patient safety in this clinical area
	(10) I receive appropriate feedback about my performance
	(11) In this clinical area, it is difficult to discuss errors*
	(12) I am encouraged by my colleagues to report any patient safety concerns I may have
	(13) The culture in this clinical area makes it easy to learn from the errors of others
Job satisfaction	(14) I like my job
	(15) Working here is like being part of a large family
	(16) This is a good place to work
	(17) I am proud to work in this clinical area
	(18) Morale in this clinical area is high
Stress recognition	(19) When my workload becomes excessive, my performance is impaired
	(20) I am less effective at work when fatigued
	(21) I am more likely to make errors in tense or hostile situations (e.g. emergency resuscitation, seizure)
	(22) Fatigue impairs my performance during emergency situations
Perceptions of management	(23) Managers support my daily efforts
	(24) Managers do not knowingly compromise patient safety
	(25) I get adequate, timely information about events that might affect my work
	(26) The levels of staffing in this clinical area are sufficient to handle the number of patients
Working condition	(27) Problem personnel are dealt with constructively
	(28) This hospital does a good job of training new personnel
	(29) All the necessary information for diagnostic and therapeutic decisions is routinely available to me
	(30) Trainees in my discipline are adequately supervised
Emotional exhaustion	(31) I feel like I'm at the end of my rope*
	(32) I feel burned out from my work*
	(33) I feel frustrated by my job*
	(34) I feel I'm working too hard on my job*

	(35) I feel emotionally drained from my work*
	(36) I feel used up at the end of the workday*
	(37) I feel fatigued when I get up in the morning and have to face another day on the job*
	(38) Working with people all day is really a strain for me*
	(39) Working with people directly puts too much stress on me*
Work-life balance	(40) Missed meals
	(41) A hasty meal
	(42) All-day work without any rest
	(43) Individual or family plan change due to work factors
	(44) Poor sleep
	(45) Less than five-hour sleep at night
	(46) Work overtime

5. Research Methods

This study was conducted at Medical University Hospital in Taichung City, Taiwan. A CSAQ-based intra-organizational survey was used to investigate patient safety culture in the hospital in December 2017. The questionnaire has been confirmed to be reliable and valid in identifying medical staff's attitude toward patient safety in healthcare organizations, which contains 46 questions, grouped into 8 dimensions: teamwork climate (questions 1 to 6), safety climate (questions 7 to 13), job satisfaction (questions 14 to 18), stress recognition (measured through reverse scored questions 19 to 22); perceptions of management (questions 27 to 30), working conditions (questions 31 to 34), emotional exhaustion (measured through reverse scored questions 31 to 39) and work-life balance (questions 40 to 46). The measures of WB was excluded since its scales were designed using frequency data, which were different from an interval scale. A total of 1500 questionnaires were issued. The final usable questionnaires were 1037, yielding a response rate of 69%. Five-point Likert scales anchored by 1 (strongly disagree) and 5 (strongly agree) were used throughout the questionnaire. After data screening, a structural equation analysis was undertaken using the maximum likelihood estimation method to check the results of proposed paths (as proposed in Figure 1).

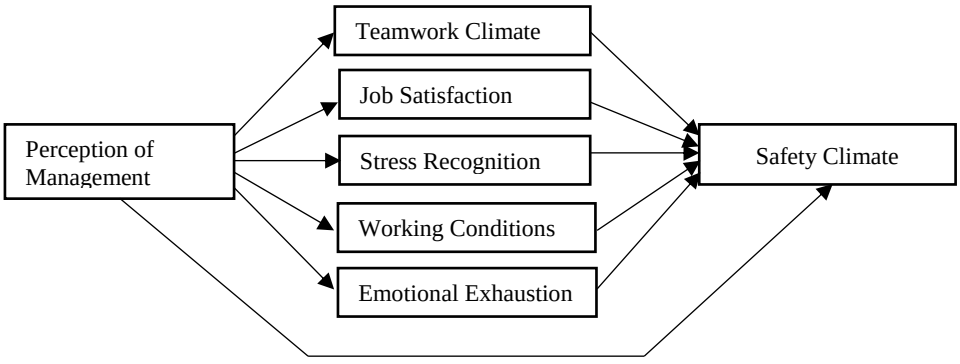


Fig. 1. Proposed framework

4. Results

As presented in Table 2, the examination of the sample profile showed that majority of respondents were female 80.2 percent), nurses (77.5 percent), age ranged from 21 to 30 years (41.7 percent), and education of the bachelor (83.9 percent) or graduate degree (15.6 percent), and more than 5 years working experiences were around 51.4 percent.

Table 2 Demographic information of the respondents (n = 1037)

Demographic Variable	Frequency	Percentage
Gender		
Male	205	19.8
Female	832	80.2
Age		
21-30 years old	432	41.7
31-40 years old	336	32.4
41-50 years old	190	18.3
51-60 years old	67	6.5
Over 61years old	11	1.1
Supervisor/Manager		
Yes	115	11.1
No	922	88.9
Job position		
Physician	233	22.5
Nurse	804	77.5
Education		
Senior High School	5	0.5
College/University	870	83.9
Master's Degree	162	15.6
Working experience in the hospital		
Less than 1 year	127	14.1
1 to 2 years	186	20.5
3 to 4 years	143	14.0
5 to 10 years	219	22.8
11-20 years	281	23.5
21 years and above	81	5.1

The structural model in Figure 2 meets satisfactory levels of overall goodness of fit, including acceptable absolute fit indices (χ^2/df = 4.065, GFI = 0.888, RMR = 0.051, and RMSEA = 0.054), incremental fit indices (TLI = 0.948, CFI = 0.954, and IFI = 0.954), and parsimonious fit indices (PNFI = 0.828 and PGFI = 0.736). All the standardized regression weights are statistically significant except for job satisfaction to safety climate, emotional exhaustion to safety climate, and perception of management to safety climate, respectively. Perception of management indicates the largest total effect on safe climate (0.914) and teamwork climate has the highest indirect effect on safety climate (0.705). Additionally, a significant indirect effect of perception of management on safety climate was found, which indicates that a noticeable mediation effect exists between perception of management and safety climate. By breaking down the indirect effect of perception of management on safety climate, 63.52 % is mediated through teamwork climate, 0.39 % through stress recognition, 29.77 % through working conditions [11].

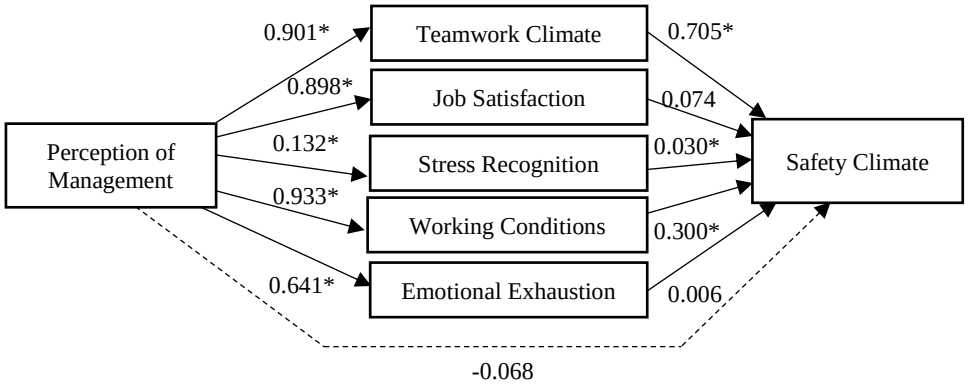


Fig. 2. Structural model

6. Discussion

In the current study, we examine the impact of management leadership on patient safety culture in the healthcare organizations by using the instrument of CSAQ. Based on the results, several managerial implements can be suggested. First, hospital managers should demonstrate their commitment decisively. In fact, management leadership on patient safety in hospitals can not only help medical staff better understand the organizational culture and leadership style, but also make them more clearly deal with patient safety issues in daily medical practice [21]. Both Huang et al. [6], and Lee et al. [22] argued that whether or not hospital has commitment to safety climate was highly related to the levels of cooperation among nurses.

Second, in line with Weng et al. [11], our study also argued that hospital managers should be aware that it would be less effective if it tries to improve patient safety culture by developing initiatives directly improve medical staff's perceptions of their leadership. It should focus on other mediator constructs such as teamwork climate, working conditions, and, stress recognition. Lee et al. [22] emphasized that hospital staff are more likely to focus on patient safety if more team building activities are developed. The degree of burnout and teamwork of the staffs are associated to healthcare-associated infection rate [23]. In work environment, high workload implies that nurses possess less resources in the working place [24]. The resources such as training program for nurses are benefit to decrease outcome such as burnout and perceived quality of care [25]. General respect for the health and wellbeing of nurses is a key part of the working conditions [26]. The present study confirms that human resource management has an important role to play in improving the patient safety climate, for instance by employing sufficient staff, by implementing a more rational salary system with performance-based bonuses, and by providing training programs for new personnel.

References

1. Lee, Y.C., Huang, C.H., Wu, C.F., Hsueh, H.W. A longitudinal study of identifying critical variables influencing patient safety culture from nurses' viewpoints in Taiwan. *Journal of Testing Evaluation*, 47(5), 3387-3398, 2019.
2. European Union Network for Quality in Health Care. Use of patient safety culture instrument and recommendations. European Network for Quality in Health Care, Office for Quality Indicators, 2006.
3. Patel, S., Wu, A. W. Safety culture in Indian hospitals: A cultural adaptation of the safety attitudes questionnaire. *Journal of Patient Safety*, 12(2), 75-81, 2016.
4. Soh, S.E., Barker, A., Morello, R., Dalton, M., Brand, C. Measuring safety climate in acute hospitals: Rasch analysis of the safety attitudes questionnaire. *BMC Health Services Research*, 16(1), 497, 2016.

5. Sexton J., Helmreich R., Neilands T., Rowan K., Vella K., Boyden J., Roberts P., Thomas E., The safety attitudes questionnaire: psychometric properties, benchmarking data, and emerging Research, *BMC Health Services Research*, 6(1), 44, 2006.
6. Huang, C. H., Wu, H. H., Lee, Y. C. The perceptions of patient safety culture: A difference between physicians and nurses in Taiwan. *Applied Nursing Research*, 40, 39-44, 2018.
7. Wu, C.F., Wu, H.H., Lee, Y.C., Huang, C.H. What attributes determine overall satisfaction in patient safety culture? An empirical study of the perceptions of hospital staff in Taiwan. *Journal of Testing Evaluation*, In press, 2019.
8. Li, A.T. Teamwork climate and patient safety attitudes: associations among nurses and comparison with physicians in Taiwan. *Journal of Nursing Care Quality*, 28(1), 60-67, 2013.
9. Lee, Y.C., Weng, S.J., Stanworth, J.O., Hsieh, L.P., Wu, H.H. Identifying critical dimensions and causal relationships of patient safety culture in Taiwan. *Journal of Medical Imaging and Health Informatics*, 5(5), 995-1000, 2015.
10. Kim, S.H., Weng, S.J. Incorporating work experience of medical staff into patient safety climate management: a multi-group analysis. *BMC Health Services Research*, 18(1), 919, 2018.
11. Weng, S.J., Kim, S.H., Wu, C.L. Underlying influence of perception of management leadership on patient safety climate in healthcare organizations–A mediation analysis approach. *International Journal for Quality in Health Care*, 29(1), 111-116, 2016.
12. Sammer, C.E., Lykens, K., Singh, K.P., Mains, D.A., Lackan, N.A. What is patient safety culture? A review of the literature. *Journal of Nursing Scholarship*, 42(2), 156-165, 2010.
13. Huang, C.H., Wu, H.H., Chou, C.Y.H, Dai, H.Y., Lee, Y.C. The Perceptions of physicians and nurses regarding the establishment of patient safety in a regional teaching hospital in Taiwan. *Iranian Journal of Public Health*, 47(6), 852-860, 2018.
14. Chi, C.Y., Huang, C.H., Lee, Y.C., Wu, H.H. Critical demographic variables on affecting patient safety culture from medical staffs' viewpoints. *Engineering Letter*, 27 (2), 76-83, 2019.
15. Laal, F., Fazli, B., Balarak, D., Dalir, F., Mohammadi, M., Mirzaei, R. Attitude toward the patient safety culture in healthcare systems. *Patient Safety & Quality Improvement Journal*, 4(2), 363-368, 2016.
16. Vincent, C. The essentials of patient safety. Blackwell Publishing, London, 2011.
17. Lee, Y.C., Zeng, P.S., Huang, C.H., Wu, H.H. Causal relationship analysis of the patient safety culture based on safety attitudes questionnaire in Taiwan. *Journal of Healthcare Engineering*, 2018.
18. Huang, C.H., Lee, N.N., Qu, B.B., Du, Q.Q., Chen, J.M., Li, L. Nurses' perceptions of patient safety culture in a tertiary hospital in China. *Forensic Science & Addiction Research*, 3(5), 1-4, 2018.
19. Poghosyan, L., Aiken, L.H., Sloane, D.M. Factor structure of the Maslach burnout inventory: an analysis of data from large scale cross-sectional surveys of nurses from eight countries. *International Journal of Nursing Study*, 46(7), 894-902, 2009.
20. Felstead, A., Jewson, N., Phizacklea, A., Walters, S. Opportunities to work at home in the context of work-life balance. *Human Resource Management Journal*, 12(1), 54-76, 2002.
21. Frankel, A., Grillo, S. P., Pittman, M., Thomas, E. J., Horowitz, L., Page, M., Sexton, B. Revealing and resolving patient safety defects: the impact of leadership WalkRounds on frontline caregiver assessments of patient safety. *Health Services Research*, 43(6), 2050-206, 2008.
22. Lee, Y. C., Huang, S. C., Huang, C. H., Wu, H. H. A new approach to identify high burnout medical staffs by kernel k-means cluster analysis in a regional teaching hospital in Taiwan. *Inquiry: The Journal of Health Care Organization, Provision, and Financing*, 53, 1-6, 2016.
23. Tawfik, D.S., Sexton, J.B., Kan, P., Sharek, P.J., Nisbet, C.C., Rigdon, J., Lee, H.C., Profit, J. Burnout in the neonatal intensive care unit and its relation to healthcare-associated infections. *Journal of Perinatology*, 37(3), 315, 2017.
24. Fagerström, L., Kinnunen, M., Saarela, J. Nursing workload, patient safety incidents and mortality: an observational study from Finland. *BMJ Open*, 8(4), e016367, 2018.
25. Paguio, J.T., Doris, Y.S.F. A mixed methods study to evaluate the effects of a teamwork enhancement and quality improvement initiative on nurses' work environment. *Journal of Advanced Nursing*, 2019
26. Gershon, R.R., Stone, P. W., Zeltser, M., Faucett, J., Macdavitt, K., Chou, S.S. Organizational climate and nurse health outcomes in the United States: a systematic review. *Industrial Health*, 45(5), 622-636, 2007.

Emotional Exhaustion Assessment from a Newly Established Hospital in Taiwan

Yi-Fang Lee[1], Yii-Ching Lee[2, 3], Jiunn-I Shieh[4], Chi-Hsuan Huang[5, 6], Hsin-Hung Wu[1]

[1]Department of BA, National Changhua University of Education, Changhua City, Taiwan
[2]Department Health Business Administration, Hung Kuang University, Taichung, Taiwan
[3]Superintendent Office, Ben Tang Cheng Ching Hospital, Taichung City, Taiwan
[4]Department M-Commerce and Multimedia Applications, Asia University, Taichung, Taiwan
[5]School of Business Administration, Hubei University of Economics, Wuhan City, China
[6]Institute for Development of Cross-Strait Small and Medium Enterprise, Hubei University of Economics, Wuhan City, China
lydia80124@gmail.com, yiiching.lee@gmail.com, jishieh@asia.edu.tw
tititacer@hbue.edu.cn, hhwu@cc.ncue.edu.tw

Abstract. Previous studies have shown that burnout could have negative impacts on healthcare quality and the increase of medical errors. Therefore, it is essential to assess medical staff's attitudes toward burnout in order to provide better medical services and reduce medical errors for patients particularly for a newly established university hospital in Taiwan. This study uses independent sample t test for mean differences and analysis of variance to examine if all of the medical staff with different demographic variables perceive emotional exhaustion from the Chinese version of the safety attitudes questionnaire statistically different. The results show that gender, supervisor/manager, experience in organization, experience in position, education, and direct patient contact are not critical demographic variables to influence employees' perceptions on emotional exhaustion statistically. In contrast, age, respondents reporting events in the past 12 months, job position, and job status are critical demographic variables that affect medical staff's perceptions in emotional exhaustion statistically. In summary, medical staff whose ages are 61 years old or more have significantly lower perceptions in emotional exhaustion. Medical staff who report 1-5 events in the past 12 months have a higher degree of burnout than those who do not report any events. Physicians seem to have lower emotional exhaustion perceptions compared with nurses, pharmacists, and others. Finally, agency medical staff have a higher degree of burnout than part-time and full-time medical employees.

Keywords. Emotional exhaustion, Chinese version of the safety attitudes questionnaire, Independent sample t test, Analysis of variance, Bonferroni.

1. Introduction

Burnout can be defined as losing enthusiasm for work, treating people as if they were objects, and having sense that the work is no longer meaningful [1]. Shanafelt et al. [2] further described that the burnout syndrome can affect workers in all fields, particularly those workers who involve intense interactions with people such as teachers, social workers, police officers, and healthcare workers. Several studies have shown that burnout would have negative influences on healthcare quality. For example, Shanafelt et al. [3] summarized physicians' overwork, stress, and fatigue are contributing factors to medical errors. West et al. [4] further

stated that physician burnout could result in negative effects on patient care, professionalism, and physicians' own care and safety such as diverse issues in mental health concerns and motor vehicle crashes, and the viability of healthcare systems such as reductions in physicians' professional work. Moreover, Cimiotti et al. [5] summarized that an increase in nurses' workloads could result in higher infections because nurse burnout is linked to job dissatisfaction and patient healthcare quality.

Lee et al. [6] pointed out that physicians and nurses are core staff in healthcare organizations having direct contacts with patients. Understanding their attitudes toward the patient safety culture particularly emotional exhaustion is essential to help hospital management enhance the patient safety culture of healthcare organizations and further reduce the potential medical errors when hospital staff having a higher degree of burnout are identified [7,8]. Huang et al. [9] stated that providing good medical services is important to a sustainable development of healthcare organizations in a highly competitive healthcare industry. For a newly established district hospital in Taiwan, it is essential for hospital management to not only observe the patient safety culture but also monitor the degree of burnout for all of the medical staff since a higher degree of burnout could result in negative influences on patient safety and healthcare quality.

Emotional exhaustion, which is one of three dimensions from Maslach burnout inventory-human service survey (MBI-HSS) and the mainstream measure to evaluate burnout, is included in the Chinese version of the safety attitudes questionnaire since 2014 [10,11]. Therefore, it would be of interest to observe how medical staff in this newly established hospital perceive in emotional exhaustion. If medical staff perceive a high degree of burnout, hospital management needs to take actions to reduce the burden in order to provide better healthcare quality and improve patient safety. This paper is outlined as follows. Section 2 reviews emotional exhaustion. Research method is described in Section 3. Results are summarized in Section 4. Finally, conclusions are provided in Section 5.

2. Review of Emotional Exhaustion

The patient safety culture assessment of healthcare organizations in Taiwan is requested by the Joint Commission of Taiwan since the development of the Chinese version of the safety attitudes questionnaire from medical staff's viewpoints annually [11]. The original Chinese version of the SAQ is developed based on the short form of the safety attitudes questionnaire by Sexton et al. in 2006, which has six dimensions and 30 questions [7,11]. The current Chinese version of the SAQ has eight dimensions along with 46 questions in total, and emotional exhaustion is included to assess medical staff's burnout since 2014 [11, 12]. Emotional exhaustion has nine questions depicted in Table 1 and is from Maslach burnout inventory-human services survey, which is a validated tool to assess burnout with three dimensions [12, 13]. These three dimensions are emotional exhaustion, personal accomplishment, and depersonalization [12, 13].

Each question in emotional exhaustion is worded negatively by a five-point Likert's type scale ranging from strongly agree to strongly disagree [11]. Thus, each respondent's answer needs to be reversed such that a smaller number represent a poor emotional exhaustion condition. In addition, each medical staff in a healthcare organization is required to answer all of the questions annually since 2014. That is, using the questions in emotional exhaustion can observe the degree of burnout from medical staff's viewpoints. The aggregated score from these nine questions can be viewed as a whole to present the level of emotional exhaustion for each medical staff.

Table 1 Nine questions in emotional exhaustion

Question
1. I feel like I'm at the end of my rope.
2. I feel burned out from my work.
3. I feel frustrated by my job.
4. I feel I'm working too hard on my job.
5. I feel emotionally drained from my work.
6. I feel used up at the end of the workday.
7. I feel fatigued when I get up in the morning and have to face another day on the job.
8. Working with people all day is really a strain for me.
9. Working with people directly puts too much stress on me.

3. Research Method

The purpose of this study is to observe how all of the medical staff in a newly established university hospital in Taichung City, Taiwan perceive emotional exhaustion. The internal survey data are from nine questions of emotional exhaustion dimension of the Chinese version of the SAQ conducted in 2017. All of the medical staff are required to fill out nine questions in emotional exhaustion by a 5-point Likert's scale ranging from strongly disagree with a numerical value of 1 to strongly agree with a numerical value of 5, respectively. By removing incomplete questionnaires and the option (selection) for each demographic variable having a frequency of one, the number of effective questionnaire in this newly established university hospital is 252. The demographic variables of the entire medical staff including physicians, nurses, technicians, pharmacists, medical administrators, and others in terms of gender, age, supervisor/manager, respondents reporting events in the past 12 months, job position, job status, experience in organization, experience in position, education, and direct patient contact are provided in Table 2.

The majority of the medical staff are female (77.8%), aged 21-40 (83.4%), not in charge of supervisors/managers (88.9%), reporting none events in the past 12 months (75.8%), nurses (61.1%), full time employees (95.6%), with either 1 to 2 years (45.6%) or less than 6 months (35.7%) in an organization, with either 1 to 2 years (44.0%) or less than 6 months (37.3%) in position, with a graduate school degree (85.7%), and very often to contact patients directly (79.0%). It is worth to note that this hospital was established in 2016, the options in experience in organization and experience in position have only three options, i.e., less than 6 months, 6-11 months, and 1 to 2 years, instead of six options.

IBM SPSS 19.0 software is used to analyze if all of the medical staff with different demographic variables have different perceptions in emotional exhaustion as a whole in this newly established university hospital. The scores from nine individual questions are aggregated into a dimension score to represent emotional exhaustion. In addition, independent sample t-test for mean differences and one-way analysis of variance (ANOVA) with $\alpha = 0.05$ are applied to nine questions in emotional exhaustion dimension. If the *p*-value is less than $\alpha = 0.05$, post hoc analysis is performed except for gender and supervisor/manager with only two levels. Bonferroni method that can reduce the probability of a Type I error and outperforms Scheffe method is chosen to perform the post hoc analysis in this study [14].

Table 2 Physicians' and nurses' demographic information of this medical center in 2017

Demographic Variable		**Frequency**	**Percentage**
Gender	1. Male	56	22.2
	2. Female	196	77.8
Age	1. Less than 20 years old	7	2.8
	2. 21-30 years old	133	52.8
	3. 31-40 years old	77	30.6
	4. 41-50 years old	20	7.9
	5. 51-60 years old	9	3.6
	6. 61 years old and above	6	2.4
Supervisor/Manager	1. Yes	28	11.1
	2. No	224	88.9
Respondents reporting events in the past 12 months	1. None	191	75.8
	2. 1-5	57	22.6
	3. 6-10	2	0.8
	4. 11-15	2	0.8
	5. More than 16	0	0.0
Job Position	1. Physician	13	5.2
	2. Nurse	154	61.1
	3. Technician	33	13.1
	4. Pharmacist	17	6.7
	5. Medical Administrator	24	9.5
	6. Respiratory Therapist	0	0.0
	7. Other	11	4.4
Job Status	1. Full Time	241	95.6
	2. Part Time	9	3.6
	3. Agency	2	0.8
	4. Contractor	0	0.0
Experience in Organization	1. Less than 6 months	90	35.7
	2. 6 to 11 months	47	18.7
	3. 1 to 2 years	115	45.6
Experience in Position	1. Less than 6 months	94	37.3
	2. 6 to 11 months	47	18.7
	3. 1 to 2 years	111	44.0
Education	1. Senior High School	0	0.0
	2. College/University	4	1.6
	3. Graduate School	216	85.7
	4. Doctoral Degree	32	12.7
Direct Patient Contact	1. No	24	9.5
	2. Rare	29	11.5
	3. Very Often	199	79.0

4. Results

Independent sample t test for mean differences is applied to test if all of the medical staff with different gender and supervisor/manager perceive emotional exhaustion differently. Levene's test is first applied to assess if the equality of variances assumption is hold for gender with $\alpha = 0.05$. The *p*-value is 0.002 which is far below 0.05 indicating the equality of variances assumption is violated. Thus, the result based on the inequality of variances assumption should

be chosen. In contrast to gender, Levene's test is used to assess if the equality of variances assumption is hold for supervisor/manager with $\alpha = 0.05$. The *p*-value is 0.309 showing the equality of variances assumption cannot be rejected. Thus, the result based on the equality of variances assumption should be chosen. Table 3 shows that the medical staff with different gender do not perceive emotional exhaustion statistically different. Besides, supervisor/manager is not significantly different for all of the medical staff in emotional exhaustion.

Table 3 Mean differences on emotional exhaustion

Demographic Variable	Assumption	t	Sig.	Post Hoc
Gender	Unequal Variance	-1.985	0.051	None
Supervisor/Manager	Equal Variance	-0.204	0.838	None

Table 4 summarizes that age has a significant influence on emotional exhaustion statistically. From the descriptive statistics, the medical staff whose ages are 61 years old and more have the lowest perception on emotional exhaustion. That is, they feel less stressful than the others in this newly established university hospital. The post hoc analysis as shown in Table 5 further depicts that the medical staff who are 21-30, 31-40, 41-50, and 51-60 years old report significantly better perceptions than those who 61 years old and more in emotional exhaustion.

Table 4 The ANOVA table for age

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	14.258	5	2.852	4.235	0.001
Within Groups	165.651	246	0.673		
Total	179.908	251			

Table 5 The post hoc test table for age

(I) Age	(J) Age	Mean Difference (I – J)	Sig.
21-30	61 and above	1.540	< 0.001
31-40	61 and above	1.450	0.001
41-50	61 and above	1.533	0.001
51-60	61 and above	1.691	0.002

Respondents reporting events in the past 12 months is a significant variable to influence medical staff's perceptions in emotional exhaustion statistically depicted in Table 6. From Table 7, medical staff reporting 1-5 events have statistically a higher perception in emotional exhaustion than those reporting none events in the past 12 months.

Table 6 The ANOVA table for respondents reporting events in the past 12 months

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	14.801	3	4.934	7.410	< 0.001
Within Groups	165.107	248	0.666		
Total	179.908	251			

Table 7 The post hoc test table for respondents reporting events in the past 12 months

(I) Events	(J) Events	Mean Difference (I – J)	Sig.
1-5	None	0.558	< 0.001

Medical staff with different job positions perceive emotional exhaustion differently in Table 8. Physicians have the lowest perceptions on emotional exhaustion from the descriptive statistics. Table 9 further summarizes nurses, pharmacists, and others have higher burnout than physicians. That is, nurses, pharmacists, and others are more stressful than physicians.

Table 8 The ANOVA table for job position

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	10.945	5	2.189	3.187	0.008
Within Groups	168.963	246	0.687		
Total	179.908	251			

Table 9 The post hoc test table for job position

(I) Job Position	(J) Job Position	Mean Difference (I – J)	Sig.
Nurse	Physician	0.818	0.011
Pharmacist	Physician	1.029	0.013
Other	Physician	1.175	0.010

Table 10 shows that job status is a critical demographic variable to significantly impact on emotional exhaustion statistically. The descriptive statistics show that the full-time medical staff have the lowest perception on emotional exhaustion, whereas the agency medical staff have the highest perception. To further perform the post hoc analysis depicted in Table 11, agency medical staff feel more stressful than both part-time and full-time medical staff statistically.

Table 10 The ANOVA table for job status

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	5.404	2	2.702	3.855	0.022
Within Groups	174.504	249	0.701		
Total	179.908	251			

Table 11 The post hoc test table for job status

(I) Job Status	(J) Job Status	Mean Difference (I – J)	Sig.
Agency	Full Time	1.650	0.018
Agency	Part Time	1.617	0.042

Employees who have different experience in organization do not report significant perceptions on emotional exhaustion as shown in Table 12. In addition, employees who have different experience in position do not have different perceptions on emotional exhaustion statistically depicted in Table 13. Moreover, from Tables 14 and 15, education and direct patient contact are not significant demographic variables to affect medical staff's perceptions on emotional exhaustion.

Table 12 The ANOVA table for experience in organization

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	2.569	2	1.285	1.804	0.167
Within Groups	177.339	249	0.712		
Total	179.908	251			

Table 13 The ANOVA table for experience in position

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	1.539	2	0.769	1.074	0.343
Within Groups	178.370	249	0.716		
Total	179.908	251			

Table 14 The ANOVA table for education

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	3.027	2	1.514	2.131	0.121
Within Groups	176.881	249	0.710		
Total	179.908	251			

Table 15 The ANOVA table for direct patient contact

	Sum of Squares	Degree of Freedom	Mean Square	F	Sig.
Between Groups	3.277	2	1.638	2.310	0.101
Within Groups	176.631	249	0.709		
Total	179.908	251			

In summary, gender, supervisor/manager, experience in organization, experience in position, education, and direct patient contact are not critical demographic variables to influence employees' perceptions on emotional exhaustion statistically. In contrast, age, respondents reporting events in the past 12 months, job position, and job status are essential demographic variables to impact medical staff's perceptions on emotional exhaustion statistically. Specifically, employees whose ages are 61 years old and more are less stressful. Their experience in reducing the burnout can be shared with the other employees in this case hospital. Physicians tend to have lower burnout than nurses, pharmacists, and others. In fact, nurses, pharmacists, and others feel more stressful from the descriptive statistics. The hospital management should take actions to reduce their degree of burnout. Finally, the agency employees feel the highest degree of burnout, while the full-time employees have the lowest degree of burnout in this case hospital by the descriptive statistics. In fact, the full-time employees have the lowest burnout statistically than agency and part-time employees. Further investigations should be examined in detail.

5. Conclusion

Assessing the medical staff's burnout is essential for hospital management particularly for a newly established district university hospital in Taiwan. A higher degree of burnout could result in higher medical errors and poor healthcare quality. This study uses nine questions in

emotional exhaustion dimension from the current Chinese version of the safety attitudes questionnaire to assess the degree of burnout from medical staff's viewpoints. The study identifies that age, respondents reporting events in the past 12 months, job position, and job status are critical variables to influence medical staff's perceptions on emotional exhaustion statistically. Besides, employees whose ages are 61 years old and more are less stressful than the others. Physicians have lower burnout than nurses, pharmacists, and others. Moreover, full-time employees have significantly lower burnout. The hospital management in this newly hospital should take actions to reduce the degree of burnout for employees with different ages except for those whose ages are 61 years old and more as well as nurses, pharmacists, and others in order to provide better healthcare quality and reduce medical errors to its patients.

References

1. T. D. Shanafelt, "Enhancing meaning in work: A prescription for preventing physician burnout and promoting patient-centered care," *The Journal of American Medical Association*, vol. 302, no. 12, pp. 1338-1340, 2009.
2. T. D. Shanafelt, L. N. Dyrbye, and C. P. West, "Addressing physician burnout: The way forward," *JAMA.*, vol. 317, no. 90, pp. 901-902, 2017.
3. T. D. Shanafelt, C. M. Balch, G. Bechamps, T. Russell, L. Dyrbye, D. Satele, P. Collicott, P. J. Novotny, J. Sloan, and J. Freischlag, "Burnout and medical errors among American surgeons," *Annals of Surgery*, vol. 251, no. 6, pp. 995-1000, 2010.
4. C. P. West, L. N. Dyrbye, P. J. Erwin, and T. D. Shanafelt, "Interventions to prevent and reduce physician burnout: A systematic review and meta-analysis," *Lancet*, vol. 388, pp. 2272-2281, 2016.
5. J. P. Cimiotti, L. H. Aiken, D. M. Sloane, and E. S. Wu, "Nurse staffing, burnout, and health care–associated infection," *American Journal of Infection Contro*l, vol. 40, pp. 486-490, 2012.
6. Y. C. Lee, C. L. Hsu, H. H. Wu, L. P. Hsieh, S. J. Weng, and 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. C. H. Huang, H. H. Wu, and Y. C. Lee, "The perceptions of patient safety culture: A difference between physicians and nurses in Taiwan," *Applied Nursing Research*, vol. 40, pp. 39-44, 2018.
8. C. H. Huang, H. H. Wu, C. Y. Chou, H. Dai, and Y. C. Lee, "The perceptions of physicians and nurses regarding the establishment of patient safety in a regional teaching hospital in Taiwan," *Iranian Journal of Public Health*, vol. 47, no. 6, pp. 852-860, 2018.
9. C. H. Huang, H. H. Wu, Y. C. Lee, and L. Li, "What role does patient gratitude play in the relationship between relationship quality and patient loyalty?" *Inquiry*, vol. 56, 0046958019868324, 2019.
10. Y. C. Lee, C. H. Huang, Y. C. Lin, and H. H. Wu, "Association rule mining to identify critical demographic variables influencing the degree of burnout in a regional teaching hospital," *TEM Journal*, vol. 6, no. 3, pp. 497-502, 2017.
11. Y. C. Lee, J. I Shieh, C. H. Huang, C. Y. Wang, and H. H. Wu, "Analyzing patient safety culture from viewpoints of physicians and nurses - A case of a regional teaching hospital in Taiwan," *Journal for Healthcare Quality*, vol. 39, no. 5, pp. 294-306, 2017.
12. Y. C. Lee, S. C. Huang, C. H. Huang, and H. H. Wu, "A new approach to identify high burnout medical staffs by kernel k-means cluster analysis in a regional teaching hospital in Taiwan," *Inquiry*, vol. 53, 0046958016679306, 2016.
13. C. Maslach, W. B. Schaufeli, and M. P. Leiter, "Job burnout," *Annual Review of Psychology*, vol. 52, pp. 397-422.
14. M. L. McHugh, "Multiple comparison analysis testing in ANOVA," *Biochemia Medica*, vol. 21, no. 3, pp. 203-209, 2011.

Identifying Care Target Combinations of Dementia Patients from a Medical Center in Taiwan by Classification and Regression Tree

Bing-Liang Lin, Hsin-Hung Wu

Department of Business Administration at National Changhua University of Education, Changhua City, Taiwan
lin8288show@gmail.com, hhwu@cc.ncue.edu.tw

Abstract. In the past, each care target is provided based upon the need from dementia patients without a comprehensive viewpoint. In practice, some care targets might be viewed as a combination for patients with dementia. This study uses classification and regression tree to find the combinations of care targets for patients with dementia in order to provide a person-centered care. Input variables include each care recipient's age, gender, and fifteen care targets developed by this case medical center, whereas four types of dementias are the target. The results show that the model has the highest accuracy value to predict patients with Alzheimer's disease. Thus, twenty-three out of thirty-one rules related to Alzheimer's disease can be the reference to form care target combinations. In general, an Alzheimer's disease patient with the age of 65-74 years who has very mild, mild, or moderate dementia requires appropriate schedule of activities. A male recipient needs to further consider introduction and referral of social resources. If a patient is less than 65 or 75 years old or more with very mild, mild, or moderate dementia, four types of care targets are required. If the patient is less than 65, 75-79, or 85-89 years old, introduction and referral of social resources should be included in the care target combination. The patient with mild dementia needs maintenance of personal hygiene, adjusting home environment safety, and regular outpatient follow-up treatment.

Keywords. Dementia, Patient with dementia, Care target, Classification and regression tree.

1. Introduction

Yan et al. [1] stated that caring for patients with dementia which is a huge task requires a health and social system to provide a wide range of care and services to meet the needs of dementia patients and their families. Song and Oh [2] and Kutsumi et al. [3] pointed out that behavioral and psychological symptoms of dementia (BPSD) is a source of psychological distress for caregivers and might cause difficulties for caregivers to care patients with dementia. In fact, these difficulties might lead to elder abuses and an increase of physical restraints [3]. In order to reduce the burden of caregivers and provide better healthcare services, Robinson et al. [4] recommended a palliative approach involving the families and healthcare staff to provide caregiving to patients with dementia.

Alzheimer's Disease International [5] concluded that quality care for dementia patients should be part of a call for action at global and country levels and the research investment should be balanced among prevention, treatment, cure, and palliative care. Obviously, taking great care of patients with dementia might be the focal point from now on when the rapid rise of dementia worldwide has become a trend. The study conducted by Yang et al. [6] identified the caring skills or related knowledge for caregivers including "activities of daily living",

"family care and social resources", "problem behaviors", and "cognitive impairments" in Taiwan. In addition, Kutsumi et al. [3] found four types of management techniques were required for staff training in order to provide better care for patients with dementia in Japan. Therefore, caregivers including care staff and families must have enough knowledge of dementia in order to offer appropriate ongoing care [1].

Clissett et al. [7] and Brooker and Latham [8] preached a person-centered care which is an ideal approach of caring for people with dementia in a long-term perspective. That is, the caring techniques or knowledge cannot be viewed individually. In contrast, the caring techniques or knowledge might be interconnected [9]. Therefore, there is a need to find essential associations of care techniques, dementia patients, and caregivers by providing an integrated approach rather than individual works to care for patients with dementia [10].

Previously, data mining has been applied to Alzheimer's disease by establishing a prediction model to assist medical professionals to predict the status of the disease based on the medical data about patients [11]. Kalyankar [12] stated that a decision tree which is a predictive modeling technique can be employed to perform classification, clustering, and prediction tasks. Moreover, decision trees have been proven to be the best algorithms for data classification in practice [13], [14]. On the other hand, fewer studies have been found to use data mining to focus on caring for dementia patients. For instance, Kutsumi et al. [3] applied the association rule to generate rules related to management techniques and BPSD so as to improve staff training. Hsu et al. [10] used classification and regression tree (CART) to identify a combination of care targets needed for each type of dementia patients when the input variables were care recipient's age, gender, and fifteen care targets and the output is the type of dementia.

The study conducted by Hsu et al. [10] only had the sample size of 173 patients. This study has the same sources of dementia patients from the same medical center in Taiwan but the sample of size is 506. Based on the same framework conducted by Hsu et al. [10], the purpose of this study is to use a larger sample size to identify what types of care targets are needed for each type of dementia by CART in order to provide better care for patients with dementia.

2. Research Method

A self-built database in a medical center in Taiwan is used with a total of 603 registered patients with dementia from October 2015 to April 2017. By removing incomplete data sets (46 patients) along with nine types of dementia patients whose frequencies are 10 or less (51 patients), the total number of dementia patients is 506. The demographic information about 506 patients is provided in Table 1. There are 172 male patients and 334 female patients. The majority of their ages falls in 75-79 and 80-84 years old (57%). The severity of patients with dementia measured by clinical dementia rating (CDR) can be classified into very mild dementia, mild dementia, moderate dementia, and severe dementia. In this study, the majority of dementia patients belong to mild dementia (51%). Four major types of dementias are found including Alzheimer's disease (71%), vascular dementia (18%), Parkinson's disease (5%), and mixed dementia (5%).

The column of data type depicted in Table 1 is to define the notations used in the analyses. For instance, M and F represent male and female patients, respectively. The numerical values of 1 to 7 in age represent the age groups from less than 65 years old to 90 years old and above. Four different types of CDR including very mild, mild, moderate, and severe are defined as 1, 2, 3, and 4, respectively. The input variables consist of each care recipient's gender, age, CDR, and fifteen care targets, whereas the target is the type of dementia. It is worth noting that these fifteen care targets depicted in Table 2 were developed

by this case medical center. The definitions and descriptions of these fifteen care targets can be found in the study conducted by Jhang et al. [15]. In this study, a value of zero is applied when this care target does not apply to a patient; whereas a value of one is used when this care target applies to a patient.

Table 1 The Information of Patients with Dementia

Demographic Variable		**Frequency**	**Percentage**	**Data Type**
Gender	Male	172	34	M
	Female	334	66	F
Age	Less than 65	17	3	1
	65-69	20	4	2
	70-74	56	11	3
	75-79	146	29	4
	80-84	144	28	5
	85-89	81	16	6
	90 and above	42	8	7
CDR	Very mild	104	21	1
	Mild	259	51	2
	Moderate	89	18	3
	Severe	54	11	4
Type of Dementia	Alzheimer's Disease	361	71	A
	Vascular Dementia	93	18	V
	Parkinson's Disease	26	5	P
	Mixed Dementia	26	5	M

Table 2 Fifteen Care Targets

No.	Description
1	Adequate nutrition and diet
2	Maintenance of personal hygiene
3	Appropriate schedule of activities
4	Adequate sleep and rest
5	Fall prevention
6	Preventing traffic accidents and getting lost
7	Preventing misusage of medications
8	Adjusting home environment safety
9	Regular outpatient follow-up treatment
10	Behavioral and psychological symptoms of dementia (BPSD) treatment
11	Referral to other specialists
12	Introduction and referral of social resources
13	Referral to family support groups and care skills training
14	Care for the mood of the caregiver
15	Education for dementia and BPSD

The purpose of this study is to identify what types of care targets (care target combinations) are needed for each type of dementia by classification and regression tree in order to provide better healthcare for patients with dementia. The sizes of training and testing data sets in this study are determined by an experiment. That is, the deigns of the training data

set are set from 50% to 90% by an incremental increase of 5% each time, while the designs of the testing data set are from 50% to 10% by an incremental decrease of 5% each time. The best combination will be determined by the ability to predict the number of types of dementia and the accuracy values of both training and testing data sets. Based on Table 3, the best combination of training and testing data sets are 60% and 40%, respectively, where the model has the ability to predict four types of dementias and has the overall better training and testing accuracy values. Classification and regression tree of IBM SPSS Modeler 18 with "simple" mode is chosen. The parameters of classification and regression tree are default values in the software. Specifically, the maximum surrogates and minimum change in impurity are set to 18 and 0.0001, respectively. The impurity measure for categorical targets is Gini. The stopping criteria are based on the percentage with minimum records in parent branch (%) of two and minimum records in child branch (%) of one.

Table 3 An Experiment Based on Different Training and Testing Data Sets

Training and Testing Ratio	**Accuracy of Training**	**Accuracy of Testing**	**Predictive Types of the Model**
50%:50%	76.17	66.80	4
55%:45%	75.71	69.47	3
60%:40%	74.34	67.82	4
65%:35%	75.08	66.47	3
70%:30%	77.62	63.40	3
75%:25%	74.80	67.67	3
80%:20%	76.98	59.13	3
85%:15%	74.94	52.75	3
90%:10%	75.40	61.19	3

3. Research Results

With the training and testing data sets of 60% and 40%, the results are depicted below. The tree depth by CART is nine. Eighteen variables are considered as important variables ranging from 0.268 to 0.014 in terms of importance values. The top four notches are age with the importance value of 0.268, CT(14) (Care for the mood of the caregiver) with 0.185, CDR with 0.152, and CT(10) (BPSD treatment) with 0.054, respectively, where CT is the abbreviation of care target. The predicted accuracy values of the model for both training and testing data sets are 74.34% and 67.82%, respectively, as shown in Table 4.

The coincidence matrices in Tables 5 and 6 shows that this established model can predict Alzheimer's disease well for both training and testing data sets with 91.2% (198/217) and 88.2% (127/144), respectively. In contrast, the accuracy values of predicting vascular dementia, Parkinson's disease, and mixed dementia for both training and testing data sets are well below 50%.

Table 4 Accuracy Values of the Model for Both Training and Testing Data Sets

Partition	Training Data Set	Testing Data Set
Correct	226 (74.34%)	137 (67.82%)
Wrong	78 (25.66%)	65 (32.18%)
Total	304	202

Table 5 Coincidence Matrix for Training Data Set

Training Data Set		Prediction			
		Alzheimer's Disease	Vascular Dementia	Parkinson's Disease	Mixed Dementia
Observation	Alzheimer's Disease	198	10	6	3
	Vascular Dementia	29	22	1	1
	Parkinson's Disease	8	4	3	1
	Mixed Dementia	13	2	0	3

Table 6 Coincidence Matrix for Testing Data Set

Testing Data Set		Prediction			
		Alzheimer's Disease	Vascular Dementia	Parkinson's Disease	Mixed Dementia
Observation	Alzheimer's Disease	127	13	2	2
	Vascular Dementia	29	10	0	1
	Parkinson's Disease	10	0	0	0
	Mixed Dementia	5	2	1	0

There are 31 rules generated including 23 rules for Alzheimer's disease, 6 rules for vascular dementia, 1 rule for Parkinson's disease, and 1 rule for mixed dementia. The CART prediction model only predicts Alzheimer's disease well but cannot accurately predict the other three types of dementias. Thus, Table 7 only summarizes 23 rules for Alzheimer's disease. For instance, the first two rules can be described as follows. A female Alzheimer's disease patient with 65-69 or 70-74 years old who has very mild, mild, or moderate dementia does not need "preventing traffic accidents and getting lost" but needs "appropriate schedule of activities." A male Alzheimer's disease patient with the age of 65-69 or 70-74 years who has very mild, mild, or moderate dementia does not need "introduction and referral of social resources" but needs "appropriate schedule of activities." From Table 7, CDR is the most essential variable in the rules followed by aged.

To sum up, an Alzheimer's disease patient with the age of 65-74 years has very mild, mild, or moderate dementia requires appropriate schedule of activities. In addition, a male recipient needs to further consider introduction and referral of social resources. When a patient is less than 65 or 75 years old or more with very mild, mild, or moderate dementia, four types of care targets are required, i.e., adequate sleep and rest, preventing traffic accidents and getting lost, referral to family support groups and care skills training, and care for the mood of the caregiver. Specifically, when the patient is less than 65, 75-79, or 85-89 years old, one additional care target is needed, namely introduction and referral of social resources. Further, when the patient has mild dementia, maintenance of personal hygiene, adjusting home environment safety, and regular outpatient follow-up treatment are required. When the patient has mild or moderate dementia, fall prevention is a required care target. When the patient has

the age of 80-84 or 90 years or more, BPSD treatment is needed. Finally, when the patient has severe dementia, fall prevention and care for the mood of the caregiver are critical care targets.

Table 7 Rules of Patients with Alzheimer's Disease

Description
CDR in [0,1,2] → Age in [2,3] → CT(3) in [1] → Gender in [F] → CT(6) in [0]
CDR in [0,1,2] → Age in [2,3] → CT(3) in [1] → Gender in [M] → CT(12) in [0]
CDR in [0,1,2] → Age in [2,3] → CT(3) in [1] → Gender in [M] → CT(12) in [1] → CT(5) in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [0] → CT(6) in [0] → CT(4) in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [0] → CT(6) in [0] → CT(4) in [1] → CT(13) in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [0] → CT(6) in [0] → CT(4) in [1] → CT(13) in [1
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [0] → CT(6) in [1] → CDR in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [0] → CT(6) in [1] → CDR in [1,2] → CT(3) in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [0] → CT(6) in [1] → CDR in [1,2] → CT(3) in [1]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [0] → Age in [1,6]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [1] → CDR in [1] → CT(2) in [0] → CT(8) in [0] → Gender in [F]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [1] → CDR in [1] → CT(2) in [0] → CT(8) in [0] → Gender in [M]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [1] → CDR in [1] → CT(2) in [0] → CT(8) in [1]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [1] → CDR in [1] → CT(2) in [1] → CT(9) in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [1] → CDR in [1] → CT(2) in [1] → CT(9) in [1]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [1] → CDR in [0,2] → CT(5) in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [1,4,6] → CT(12) in [1] → CDR in [0,2] → CT(5) in [1]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [5,7] → CT(10) in [0] → Gender in [F]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [5,7] → CT(10) in [0] → Gender in [M] → CT(4) in [0]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [5,7] → CT(10) in [0] → Gender in [M] → CT(4) in [1]
CDR in [0,1,2] → Age in [1,4,5,6,7] → CT(14) in [1] → Age in [5,7] → CT(10) in [1] → CT(13) in [1] → CT(5) in [0]
CDR in [3] → CT(14) in [0] → Age in [2,4,5,7]
CDR in [3] → CT(14) in [1] → CT(5) in [1]

4. Conclusions

Based on the study conducted by Hsu et al. [10], this study uses a larger sample size from the same database of a medical center in Taiwan to identify essential care target combinations for each type of dementia patients. The model established by classification and regression tree can accurately predict the patients with Alzheimer's disease with 23 rules. These rules can be the guidelines for caregivers to follow. Generally speaking, CDR is the most essential variable in the rules followed by age. When comparing the results conducted by Hsu et al. [10], our findings are somewhat different. First, the predicted accuracy values in our study are relatively lower. This might indicate the data are heterogeneous or the diagnoses of types of dementias by physicians might not be consistent in practice because there are several physicians who are in charge of diagnosing patients with dementia. Second, caring for patients with dementia might be very complicated. The associations among care recipients, caregivers, and care targets cannot be found in general due to the heterogeneous data set. That is, different approaches or philosophies might be used for analyses. One possible philosophy is to identify a combination of care targets for different types of dementias. For instance, one might pay much attention to the care target combination for each individual dementia rather than for a combination of different types of dementia for analyses. In doing so, the data might be more homogeneous. More importantly, a more customized care target combination can be identified to meet the special needs of patients with dementia based on a person-centered care philosophy.

References

1. G. J. Yan, W. F. Wang, K. M. Jhang, C. W. Lin, and H. H. Wu, "Associations between patients with dementia and high caregiving burden for caregivers from a medical center in Taiwan," *Psychology Research and Behavior Management*, vol. 12, pp. 55-65, 2019.
2. J. Song and Y. Oh, "The association between the burden on formal caregivers and behavioral and psychological symptoms of dementia (BPSD) in Korean elderly in nursing homes," *Archives of Psychiatric Nursing*, vol. 29, no. 5, pp. 346-354, 2015.
3. M. Kutsumi, M. Ito, K. Sugiura, M. Terabe, and H. Mikami, "Management of behavioral and psychological symptoms of dementia in long-term care facilities in Japan," *Psychogeriatrics*, vol. 9, no. 4, pp. 186-195, 2009.
4. A. Robinson, C. Eccleston, M. Annear, K. E. Elliott, S. Andrews, C. Stirling, M. Ashby, C. Donohue, S. Banks, C. Toye, and F. McInerney, "Who knows, who cares? Dementia knowledge among nurses, care workers, and family members of people living with dementia," *Journal of Palliative Care*, vol. 30, no. 3, pp. 158-165, 2014.
5. Alzheimer's Disease International, *World Alzheimer Report 2015 The Global Impact of Dementia: An Analysis of Prevalence, incidence, Cost and Trends*, Alzheimer's Disease International, London, 2015.
6. J. F. Yang, J. J. Wang, M. F. Yang, L. H. Chang, W. Y. Cheng, and M. C. Pai, "An integrated multi-disciplinary consultation clinic for dementia care," *Formosan Journal of Medicine*., vol. 19, pp. 394-401, 2015.
7. P. Clissett, D. Porock, R. H. Harwood, and J. R. Gladman, "The challenges of achieving person-centred care in acute hospitals: a qualitative study of people with dementia and their families," *International Journal of Nursing Studies*, vol. 50, no. 11, pp. 1495-1503, 2013.
8. D. Brooker and I. Latham, *Person-Centred Dementia Care: Making Services Better with the VIPS Framework*. 2nd ed. Philadelphia, PA: Jessica Kingsley Publishers, 2016.
9. W. F. Shih, C. W. Lin, W. F. Wang, and H. H. Wu, "Association rule mining of care targets from hospitalized dementia patients from a medical center in Taiwan," *Journal of Statistics and Management Systems*, vol. 21, no. 7, pp. 1299-1310, 2018.

10. Y. Y. Hsu, F. M. Chen, W. F. Wang, C. W. Lin, and H. H. Wu, "Using classification and regression tree to identify care target combinations for different dementia patients from a medical center in Taiwan," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 4, pp. 701-705, 2018.
11. D. Al-Dlaeen and A. Alashqur, "Using decision tree classification to assist in the prediction of Alzheimer's disease," *Proceedings of 2014 6th International Conference on Computer Science and Information Technology*, pp. 122-126, 2014.
12. N. V. Kalyankar, "Effect of training set size in decision tree construction by using GATree and J48 algorithm," *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2018*, 4-6 July, 2018, London, U.K., pp. 193-196.
13. N. Prasad L V and M. M. Naidu, "CC-SLIQ: Performance enhancement with 2k split points in SLIQ decision tree algorithm," *IAENG International Journal of Computer Science*, vol. 41, no. 3, pp163-173, 2014.
14. Y. H. Liu, K. Y. Kou, H. H. Wu, and Y. C. Nian, "Using classification and regression tree and dimension reduction in analyzing motor vehicle traffic accidents," *Applied Mathematics & Information Sciences*, vol. 10, no. 2, pp. 639-646, 2016.
15. K. M. Jhang, M. C. Chang, T. Y. Lo, C. W. Lin, W. F. Wang, and H. H. Wu, "Using the Apriori algorithm to classify the care needs of patients with different types of dementia," *Patient Preference and Adherence*, vol. 13, pp. 1899-1912.

Contemporary Trends in Wearable Lead-Free Composite Materials for Radiation Shielding- A Review

Digvijay Yadav[1], Mimangsha Sarma[1], Shannon Torcato[1], Shubhankar Shanware[1], R. Sivakumar[2]

[1] School of Electronics and Engineering (SENSE), Vellore Institute of Technology, India
[2]Department of Sensor and Biomedical Technology (SENSE),
Vellore Institute of Technology, Vellore, India
rsivakumar@vit.ac.in

Abstract. The paper reviews and analyzes contemporary trends in research in the field of non-lead-based aprons for radiation shielding for the purpose of extended duration of wear for physicians and increased comfort. It follows the emergence and development of composite radiation shielding materials that are lightweight and non-toxic. Attenuation coefficient(s) for the composite material(s) is analyzed and compared to that of conventional lead aprons and a suitable material is proposed. Promising directions for future research are outlined.

Keywords. Wearable lead free composite materials, Radiation shielding, Conventional lead aprons, Coefficients for the composite materials

1. Introduction

Radiation is the energy that propagates through matter or space. The energy of radiation can be either electromagnetic or particulate. Radiation can be classified into non-ionizing and ionizing. Of the two radiations, ionizing radiation is more dangerous as it has the ability to remove electrons from the atomic structure, this ionization causes a change in the DNA structure that can lead to genetic abnormalities. Radiation comes from many sources, natural(cosmic rays) or man-made(nuclear power plants, medical imaging procedures)

Symptoms of radiation sickness, which include fatigue, bleeding from the nose, skin burns show up when you're exposed to levels of more than 500 millisieverts (mSv), or half a sievert. More than 4 to 5 Sv is likely to be fatal. On the other hand, Radiation finds itself in a variety of applications in the medical field. From checking of fractures in bones to destroy tumor cells as a treatment for cancer in our body. One could say that ionizing radiation has been tamed by humans using the technology of the 21st century but one problem still persists, and that is the protection from it.

With the advancements in current medical imaging equipment, it is common for patients to have multiple tests for their medical problems. While such tests and procedures help physicians to recognize risks and intervene to avoid serious health issues, it is easy to forget that radiation is strong, cumulative, and proper procedures need to be taken if one were to be safe from it. Patients are also not the only ones affected. Exposure to radiation is a concern for doctors, ancillary staff and patients alike. As of 2012, in interventional cardiologists, nine cases of left-sided brain / head-and-neck tumors had been reported. This seems to support studies of doctors that use radiation therapy showed an increased risk of brain tumors.

The conventional method that we have been using in modern medicine is the use of Lead apron. Lead is known for its superior stopping power due to its high molecular density and high atomic mass, thus it is used radiation protection in applications ranging from X-Ray imaging and PET rooms to nuclear reactors. The only set of problems that arise with this accessory is that the material is heavy and toxic to the environment. This means that physicians who wear the apron for longer durations for performing interventional procedures, may suffer from fatigue and the disposal of these aprons after its lifespan (usually after 5 years) will be harmful to the environment.

This paper is an extensive review of this problem and proposal of a new wearable accessory that has a high attenuation coefficient, lightweight and non-toxic.

2. Current Trends

Growth in demand for radiation safety aprons in hospitals, development, and advancement in those products, and an increase in the number of diagnostic facilities facilitates the worldwide radiation safety aprons marketplace. The technological improvements in radiation safety apron and high or sudden thrust in the number of injuries and patient hospitalizations are the other main factors. The main constraints of the market are the development of other alternative technologies and severe opposition among existing producers. Moreover, product recalls, government guidelines for discount of healthcare fees, and the increase in demand via organization shopping corporations could hamper the marketplace economy.

With all these parameters taken under consideration for the development of effective radiation shielding aprons, these can be of 3 different types

1. Front Protection Aprons
2. Vest and skirt aprons
3. Additional aprons

The current aprons that are conventionally being used are lead-based aprons. However there have been attempts to develop certain light lead composite aprons with either bilayers or coating with textiles or nano-coating etc. lead-free aprons have also been developed with little or no lead been embedded in it, examples of it include the Bismuth-silicone based aprons, textiles based aprons like EarthSafe and Demron. Certain PVC based polymers for effective radiation shielding have been developed and have also shown Bio-compatibility properties as well.

3. Result

Textile Based Aprons

All materials had effectiveness at 100kev which is a clinically relevant exposure. 0.5 mm lead equivalency was demonstrated by materials EarthSafe and Xenolite at 80-100kev energy levels but were not able to demonstrate the same results beyond 100 keV energy level. Normal radiological operations were performed using energy levels beyond 100kev with the help of advanced imaging equipment. Demron was able to demonstrate the desired effect by shielding ionizing radiations beyond 100 Kev energy levels. All these materials were found to be light-weight.

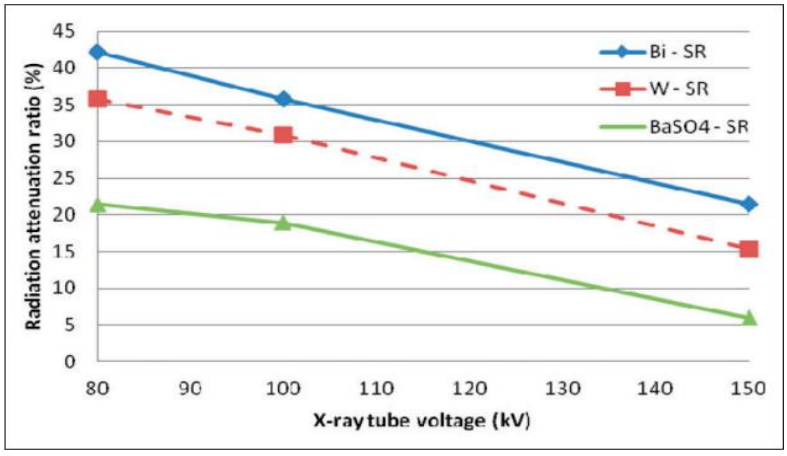
"Evaluation of non–lead-based protective radiological material in spinal surgery *Gaetano J. Scuderi et al., MD Georgiy V. Brusovanik et al., MD David R. Campbell et al., MD, Robert P. Henry et al., MD, Brain Kwon et al. Alexander R. Vaccaro et al., MD"*

Table 1
Comparison of transmission, attenuation and lead equivalencies

Energy (KeV)	Transmission (%)	Attenuation (%)	Lead equivalency (mm)
Demron			
60	0.55	99.45	0.39
80	2.21	97.79	0.44
100	3.14	96.86	0.54
120	4.00	96.00	0.55
130	4.50	95.50	0.54
EarthSafe			
60	0.46	99.54	0.42
80	1.47	98.53	0.53
100	3.67	96.33	0.51
120	6.78	93.22	0.42
130	8.43	91.57	0.39
Xenolite			
60	0.4	99.6	0.44
80	1.75	98.25	0.5
100	3.46	96.54	0.52
120	5.89	94.11	0.45
130	7.18	92.82	0.44
Lead			
60	0.23	99.77	0.5
80	1.80	98.2	0.5
100	3.50	96.5	0.5
120	4.80	95.2	0.5
130	5.20	94.8	0.5

Coated Textile Based Aprons

A textile that would be eco-friendly and flexible was chosen. These were used with tungsten, barium sulfate and bismuth to act as an alternative to lead aprons. Using the standard medical protection method the X-ray attenuation ratio was measured at 80,100 and 150kv energy. 1.55mm bismuth-based material at 60% weight ratio had attenuation of 90% for 100kv energy the same was achieved with 1.73mm of tungsten. Bismuth-silicone rubber blend reached better attenuation values.



"An alternative X-ray shielding material based on coated textiles
Nebahat Aral et al., F Banu Nergis et al. and Cevza Candan et al."

Bilayers

These materials had good attenuation properties because of low-Z upstream and high-Z downstream, the values obtained were much more than the ones in reverse order. X-ray beam quality also determined the optimal percentages of each metal in the bilayer.

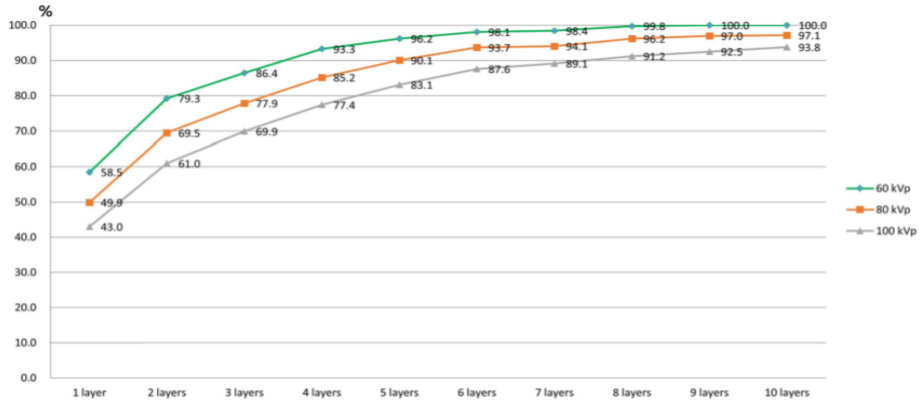
Biocompatibility estimation Non-Lead based X-ray shielding polymer

As per the conventional system for materials that can shield against x-rays, they are to be equivalent to the weight of Lead or of higher atomic mass. Certain composite lead-free materials were developed which displayed properties of Lead-based polymers. They had low attenuation at 400phr but high attenuation at 1000phr.

A composite polymer with a 0.25mm lead can be used. the lead-free contents of the polymer were able to achieve thermal-gravimetric analysis with high temperatures of 800 degrees at 92.4% for 1000phr. It also showed homogeneity implying that it had a uniform film structure without any aggregation. The lead-free PVC sheet was able to achieve standard tests for biocompatibility by undergoing tests of sensitivity on guinea pigs for clinical signs, mortality, etc.

Bismuth-Oxide

It was found that the radiation-attenuation for a single layer of Bismuth-Oxide was 58.5%,49.9% and 43.0% at tube voltages of 60,80 and 100 kVp, respectively, which follows a uniform dispersion of attenuation effectiveness. After lamination with multiple layers, it was found out that the radiation-shielding effect of Bismuth Oxide gradually increased with the increase in the layers. On conducting the flexural resistance test and abrasion strength test, the results obtained were a test performance of Class 6 and mlevel 3 respectively based on more than 100,000 bending tests and 2000-8000 abrasion cycles respectively.



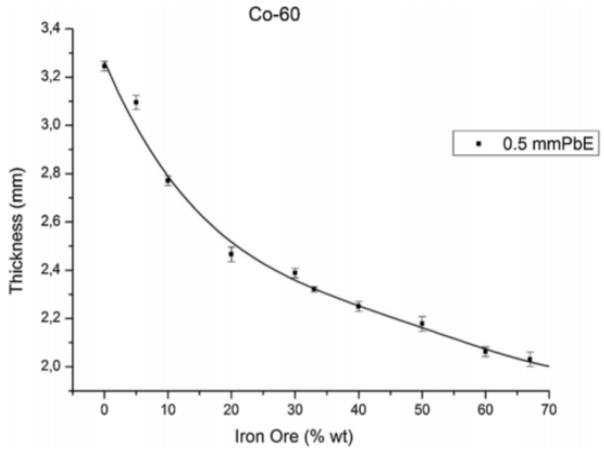
"Protection evaluation of non-lead radiation-shielding fabric: preliminary exposure-dose study .Ju Hee Kang et al"

Iron-Ore Imbedded Silicone Rubber Material

On increasing the iron ore content, lower relative intensity values were obtained. This concludes that greater shielding capabilities can be obtained with higher iron ore content.

The half-value layer (HVL), tenth-value layer (TVL), mean free path (MFP), and effective atomic number (Zeff) values of the materials were calculated. It was observed that the values of HVL, TVL, and MFP of the materials were dissimilar to those of pure lead, but

similar to those of 0.5mm PbE lead apron. It can be concluded that values to those of lead apron can be seen by increasing the iron ore ratio. It was obtained that the linear attenuation coefficient could be increased by embedding iron-ore into silicone rubber material. Iron-ore-embedded silicone rubber materials were compared with the 0.5 mm PbE lead standard for radiation protection against the Co-60 gamma source. The material thicknesses for 0.5 mm PbE show that iron-ore-embedded silicone rubber materials are good alternatives to lead as radiation shielding materials.



"Preparation and characterization of iron-ore-imbedded silicone rubber materials for radiation protection Bulent Buyuk"

Nano metal precipitates

Nano metal precipitates were dispersed in epoxy paints to evaluate their attenuation properties. Nano tungsten trioxide (15%) in addition to nano tin dioxide (85%) were the two most suitable candidates for radiation protection. The coatings made of combination nano tungsten trioxide (15%) and nano tin dioxide (85%) in an epoxy paint medium could be considered as a non-toxic, resilient to erosion, lightweight and with widely applicable for radiation shields. Radiation intensity after passing through various layers of paint is shown in below table using MCNP modeling. The highest attenuation was obtained in a combination of two non-lead metals (15% nano tungsten trioxide and 85% nano tin dioxide). The attenuation of lead shields was much lower than that of the non-lead paint layer. The densities of nano lead and nano tungsten trioxide-nano tin dioxide paints were 4.79 and 4.16 g/cm3, respectively.

"Novel paint design based on nanopowder to protection against X and gamma rays. Indian J Nucl Med .Movahedi MM et al."

100 kVp	I	I0	I/I0	% metals	% metals in epoxy paint	Density (g/cm³)
1	4.47×10^{-07}	7.71×10^{-06}	0.058	15 Nano WO_3, 85 Nano SnO_2	13.02 Nano WO_3 73.78 Nano SnO_2	4.16
2	5.31×10^{-07}	7.71×10^{-06}	0.069	100 WO_3	86.8 Nano WO_3	5.64
3	5.62×10^{-07}	7.71×10^{-06}	0.074	35 Nano WO_3, 65 Nano SnO_2	30.38 Nano $WO_3$56.42 Nano SnO_2	4.41
4	5.39×10^{-07}	7.71×10^{-06}	0.071	28.5 Nano WO_3, 71.5 Nano SnO_2	24.74 Nano WO_3 62.06 Nano SnO_2	4.34
5	6.01×10^{-07}	7.71×10^{-06}	0.073	25 Nano WO_3, 75 Nano SnO_2	21.7 Nano $WO_3$65.1 Nano SnO_2	4.29
6	6.86×10^{-07}	7.71×10^{-06}	0.076	100 Nano SnO_2	86.8 Nano SnO_2	3.77
7	7.24×10^{-07}	7.71×10^{-06}	0.090	100 Nano Pb	86.8 Nano Pb	4.79

MNCP: Monte carlo N-particle

Hydrogenated styrene –butadiene – styrene copolymer

Shielding ability of the new material can be improved by increasing the percentage of tungsten, but the bremsstrahlung will also get higher at the same time. The shielding ability of the new material with 85% tungsten is much larger than that of lead. The new material composed of tungsten and hydrogenated styrene –butadiene – styrene copolymer (SEBS) is capable of radiotherapy when protecting against electrons with low energy and thus also reduced the Bremsstrahlung radiation. It happens to be flexible thus can be easily customized into random shapes. However, when the incident energy is more than 12 MeV, the shielding ability of the material is worse than that of lead and cannot meet the requirement of radiotherapy. Though this is a factor not to be worried about because the highest energy of radiation used in the medical field does not usually exceed 511KeV. The tables offer supply credible proof of the MC calculation here. Shielding electrons with a similar energy, the larger the proportion of tungsten contained within the material, more energy is shielded.

Table 4. Ratio of transmitted doses with the new material (75% W + 25% SEBS) shielding.

	0.3 ± 0.01 cm	0.6 ± 0.02 cm	1.2 ± 0.04 cm	1.8 ± 0.06 cm
9 MeV	45%	80.2%	97.8%	98.5%
12 MeV	15.1%	46.5%	89.5%	96.4%
20 MeV	1%	4.1%	31.6%	64.2%

Table 5. Ratio of transmitted doses with the new material (80% W + 20% SEBS) shielding.

	0.3 ± 0.01 cm	0.6 ± 0.02 cm	1.2 ± 0.04 cm	1.8 ± 0.06 cm
9 MeV	60.4%	91.9%		
12 MeV	27.3%	67.1%	89.1%	96.4%
20 MeV	1.7%	12.2%	32.7%	73.5%

Table 6. Ratio of transmitted doses with the new material (85% W + 15% SEBS) shielding.

	0.3 ± 0.01 cm	0.6 ± 0.02 cm	0.9 ± 0.03 cm	1.8 ± 0.06 cm	2.7 ± 0.09 cm
9 MeV	32.3%	84%	95.4%	—	—
12 MeV	—	86.5%	95.8%	—	—
20 MeV	—	28.0%	56.6%	91.8%	94.3%

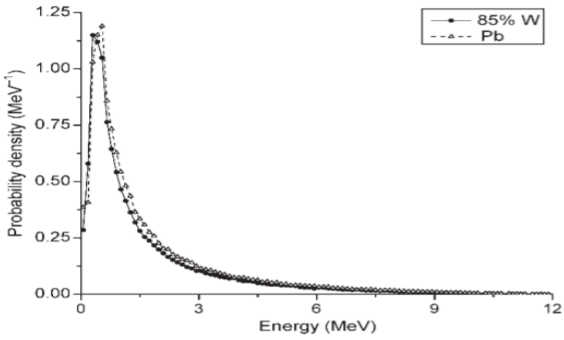


Figure 4. Energy distribution of transmitted photons for 12 MeV incident electrons.

"Radiation Protection Dosimetry, Volume 133, Issue 4, Kun Yue et al."

Moreover, shielding by a similar material, the fabric with additional thickness offers better shielding. In radiotherapy, it's needed that the protecting material ought to a minimum

of shield about 95% of the incident energy. The new material, particularly the fabric containing 85% tungsten in weight, is competent for radiotherapy once using electron incidence of 9 and 12 MeV. However, when shielding against electrons with higher energy, the shielding ability of the new material cannot meet the need. Therefore, the part of the fabric ought to be additional developed.

4. Conclusion

Lead-based aprons are one of the best methods of shielding against radiation however these are not environmentally friendly and also due to their weight, it also can cause back problems. so a textile that would be eco-friendly and flexible was chosen. these were used with tungsten, barium sulfate and bismuth to act as an alternative to lead aprons. Using the standard medical protection method the X-ray attenuation ratio was measured at 80,100 and 150kv energy. 1.55mm bismuth-based material at 60% weight ratio had attenuation of 90% for 100kv energy same was achieved with 1.73mm of tungsten.however bismuth-silicone rubber blend reached better attenuation values. If we were to make any choice of metal-embedded elastomer layers so as to decrease the weight of radiation shielding garments X-ray beam quality was used to determine the appropriate choice and as a result, metal-embedded elastomer layers were developed that had weight 25% lesser than their lead-based elastomer garments providing equal attenuation.

The materials must be light-weight, biocompatible, cost-effective, non-toxic and most importantly resistant to x-ray radiation. Some of the materials used are bismuth oxide, silicon rubber, tungsten oxides, hydrogen styrene, and PVC based polymer.

References

1. Buyuk, B et al. Preparation and characterization of iron-ore-imbedded silicone rubber materials for radiation protection. NUCL SCI TECH 29, 135 (2018)
2. Kang et al., J.H. et al., Oh et al., S.H. et al., J. et al. Protection evaluation of non-lead radiation-shielding fabric: preliminary exposure-dose study. Oral Radiol 35, 224–229 (2019)
3. S.M.Tajudin et al., A.H.A.Sabri et al., M.Z.Abdul Aziz et al., S.F.Olukotun et al., B.M.Ojo et al., M.K.Fasasi, et al.'Nuclear Engineering and Technology'.Vol.51, Issue 6,(September 2019).
4. Gaetano et al., J.Scuderi MD et al., Georgiy et al.,V.Brusovanik MD et al., David R.Campbell MD et al., Robert P.Henry MD et al., BrainKwon et al., Alexander R.Vaccaro MD et al. Vol.6. Issue 6,(September-October 2006).
5. Nebahat Aral et al., F.Banu Nergis et al., Cevza Candan et al. Volume: 86 issue: 8, page(s): 803-811, Article first published online: June 17, 2015; Issue published: May 1, 2016
6. Movahedi MM et al., Abdi A et al., Mehdizadeh A et al., Dehghan N et al., Heidari E et al., Masumi Y et al., Abbaszadeh M, et al. *Novel paint design based on nanopowder to protection against X and gamma rays.* Indian J Nucl Med 2014;29:18-21
7. Kun Yue et al., Wenyun Luo et al., Xiaoqing Dong et al., Chuanshan Wang et al., Guohua Wu et al., Mawei Jiang et al., Yuanzi Zha et al. A new lead-free radiation shielding material for radiotherapy, *Radiation Protection Dosimetry*, Volume 133, Issue 4, February 2009, Pages 256–260.
8. Anil Kumar Singh et al, Rakesh Kumar Singh et al, Bhupesh Sharma et al, Ajay Kumar Tyagi et al Characterization and biocompatibility studies of lead-free X-ray shielding polymer composite for healthcare application Received 12 December 2016, Revised 7 April 2017, Accepted 17 April 2017, Available online 28 April 2017.
9. Soares et al., F.A.P et al., Pereira et al., AG et al., & Flôr et al., R. de C. (2011). *Use of radiological protective clothing for absorbed dose reduction: an integrative literature review. Brazilian Radiology, 44 (2), 97–103.* doi: 10.1590 / s0100-39842011000200009

3rd Dimensional color X-ray based on Spectral Imaging – A Review

C. Parikshith[1], Rajat[2], R. Siva Kumar[3]

Dept. of Sensors and Biomedical Technology, School of Electronics Engineering,
Vellore Institute of Technology, Vellore, India
chavakula.parikshith2016@vit.ac.in, thakur.03rajat@gmail.com, rsivakumar@vit.ac

Abstract. From the Large Hadron Collider at the CERN laboratory to finding the God's particle; to the use in the latest trends of spectral Imaging break-through for a 3D X-ray diagnostic purpose. The paper reviews these new trends of medical imaging using spectroscopic imaging, and also the recent development in microelectronic technology which allows the design of imaging detectors with single photon processing capabilities. The paper studies the use of AISC's, other microelectronic technologies used in spectroscopic imaging. Its advantages and disadvantages compared to other traditional x-ray diagnostics.

Keywords. CERN, LHR, Spectral imaging, 3D x-ray diagnostics, AISC's (Application-specific integrated circuit), Medipix chips, Single photon detectors.

1. Introduction

2.1. Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest and most dominant particle accelerator. It comprises of the rings of superconducting magnets with various accelerating structures to support the energy of the particles en route. Inside the accelerator, two high energy particle shafts head out at near the speed of light before they are made to impact with each other. At that point, the beams travel on the contrary headings in discrete shafts and pipes which are kept at ultrahigh vacuum. At that point they are guided around the accelerator ring by a solid magnetic field kept up by superconducting magnets.

Since the particles accelerate at super high speeds, thus contain massive kinetic energies. These particles when are made to collide dis-integrate into many different particles with comparatively lesser energies. The LHR is set up with many highly sensitive particles/pixel detectors, which are capable of detecting these particles which disperse in nano-seconds. These detectors developed by CERN are able to detect these particles in such nano intervals of time. Not only that particle detectors need to have particles detection capabilities but also in a variety of spectrum (spectroscopic analysis) to be able to find new elementary particles. In 2012 the CERN facility and the collaborative teams successfully detected the Higgs Boson (a.k.a God's particle) which can give us information how our universe might have sprung to life [1] (Fig. 1)

2.2. Medical diagnostics

In the recent times x-ray spectroscopy found a major application in the medical field for diagnostic purposes. Unlike the traditional x-ray diagnosis available clinically. The amount of information that can be obtained is far greater using x-ray spectroscopy then a normal x-ray CT procedure. The 3D constructed images from a x-ray spectroscopy if far more superior than

a 3D image constructed using traditional x-ray imaging techniques (Fig. 2).

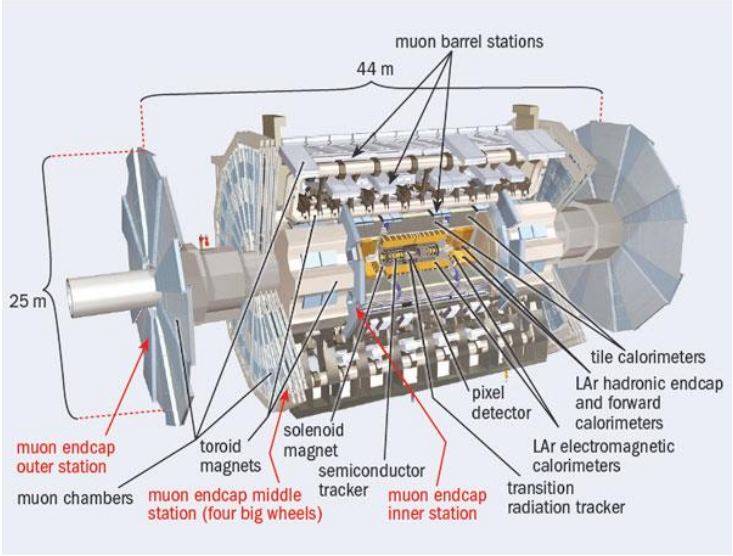


Fig. 1a. Particle detector set-up in CERN's atlas experiment.

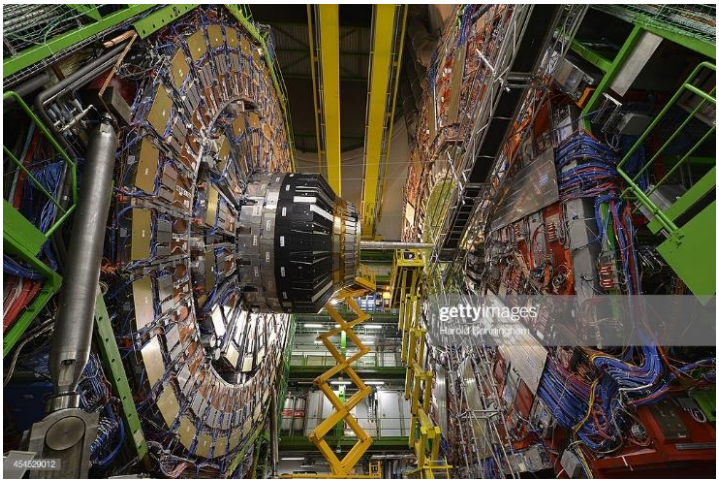
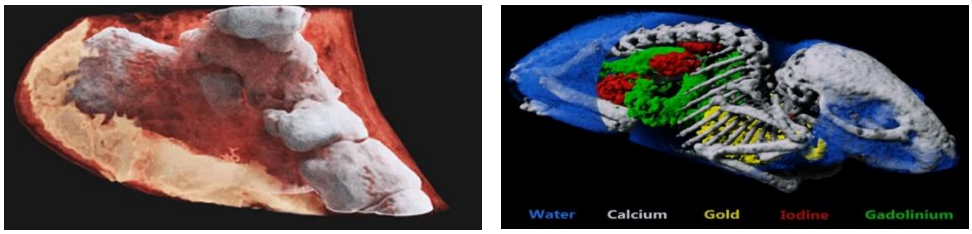


Fig. 1b.: image of actual LHR set-up and detector.

These particle pixel identifiers made out of the MEDIPIX3RX chips which were developed for the LHR can be utilized in the diagnostic purposes whenever utilized in a controlled manner utilizing the blend of spectral imaging. Likewise, this procedure utilizes the energy reliance of X-beam constriction to either expand the difference to-commotion ratio or to give quantitative picture information and diminish picture ancient rarities by supposed material decomposition. Spectral imaging isn't new yet has picked up enthusiasm for ongoing years on account of the quickly expanding accessibility of spectral CT, dual-energy CT and with the beginning of energy-settled photon-counting detectors [3, 4].



(2a) (2b)

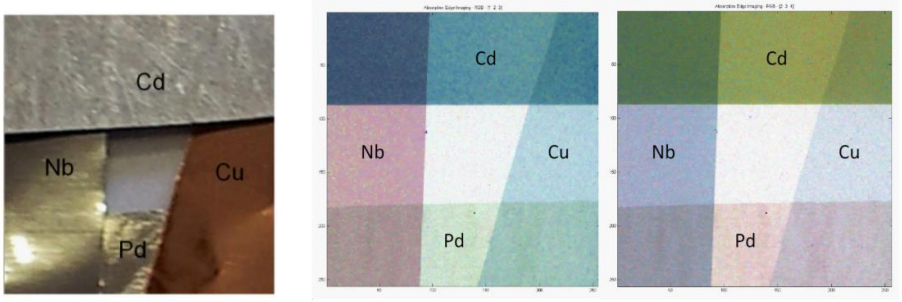
Fig. 2. (2a): X-ray with different spectral components identified, each component when assigned a color gives information about a different material namely muscles(red), bone(white), cartilage (brownish white) etc. as in the above image; (2b): 3D color x-ray of a small animal i.e. Mouse with contrast agents introduced in its body.

2. Related Work

2.1. X-ray Spectroscopic imaging

X-ray spectroscopy is a system that identifies and quantifies photons, or particles of light, that have wavelengths in the X-beam segment of the electromagnetic range (Fig. 3). It's utilized to assist researchers with understanding the synthetic and elemental properties of an item

Spectral imaging will be imaging that uses various bands over the electromagnetic spectrum. While a standard camera catches light crosswise over three wavelength bands in the visible spectrum, red, green, and blue (RGB), spectral imaging includes a wide assortment of strategies that go past RGB. Spectral imaging may utilize the infrared, the visible spectrum, the bright, x-beams, or a blend of the abovementioned. It might remember the procurement of image information for visible and non-visible bands at the same time, brightening from outside the visible range, or the utilization of optical channels to catch a particular spectral range. It is likewise conceivable to catch several wavelength bands for every pixel in an image.



(3a) (3b)

Fig. 3. (3a) and (3b): X-ray Spectral imaging of various metals and specific material detection after x-ray imaging [2]

For example, in the about figure 2, the x-ray imaging done using the special pixel detectors works as follows:

- The x-ray is made to fall on the materials to be imaged.
- Let there be an x-ray of a known frequency and energy, when these x-rays fall on the materials the frequency of these x-ray are change because of the materialistic energy attenuation in the materials of the x-rays.

- Because of this when these x-rays are re-measured with the single photon spectral detectors which have the ability to analyze the single x-ray photons.
- X-ray rays are re-measured with the single photon pixel detector and the change in their individual photon energy is taken note of.
- As the x-ray's energies are measured unlike the traditional x-ray which can count only the number of photons falling, these detectors can measure individual photon energies too.
- Because of which each material the x-ray pass through can be identified through image processing and back propagation techniques.
- The image is reconstructed as follows in figure 2b, otherwise using traditional methods the image would have been just been in a grey scale only [2].

2.2. Technology used

The major technology used is the Medipix3 and the Timepix3 chip; these chips together give the proper color illustrations.

Medipix3 is a CMOS pixel detector readout chip designed to be connected to a segmented semiconductor sensor. Like its predecessor, Medipix2, it acts as a camera taking images based on the number of particles which hit the pixels when the electronic shutter is open. However, Medipix3 [4] goes much further than Medipix2 permitting color imaging and dead time free operation (Fig. 4). A novel charge summing and allocation scheme is implemented at the pixel level permitting proper binning of the energy of incoming photons overcoming the effects of fluorescence and charge diffusion. As there are 2 counters in each 55μm pixel the chip can be programmed such that one counter is being read out while the other is counting. It is also possible to connect the chip to a sensor matrix with a pitch of 110μm. In this way, up to 8 counters are available per pixel [7].

Fig. 4. Medipix3 Chip ©CERN

These scanners go beyond the traditional black and white X-ray to produce color images where different materials can be separated. In the images separate metal, soft tissue, fat and bone are all being identified. This has a small bore scanner and is currently developing a human-scale machine.

Here they use math algorithms to take the energy information form the color of the x-ray to find the materials present. They create separate image layers for each material. The

Medipix3 chip can choose any color, but often they choose to use colors that look similar to what you would expect to see. Then take all the materials and combine them to produce a single color image (Fig. 5).

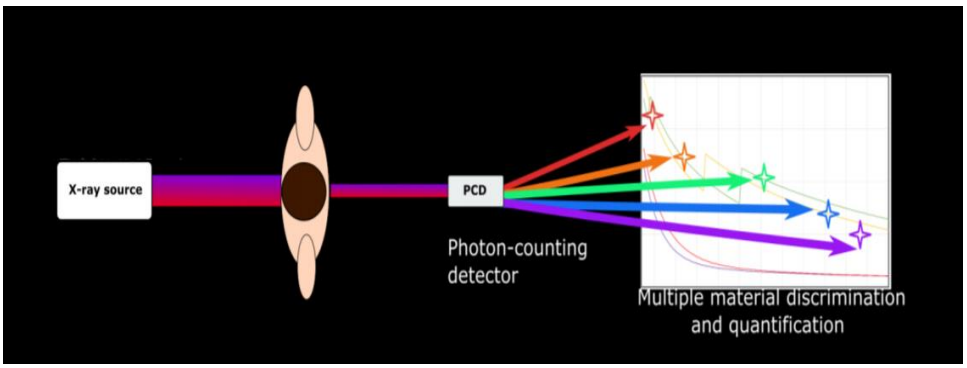


Fig. 5. Schematic illustration of work flow using x-ray spectroscopy and MEDIPIX3RX chip based photon-counting detectors [5].

3. Comparison

Here, a comparison is made between normal single energy x-ray and spectroscopic x-ray (Fig.6 and Fig. 7):

Normal X-ray 3D CT scan	Spectroscopic X-ray CT scan
○ Number of x-ray photons hitting the indicator is only measured. ○ Only one x-ray frequency measured. ○ Different materials can't be identified. ○ Gives only density related info. ○ Uses normal x-ray detectors ccd based etc. ○ 2D pattern of the scattered X-rays. ○ Cannot differentiate between the different layers of the body. ○ Traditional X-ray has poor soft tissue contrast and is limited by its resolution ○ Two different materials with same densities can't be identified.	○ Frequency/intensity of each x-ray photon is also measured. ○ Different Spectral Attenuations measured ○ Different materials are identified. ○ Gives density as well as material info. ○ Uses modified Pixel detectors similar to LHR. ○ 3D pattern of the scattered X-rays ○ Can differentiate between the different layers of the body. ○ Color x-ray also offers opportunities to use next generation contrast agents such as gold nanoparticles to detect cancer. ○ Two different materials with same densities can be identified.

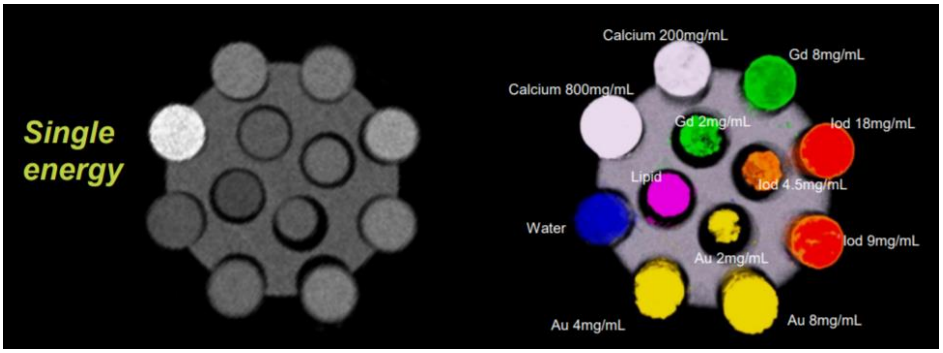


Fig. 6. Normal single energy x-ray vs spectroscopic x-ray of a phantom with 6mm tubes filled with different elements of various densities, using spectroscopic x-ray we can find the partial densities as well as distinguishing different materials.

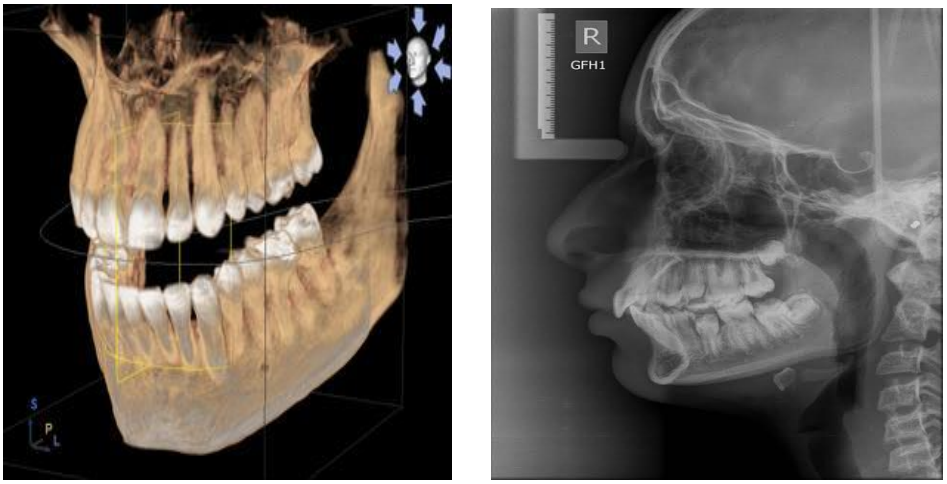


Fig. 7. Normal X-ray Analysis of the Jaw vs. the 3D Spectral Analysis of the Jaw

4. Pros and Cons of Spectral Analysis

4.1. Advantages

- Spectral imaging creates images using hundreds of thousands of narrow band they have or uses 3 to 10 wider bands.
- Having a higher level of spectral detail in spectral images gives better capability to see the unseen.
- For example spectral remote sensing distinguished between 3 layers of fat because of its high spectral resolution.
- It also adds a level of complexity and another advantage to spectral imaging is that because an entire spectrum the better image is formed.
- Spectral imaging can also take advantage of the spectra in a neighborhood, allowing more elaborate spectral structural models for a more accurate analysis and classification of the image [8].

4.2. Disadvantage

- The primary disadvantage is cost and complexity.
- Fast computers, sensitive detectors, and large data storage capacities are needed for analyzing spectral data.
- Significant data storage is necessary since spectral cubes are large multidimensional datasets, probably passing hundreds of megabytes.
- All of these factors greatly increase the cost of getting and processing spectral data.
- One of the obstacle researchers have had to face is finding ways to program large spectral images to sort through data on their own and transmit only the important images [8].

5. Application of Spectral Imaging in Other Field

- Spectral imaging technology can be used **in remote sensing** and it helps in very important to distinguish earth surface features.[10]
- Using Spectral imaging one can conclude that whether those seed are viable or not viable that is it can tell **the viability of seed.**
- Can be used in the application of **image texture for sorting tea categories** using multispectral imaging technique.[9]
- This technology has become popular in the **biological and medical application** in order to acquire the large data used in laboratory.
- Spectral imaging has also become very popular in **environmental monitoring** to understand the surface CO_2 emissions.[10]
- This can also be widely used **in food sector** for detection of bruise in apple or freshness of the fish etc.
- Spectral imaging technology is used in **exploration of oil and gas** due to this imaging sensor it is possible to detect on shore oil seeps.
- This imaging technique can also be used in the **advance solar physics examination** and exploring solar theories
- This can be also used **to measure multiple contrast agents simultaneously** this helps in tracking multiple biomarker simultaneously provides a way to monitor processes non-invasively.
- Spectral analysis provides **a very new way to image cancer** this technique offers a method to track nanoparticles, allowing preclinical to confirm that their treatment has reached their target cells (Fig.8).

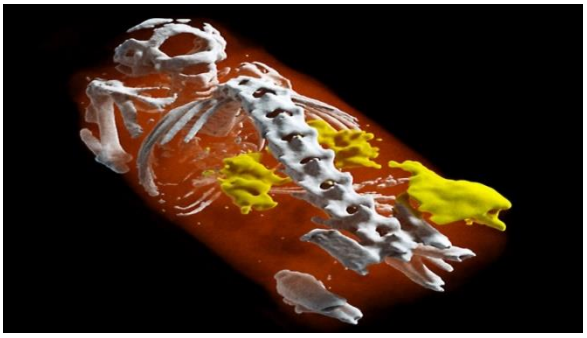


Fig. 8. Gold nanoparticles used as tracer to detect cancer

- This enables both **structural and material** information to be measured simultaneously (Fig. 9)**.**

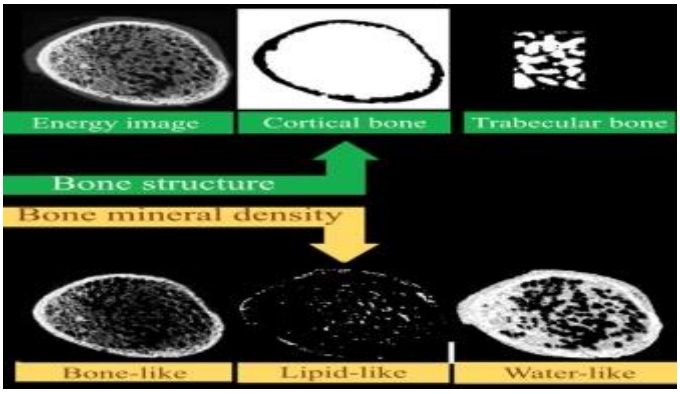


Fig. 9: Bone mineral density

- This technique provides **better soft tissue contrast** that is available with traditional x-ray system and it can be used to better characterize muscles, bone and fat (Fig. 10).

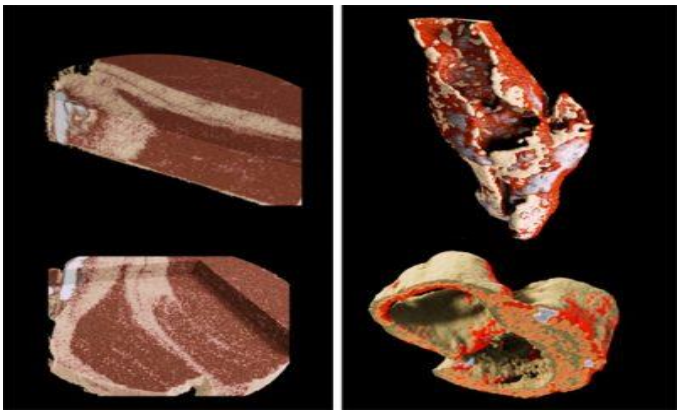


Fig. 10. Soft tissue in lamb steak and excised atherosclerotic

6. Conclusion

The medical field is constantly being influenced by technologies from other fields of study. Newer and more advanced technologies for diagnostics and therapeutics are being developed. And with the rise of better manufacturing technologies and material advancements, it is possible for technologies from other fields to be used in the medical field for various present problems and also to be able to give better diagnosis on various known and unknown diseases.

This 3D color x-ray spectroscopic imaging technology which is successfully influenced from the technologies used in the ATLAS experiment is a new era of modern technology in radiography can be by far the most efficient means of x-ray diagnosis to find more accurate and to be more amount of information from given x-ray imaging, to facilitate better diagnosis and to suggest better therapeutics. The MEDIPIX3RX chip which allows

detection of not only the no of x-ray photons but also their individual energies which has been successfully used in this imaging technology is by far the most advanced chip technology to exist in today's date. This 3D x-ray spectroscopy is not yet fully available for clinical trails for the use of general public but will be very soon available.

Thus, the 3d spectroscopic technology made with collaborations of CERN and other officials has been reviewed.

References

1. ATLAS Collaboration 2008 "T*he ATLAS Experiment at the CERN Large Hadron"* Collider *JINST* 3 003
2. R. Ballabriga,J. Alozy,G. Blaj,M. Campbell,M. Fiederle,E. Frojdh E.H.M. Heijne, X. Llopart, M. Pichotka, S. Procz, L. Tlustos and W. Wonga ; "*The Medipix3RX: a high resolution, zero dead-time pixel detector readout chip allowing spectroscopic imaging";* CERN, 1211 Geneva, 23, Switzerland IEAP-CTU, Horska 3a/22, 128 00 Praha 2, Czech Republic ´ NIKHEF, Science Park 105, 1098 XG Amsterdam, The Netherlands Freiburger Materialforschungszentrum FMF, Albert-Ludwigs-Universitat Stefan-Meier-Straße 21, D-79104 Freiburg, Germany.
3. Ostadhossein, F., Tripathi, I., Benig, L., LoBato, D., Moghiseh, M., Lowe, C., Aamir, R., Butler, A., Panta, R., Anjomrouz, M., … Pan, D. (2019). "*Multi-"color" delineation of bone microdamages using ligand-directed sub-5 nm hafnia nanodots and photon counting CT imaging. Advanced Functional Materials, 1904936. Advance online publication"*. doi: 10.1002/adfm.201904936
4. M F Walsha, S J Nikb, S Proczc, M Pichotkac, S T Belld, C J Batemana, R M N Doesburgb, N De Ruitera, A I Chernoglazove, R K Pantaa, A P H Butlera, and P H Butlerb, *"Spectral CT data acquisition with Medipix3.1"* Published 11 October 2013 • © CERN 2013 Journal of Instrumentation, Volume 8, October
5. *"Spectral x-ray imaging in medicine Anthony Butler",* On behalf of the MARS Collaboration https://indico.cern.ch/event/782801/contributions/3470098/attachments/1909624/3160690/2019.09_Medipix_20_years_public.pdf
6. *"MARS Spectral Imaging: From High Energy Physics, To a Biomedical Business"* https://indico.cern.ch/event/616390/attachments/1454028/2260470/2017.05.19.MARS_CERN.pdf
7. MARS small animal "*spectral CT harnesses Medipix3RX, the world's most advanced energy resolving photon-counting detector".* The scanner is fully self-contained & fully automated for biomedical research, https://www.marsbioimaging.com/mars/wp-content/uploads/2018/07/MARS_Electronic.pdf
8. Richard Nguyen Updated 10 April2017 "*What are the advantages and disadvantages of using hyperspec"*, https://prezi.com/gaygsbqa789k/what-are-the-advantages-and-disadvantages-of-using-hyperspec/
9. D Wu, H Yang, X Chen, Y He, X Li *"Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine*" Journal of food engineering, 2008 – Elsevi
10. N Gat, S Subramanian, J Barhen "*Spectral imaging applications: remote sensing, environmental monitoring, medicine, military operations, factory automation, and manufacturing"* 1997 - spiedigitallibrary.org

Wearable TENG Sensor Embedded in Kinesio Tape as a Self-Powered Active Biomechanical Joint Motion Sensor

Theertan Ananda, C. M. Dikshitha, Charanya Sukumaran, Chandrasekhar Arunkumar

Nanosensors and Nanoenergy Lab., Department of Sensors and Biomedical Engineering, School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

anandtheertan@gmail.com, dikshithac.2017@vitstudent.ac.in
s.charanya2018@vitstudent.ac.in, arunkumar.c@vit.ac.in

Abstract. Triboelectric Nanogenerators (TENG) in field of wearable human motion sensors has shown significant potential in biomedical monitoring and athlete posture sensing which requires the proper power management circuit. Hence the idea of scavenging the biomechanical energy would minimise this issue thereby giving a new insight towards self-powered devices.We propose a flexible, wearable Kinesio tape based triboelectric nanogeneartor (KT-TENG) which converts the biomechanical energy of knee joints to electrical energy that is further programmed to monitor the athlete's motion. This works on contact and separation mode where Ecoflex and copper are the negative and positive tribolelectric layers of KT-TENG which is fabricated using a cost-effective approach. The device generates a maximum voltage of ~5V and ~1μA of current respectively.The performance analysis proves that KT-TENG has the potential to contribute towards the energy conversion process providing original intended benefits to athletes.

Keywords. Triboelectric Nanogenerator, KT Tape, Self-powered Sensors, Soft-wearable sensor platform, Biomechanical energy

1. Introduction

Wearable electronics have been consistently developed since the beginning of 21st century and are ever since preferred for their emerging applications in military, commercial, medical and industrial sectors of applications. Accurate Human motion sensing and movement tracking has very recently evolved as one of the most promising sub-categories which consumes huge amount of power and requires constant external supply. Many researchers across the globe are constantly working on this domain to minimise this effect that gave rise to Self-powered technology which is the best alternative for battery driven wearable devices.[1,2] Extended from the phenomenon of Displacement current, modern nanogenerators (NG) are based on the observable triboelectric and piezoelectric effects optimized to efficiently scavenge mechanical energy from the available energy sources. This can be used to develop battery-free active self-powered sensors with high stability, cost-effectiveness and efficiency.

Triboelectric Nanogenerators (TENGs) operate on the principle of electrostatic induction and triboelectrification where high efficiency, substantial output, light weight, less cost, eco-friendliness and simple fabrication are the fascinating features. [3]. Ever since emergence of TENG devices in 2012, efforts have been dedicated to field this technology thereby demonstrating their wide field of applications, ranging from self-powered active sensors to self-charging systems [4-6]. Specifically, in the advancement of wearable human

motion tracking sensor, TENGs have been deployed as self-charging appliances for sensing applications and real-time monitoring [7,8]. Researchers namely Wang et.al [9] have developed KT-TENG sensors for Self-Powered Human Motion Sensing. Liu et.al [10] have fabricated PDMS strain sensors interfaced with Kinesio Tapes. Development of such composite sensors provide useful biomechanical data as signal waveform while simultaneously providing all benefits of a regular Kinesio-Tape. We present a wearable, embedded and flexible Kinesio-Tape-based TENG (KT-TENG) based on contact-and-separation working mode as a self-driven smart human motion tracking sensor for athletes. A real-time motion sensing resulted in maximum voltage of ~5V and ~1μA of current respectively.

2. Materials and Methodology

Kinesio and Copper Tape were cut into required dimensions which is further applied on athlete's skin surface. The polymer (Ecoflex) was prepared by mixing monomer and hardner in equal ratios which is poured onto sandpaper of very fine grit sizes which ensures that the resultant film has irregularities contributing to higher charge generation.The sandpaper layout is degassed at a temperature of 70 ºC for 1 hour to remove the air bubbles. Ecoflex is then peeled off from the mould to obtain a thin film. Two conducting copper wires were connected as the connective mounted leads for measuring electrical activity as output. Finally, the entire sensor device is embedded and encapsulated using kinesio tape (Fig.1).

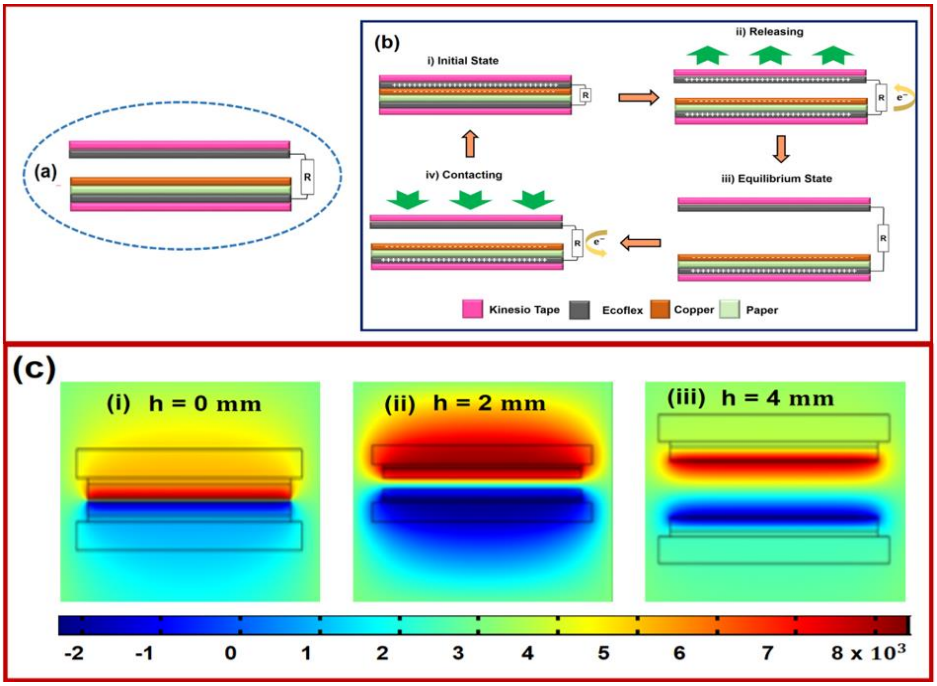


Fig. 1. (a) Proposed Device Structure of KT-TENG; (b) Working principle of the device; (c) Comsol simulations of KT-TENG

3. Results and Discussion

The performance of KT-TENG has been thoroughly analysed which works on fundamental contact and separation mode as illustrated in Figure.1 [11]. Initially the triboelectric layers are in contact signifying the zero potential in fig.1(b)(i) Whenever an external force is applied via human knee flexion or extension, a potential difference is developed which provides a path for the charges to flow between the layers thereby contributing to the positive cycle of AC signal as shown in fig.1(b) (ii) Equilibrium state is achieved when both the substrates attain their original state as highlighted in fig.1(b)(iii).When external force is applied on the layers at this condition, the potential difference decreases as the electrons the flow in the opposite direction, completing the AC signal which is demonstrated in fig.1 (b); (iv). When the layers are in full contact again, the charges are neutralised and no output is generated as shown in fig.1 (b)(i). This theoretical approach is validated using COMSOL software and the simulation results are presented in fig.1(c)(i-iii).

Considering the empirical implementation of the wearable KT-TENG sensor, the device works effectively under flexion and extension movements. In addition to this, a real-time knee joint motion sensing was performed which resulted in ~5V and ~1μA of voltage and current respectively thereby possessing good stability, simple fabrication, compact structure, superior flexibility which gives a deeper insight towards recyclable or eco-friendly approach to fabricate self-powered wearable devices [12] (Fig. 2).

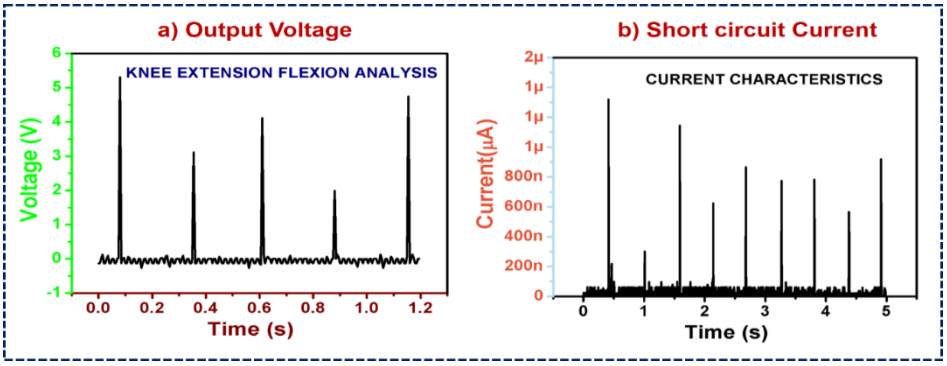


Fig. 2. Electrical analysis of KT-TENG (a) Output voltage. (b) Short circuit current

4. Conclusion

We have successfully fabricated flexible, cost-effective self-powered wearable KT-TENG which scavenges the biomechanical motion of knee joints and enhances the energy conversion process. This device comprises of two triboelectric layers (Ecoflex and Cu Tape) working on contact and separation mode. Cost effective approach was adopted to fabricate the device. Stretching the TENG device proportional to certain displacements of knee joint, an accurate output electric signal is generated which is recorded using a digital oscilloscope thereby giving a new insight towards self-powered sports, entertainment, and healthcare appliactions. Kinesio Tapes enables the improvement in athletes performnace which includes the delivery of competitive feedback during their intensive training.

Acknowledgement

The authors thank VIT for providing 'VIT SEED GRANT' for carrying out this research work.

References

1. Guo, H.; Yeh, M.H.; Lai, Y.C.; Zi, Y.; Wu, C.; Wen, Z.; Hu, C.; Wang, Z.L. All-in-one shape-adaptive self-charging power package for wearable electronics. ACS Nano 2016, 10, 10580–10588. [CrossRef] [PubMed]
2. Lin, Z.; Chen, J.; Li, X.; Zhou, Z.; Meng, K.; Wei, W.; Yang, J.; Wang, Z.L. Triboelectric nanogenerator enabled body sensor network for self-powered human heart-rate monitoring. ACS Nano 2017, 11, 8830–8837. [CrossRef] [PubMed]
3. Wang, Z.L. Triboelectric nanogenerators as new energy technology for self-powered systems and as active mechanical and chemical sensors. ACS Nano 2013, 7, 9533. [CrossRef] [PubMed]
4. Chen, J.; Yang, J.; Li, Z.; Fan, X.; Zi, Y.; Jing, Q.; Guo, H.; Wen, Z.; Pradel, K.C.; Niu, S. Networks of triboelectric nanogenerators for harvesting water wave energy: A potential approach toward blue energy. ACS Nano 2015, 9, 3324–3331. [CrossRef] [PubMed]
5. Fan, F.R.; Lin, L.; Zhu, G.; Wu, W.; Zhang, R.; Wang, Z.L. Transparent triboelectric nanogenerators and self-powered pressure sensors based on micropatterned plastic films. Nano Lett. 2012, 12, 3109. [CrossRef] [PubMed]
6. Lin, L.; Xie, Y.; Wang, S.; Wu, W.; Niu, S.; Wen, X.; Wang, Z.L. Triboelectric active sensor array for self-powered static and dynamic pressure detection and tactile imaging. ACS Nano 2013, 7, 8266–8274. [CrossRef] [PubMed]
7. Cai, F.; Yi, C.R.; Liu, S.C.; Wang, Y.; Liu, L.C.; Liu, X.Q.; Xu, X.M.; Wang, L. Ultrasensitive, passive and wearable sensors for monitoring human muscle motion and physiological signals. Biosens. Bioelectron. 2016, 77, 907–913. [CrossRef] [PubMed]
8. Chen, X.X.; Miao, L.M.; Guo, H.; Chen, H.T.; Song, Y.; Su, Z.M.; Zhang, H.X. Waterproof and stretchable triboelectric nanogenerator. [CrossRef] [PubMed]
9. Wang, S.; He, M.; Weng, B.; Gan, L.; Zhao, Y.; Li, N.; Xie, Y. Stretchable and Wearable Triboelectric Nanogenerator Based on Kinesio Tape for Self-Powered Human Motion Sensing. Nanomaterials 2018, 8, 657. [CrossRef]
10. Z. Liu, D. Qi, G. Hu, H. Wang, Y. Jiang, G. Chen, Y. Luo, X. J. Loh, B. Liedberg, X. Chen, Surface Strain Redistribution on Structured Microfibers to Enhance Sensitivity of Fiber-Shaped Stretchable Strain Sensors. Adv. Mater. 2018, 30, 1704229. [CrossRef]
11. Shi, Q.; Wang, H.; Wang, T.; Lee, C. Self-powered liquid triboelectric microfluidic sensor for pressure sensing and finger motion monitoring applications. Nano Energy 2016, 30, 450–459. [CrossRef]
12. Maselli, M.; Mussi, E.; Cecchi, F.; Manti, M.; Tropea, P.; Laschi, C. A wearable sensing device for monitoring single planes neck movements: Assessment of its performance. IEEE Sens. J. 2018, 18, 6327–6336. [CrossRef]

Training to Work with ERP Systems with Virtual Reality

Emil Denchev

University of National and World Economy, Sofia, Bulgaria
emild@unwe.bg

Abstract. The aim of the present article is to improve the level of information provision of the training to work with ERP system with virtual reality (VR). The tasks of the research are the specific requirements for training staff to work with ERP systems and how it can be realized with virtual reality. The methodology used is methods of analysis and synthesis.

Keywords. Training, ERP system, Virtual reality.

1. Introduction

„The performance evaluation of IT projects and their influence over the increase of efficiency of the company itself is evident. One of the main methods for evaluation of already produced IT solutions is their Total Cost of Ownership (TCO)“ [1].

„The cost of the users education should be reported in the performance evaluation of the IT project“ [1].

One of the main aspects of introducing an ERP system seems to be educational costs. The usage of the ERP system cannot be successful if the system works exemplary, yet the staff are not aware of how to use it to its maximum. The education related to the ERP system is not only concerning the users of the ERP system but also the people who are in charge of its implementation. Education regarding work with a specific informational system is often viewed as part of the implementation project, but when the employees change and new people are employed they also need to get acquainted with that informational system. Apart from that, systems are changed every two to five years and each new functionality means an additional education process for the employees. The study of working with an ERP system is a process that is prolonged throughout the whole life cycle of the ERP System.

2. Planning

The execution of the staff education curriculum is tightly connected to a lot of planning being done beforehand on the following topics: Period of Execution of the system, Volume, Type of Learning Materials, Compensation of the time of learning, Motivation, Expenses, Reach, Structure of the Education, Digital Technologies used for “Speed” of the education, Induction training, Importance.

Period of Execution

The knowledge and skills acquired from an education which is held long before the start of the system would be forgotten. An education started too late, will not end in time and the “Stabilization” period might be delayed. For a training to take place before the start of the system, the firm is needed to confirm how much time there is before the start and how much time will be needed for the end users.

Volume

The volume of the necessary education depends on the specific module or modules, the purpose of which is the training of the employees. In some cases, for the end user to use the ERP system in a confident and comfortable manner, it may take up to six months.

Compensation for training time

There are different solutions for freeing sufficient time for employees training. This time can be part of their daily work hours or be added as additional time for carrying out the education. Some firms recruit temporary mentors for the duration of the training of their specialists.

Importance

The time spent for education of the staff, shows how important it is for the implementation period. Staff education, planned during work hours speaks for its importance, while conducting it outside of business hours, implies lesser importance than regular business as usual tasks and responsibilities take priority. A commonly accepted way of conducting education is teaching a member of the client's staff as a "super user", who in terms can educate other users about the things he has learnt. This kind of education by a "super user" has its own perks and advantages. We will cover the most important three advantages. First, the super user can ensure the support of the other users, as he is an employee of the same firm as the rest of the staff. Second, creating such a "super user" shows the rest of the personnel, which are treated as "basic" users, that this education regarding the new software is an important task. Third, the super user can showcase a certain level of the education system that will be sufficient to be considered, that a basic user understands how to work with the software.

Motivation

The biggest attention often falls on the task which cannot be postponed – often at the cost of the time for education, which at this point in time is of a lower priority. This means that the business as usual work is commonly treated with more attention and detail. If the company in question does not put enough emphasis on the education process for implementing a new ERP system, the users might treat it as of a much lesser priority thus being pushed back as plan B. As a result, companies often introduce either punishment for neglecting these education materials/sessions or bonuses to emphasize on their importance. An example of this could be that this education might be mandatory to specific key users.

Expenses

The expenses for the education in question might vary depending on the project. Usually the budget could make up of up to 10% or more of the total project budget. A lot of companies and firms go into unnecessary expenses as much as educations are concerned.

Range

The education of the staff to work with an ERP system should include technologies, business processes and specialized insight, to ensure sufficient information on the system. On the other hand, the education process might become too generic, not enough specialized in the ERP system that will be in use.

Type of the educational materials

There are two commonly used types of materials for education: Generic ones and Scenarios. The generic materials usually are not affected by the context, but contain lists of possibilities and some of the terms with explanations, similar to a HELP documentation. Scenario guidelines are created for the users, commonly used to take a user through the whole process, step by step, until the end of the whole business process. These guidelines are available to the users in the local network. This approach has a few advantages that we will cover. First, in this kind of guidelines a user might often find a page with frequently asked questions (FAQ). Second, they are also used as documentation for system usage with tips and tricks. Third, these materials are solely based on the context in which the ERP system is used. When using predefined scenarios there is one major flaw – when encountering a situation that is not listed in the scenario guidelines, then there is no information on what to do.

Structure of the training program

Usually the major part of the training program is based on what the company in question needs (are the systems are created using a "module" principle, could it be integrated with other IT system, is the range of the company regional, national or international placement) and what is needed to be done.

Digital technologies for education

Almost all of the education of an ERP system is conducted in auditoriums in a digital form of education. Although there are other forms of education like education through Internet, Self-learning and new digital technologies, like virtual reality.

Virtual reality (VR) is a computer generated environment, which allows for the users to be showcased everything that they have access to as education, with the help of special VR Glasses and 3D sound (or a special VR Headset). The VR video marketing is a much better marketing tool than the regular video, because it generates up to 6 times more viewership, up to 7 times more shares in social medias, has a much stronger influence in the viewers and is an innovative technology, being the most promising way of communication in the future. [2]

The benefits of the usage of virtual reality for education purposes are: quite an effective way of informing the trainees as well as influencing them; times better of a visual experience; encourages the continuation of the education process; a complete overview of the informational system by all of the employees from everywhere; possibility for online viewing of the video in the office, storage unit, or the shop; usage of computer generated graphic for adding non existing objects and attaining captivating visual effects. [3]

The VR goggles are a much more attractive way of the company to showcase all that it offers to the clients. This offers a virtual walkthrough across all the modules of the system and all of the stages of the business processes, regarding finances, accounting, production of goods, storage, delivery, sales, services/support center, material planning, analyses, administration, while the clients are hosted in a specific VR space in the entity. Via computer graphic, we can ensure the creation of many different scenarios, future innovations, new services (functionalities) in the system or other systems, integrated with it even before their implementation. The staff is encouraged for potential future participation in live/real education on this topic. [4]

„Speed" of the education

The project of the implementation of the ERP system often falls behind schedule, thus the firm can try to speed up the learning process. The capability of the firm to actually speed

up the learning process depend on the needs of the firm, on the staff and on the education had beforehand. Sometimes, one part of the personnel gets educated ahead of the schedule, or has more education time compared to the other part of the staff. The education is conducted slower or faster depending on the similarities of the new system to the old one which the users are adept at using. The attempt of speeding up the education however, has potential dangers and risks in itself as it might hinder the whole project or cause it to fail.

Training of the induction employees

The users of the ERP systems are not the only ones who need to get educated to work with it. Apart from them, a different but corresponding education needs to be held for the employees responsible for the implementation of the ERP system.

The specifications of the education for work with the distinct ERP system version, could affect the speed and efficiency of the implementation process. The inadequate education might negatively affect the implementation process. An example of this would be, the users lack of knowledge of the system capabilities, which in turn might lead to unnecessary changes, delaying the implementation process and increase in expenditures. The lack of adequate preparation might also impact the volume and quality of changes and fixes needed to be implemented, conducted in the ERP system during the implementation process. [5]

3. Conclusion

The education is a key element for the successful implementation and afterwards work for each specific informational system, including ERP systems as well. For its conducting multiple parameters are planned and modern digital technologies can be used to ease the process. The Virtual Reality creates a more interactive learning environment and can present, in an easy to understand way, all the processes that are carried out throughout the whole business cycle and their respective core responsibilities in regards to usage of the modules of the ERP system and other informational Systems, integrated in it (including incoming, intermediate and outgoing documents, internet portals and internet networks, etc.).

References

1. CIO Media 2014. Filosofiia na efektivnostta, Management https://cio.bg/management/2014/05/27/3440814_filosofiia_na_efektivnostta/
2. Tsaneva M., 2019. A Practical Approach for Integrating Heterogeneous Systems, Business management, Year XXIX, Book 2, pp. 5 - 15, https://dlib.uni-svishtov.bg/bitstream /handle/10610/4097/55e8a0376a87fe55ed48fcd15ea8aaa2.pdf?sequence=1&isAllowed=y)
3. Ivanov V., M. Tzaneva, A. Murdjeva, V. Kisimov, 2010. Securing the Core University Business Processes, Open Research Problems in Network Security, IFIP WG 11.4 International Workshop, iNetSec 2010, Sofia, Bulgaria, March 5-6, 2010, Revised Selected Papers, pp. 104-116, https://link.springer.com/content/pdf/10.1007/978-3-642-19228-9_9.pdf
4. Tzaneva M., S. Kouzmanov, 2016. Applicability of Business Information Systems Integration Approaches from Technical and Management Perspective, Proceeding of the 6th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE 2016), 2-3 December 2016, Sofia, Bulgaria, pp. 538-540
5. Tsaneva M., 2019. Challenges of GDPR compliance in consumer financing companies. Conferences of the department Informatics, Publishing house Science and Economics Varna, issue 1, pp. 103-115, http://informatics.ue-varna.bg/conference19/Conf.proceedings_Informatics-50.years%20103-115.pdf

Data Mining Techniques for Detection of Credit Card Fraud

Angelin Lalev

D.A.Tsenov Academy of Economics, Svishtov, Bulgaria
a.lalev@uni-svishtov.bg

Abstract. In this paper, we present preliminary results from ongoing study on applicability of deep neural networks and other data mining techniques on detection of credit card fraud. We examine the effects of varying both depth and learning parameters of the networks and compare their efficiency. Our results indicate that neural networks of moderate depth are superior to other approaches. We also observed that generative-adversarial approach in designing networks for detection of credit card fraud seems to induce instabilities in the learning process of the network.

Keywords: Data mining, Neural networks, DNN

1. Dataset

We use available dataset with 280 000 transactions, occurring in the course 48 hours (The dataset is available through Kaggle at https://www.kaggle.com/mlg-ulb/creditcardfraud). The data comprises mostly of transactions, made by European cardholders. Fraudulent transactions in the entire dataset amount to 492 or 0.17% of the entire dataset.

Before being published, the dataset has been anonymized and most of the columns of the original data have undergone PCA dimensionality reduction. The resulting data consists of 31 columns. 28 of these columns contain normalized values which fall mostly between -1 and 1 and the other three contain time of the transaction (in seconds after the beginning of the period), the amount of the transaction, and a number 0 or 1 which indicates respectively legit or fraudulent transaction.

As part of the data preparation, we introduced two further transformations of the dataset. Based on the hypothesis that time of day may be very slightly related to the possibility of the fraud occurring, the time column has been converted in values between 0 and 1 which represent the time since the start of each 24-hour period.

The other transformation concerns avoidance of roundoff errors. The amount column must be mapped to values between 0 and 1 to facilitate neural network training, but this is problematic due the presence of outliers as big as $25000. Most of the transactions in the dataset do not exceed $100, and direct normalization would result most values being too close to zero. This in turn would lead to roundoffs when computing gradients and would prevent proper neural network training.

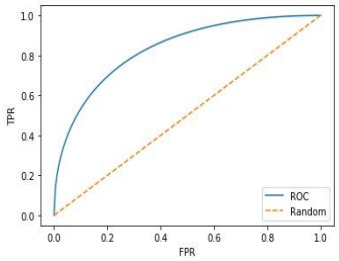
To deal with this problem, the values of this column have been capped to $5000 before normalization. The cap is somewhat arbitrary, but it can be rationalized by the hypothesis that probability $5000 transaction being fraudulent should be basically the same as the probability $25000 being fraudulent. This hypothesis is itself based on the observation that in general fraudsters prefer stealing smaller sums at once to avoid detection. Only 55 transactions in the entire dataset are over $5000. All of them are legit.

To measure the performance of each DNN architecture, the dataset has been split into training and testing sets, and ¼ of the records in the original dataset have been assigned to the testing set, based on random selection.

2. Metrics

The task of separating fraudulent and legit transactions is a binary classification. Binary classification models that produce discrete output are measured in terms of the well-known confusion matrix (see Fig.1a).

		Actual	
		True	**False**
Predicted	**True**	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)



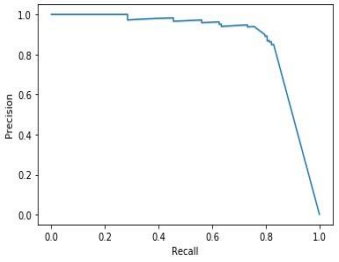


Fig. 1. (a) Confusion matrix /left/; (b) ROC curve /center/ and (c) PRC curve /right/

In time, different problem domains have produced different (often redundant) metrics to measure quality of binary classifiers, but all of them can be related to the confusion matrix components. **True positive rate** (TPR), known also as Sensitivity or **Recall** measures how much of the members of the positive class are correctly classified by the model. **False positive rate** (FPR) measures how many members of the negative class are incorrectly labeled as positive by the classifier and **Precision** or Positive Predictive Value (PPV) measures the proportion of true positives to the sum of true and false positives. That is:

$$TPR = \frac{TP}{TP + FN}; FPR = \frac{FP}{FP + TN}; PPV = \frac{TP}{TP + FP}$$

In most cases though, binary classifiers result from models, which output values between 0 and 1, interpreted as score or probability that some sample belongs to the positive class. To make classifications then, a threshold is chosen and all data points that result score above the threshold are assigned to the positive class. Each different threshold results different confusion matrix, which makes the aforementioned metrics impractical when comparing classifiers themselves.

Standard approach when comparing the performance of binary classifiers is to use Receiver Operating Characteristics (ROC) curve (see fig.1b). ROC is a visual tool that plots true positive rates vs. false positive rates as the threshold goes from 1 (at witch point no element of the dataset is assigned to the positive class) to 0 (where all the elements in the dataset are assigned to the positive class). It has many advantages, one of which is that ROC curve is very intuitive. ROC curves can be transformed into quantitative measure of classifier performance by measuring the area under the ROC curve. This measure is abbreviated as "AUROC"), 1 being AUROC score of a perfect classifier.

Unfortunately, with imbalanced datasets where the members of the negative class are much more prevalent, ROC and AUROC give over-optimistic assessment of the quality of the model. This results from the fact that absolute increase of the false positives will raise FPR much slower than the decrease of false negatives would raise TPR. This results higher situated ROC curves "by default" and consequently - larger AUROC scores that differ very slightly between various different models.

Our dataset is greatly imbalanced, with positive class consisting of only 0.17% of the records. Because of this, as recommended by the providers of the dataset, we use Precision-Recall curve (PRC) which plots Recall vs Precision (see fig. 1c) as threshold goes from 0 (maximum recall) to 1 (maximum precision). This approach is used less often, but unlike the case with ROC curve, the number of true negatives does not participate in computation of PRC, which helps avoid problems with imbalance of the classes. As with ROC, PRC curve can be converted in quantitative measure when the Area under Precision-Recall Curve (AUPRC) is measured. Again, AUPRC of 1 is the score of a perfect classifier.

3. Architectures

One of the main goals of our study is to determine if additional hidden layers with different dimensions can help improve the quality of DNNs' predictions about credit card fraud. This is valid question, since unlike the field of image recognition, where DNNs were used quite successfully, financial data is much less dimensional. And unlike image recognition, where first layers of the neural network are supposed to learn minor features and the latter layers are supposed to make generalizations, there aren't any recipes for choosing DNN architecture when dealing with task as specific as detection of credit card fraud.

Nonetheless, we can make some reasonable assumptions based on the general properties of the dataset. The most important of these assumptions is that imbalance of the dataset will greatly increase the tendency of the network to overfit the data. Addition of further hidden layers, as well as increasing each layer's dimensions, will further contribute to this tendency.

This probably means that there will be need of measures to prevent overfitting. In our study we have explored two methods to reduce the overfitting - adding of dropout layers, which turn some neurons off at each iteration, and usage of oversampling to balance the dataset.

In addition to different layers and different number of neurons in each layer, we have experimented with two activation functions, suited for our purpose - Rectified Linear Unit (ReLU) and Hyperbolic Tangent (tanh). Both functions are tried-and-true representatives of their respective classes. In principle, ReLU has reputation of the better activation function for DNNs, because it does not suffer from vanishing gradients problem. ReLU has its own disadvantages though - is not differentiable at 0 and in rear occasions it can lead to situations where part of the neural network "gets stuck" and do not participate in further training.

Table 1 summarizes the architectures tested so far.

Table 1. Summary of tested DNN architectures

Short moniker	Layers				
	0	1	2	3	4
	ReLU	Dropout	ReLU	Dropout	ReLU
16-48-1	16	x	48	x	1
32-72-1	32	x	72	x	1
32-72-1	32	0.2	72	0.2	1
32-72-1	32	0.3	72	0.3	1
32-72-1	32	0.4	72	0.4	1

In addition to these more conventional architectures, we have devised additional one which emulates the approach, adopted by special form of neural networks - the so called "Generative Adversarial Networks" (GANs). The idea behind this architecture is to teach additional deep neural network how to generate synthetic data which has the distribution of the original positive class.

Then this synthetic data is used along with the real data points to train our classifier. In essence, this is oversampling method, which helps balance our training set.

4. Results

Of the tests we conducted on DNNs with no drop layers, the best results on the test set were achieved by 24-56-1 network, that is, a network with two hidden layers with respectively 24 and 56 neurons, which achieved AUROC of about 0.832. These results were closely matched by the 32-72-1 network that, over 900 epochs achieved AUROC of about 0.826.

The AUPRC of 16-48-1 network failed to converge over the same number of epochs and got its best results early in the training, which seems to indicate that the network did overfit the data.

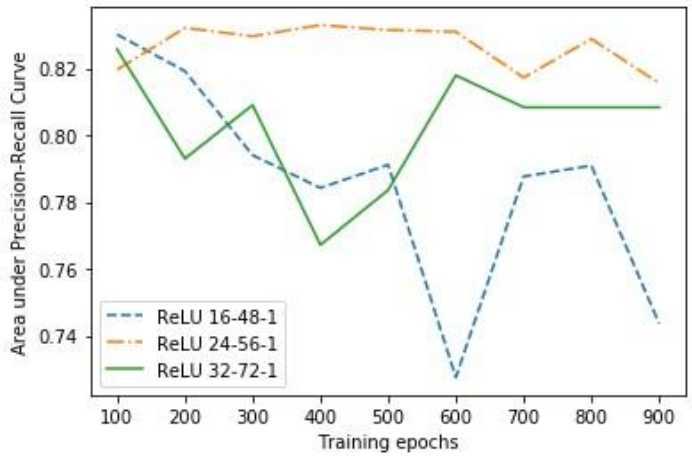


Fig. 2. Comparison of DNNs with various layer widths

Addition of the dropout layers to the 32-72-1 network produced by far the best results from DNNs that do not use oversampling. Of the three scenarios tested - adding layers with dropout of respectively 0.2, 0.3 and 0.4 between the hidden layers - the one with dropout of 0.3 produced the best result with AUROC of 0.877 within the first 200 epochs of training. In the remaining 700 epochs the results slowly deteriorated, probably because of overfitting on the training set.

Increase of dropout to 0.4 produced AUROC of over 0.874 on the test set in the first 200 epochs of training. Reducing dropout to 0.2 produced best result of 0.853 AUROC.

To get a baseline, we fitted basic logistic model on the training set, which achieved about 0.76 AUROC on the test set. In addition we experimented with conventional, "shallow" neural network 32-1 of only one hidden layer, which produced best AUROC of 0.81 over 900 iterations.

Figure 3 shows comparison of the best results, achieved by means of logistic regression, neural networks without dropout and neural networks with dropout.

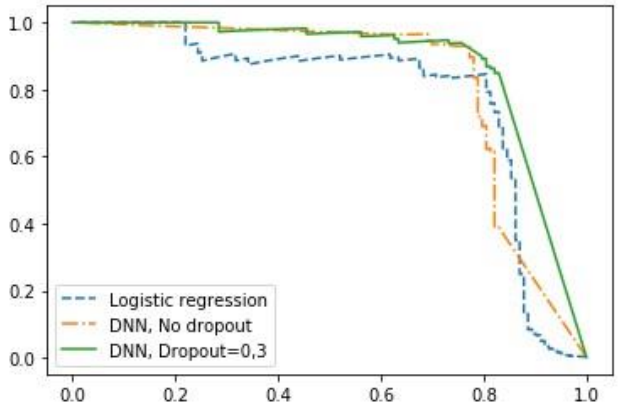


Fig. 3. Precision-recall curves of best architectures with dropout and without dropout vs. logistic regression as baseline

These results seem to indicate that neural networks are more than credible method for detection of credit card fraud. The neural networks have demonstrated marked advantage over basic logistic regression and the deep networks with dropout have performed much better than neural networks without dropout.

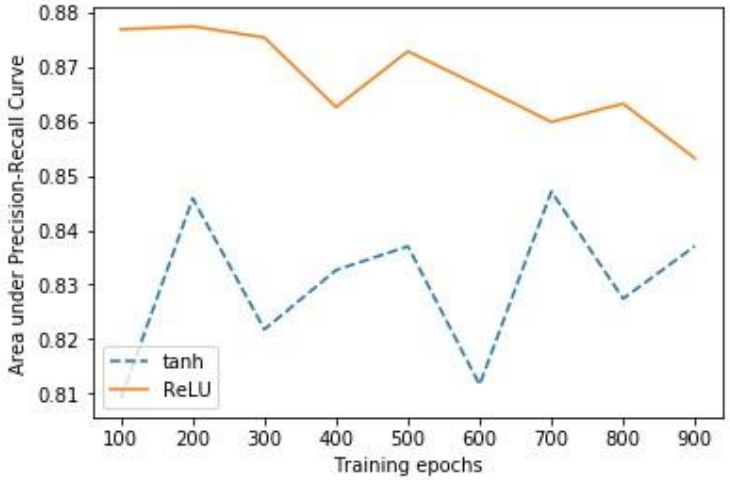


Fig. 4. Hyperbolic tangent vs Rectified Linear Unit activations

The results, described above, were achieved by using ReLU activation function. Most of the experiments were conducted with both tanh and ReLU activation functions. In practically all tested scenarios, ReLU produced better results than tanh. Figure 4 shows the difference between usage of ReLU and tanh on the best performing architecture - 32-72-1 with dropout of 0.3.

To try to further improve detection of fraudulent transactions, we tried to combine the best DNN architecture - 72-32-1 with methods for oversampling. To get a baseline we first tried naive oversampling, which consisted of balancing the training set by repeating the members of positive class. This time the process of training the network was separated into batches of 200 legit and 200 fraudulent transactions, picked at random from the training set.

The training was repeated for over 1000 epochs, consisting of 1000 batches. These numbers were picked to get approximately the same number of individual iterations as in previous scenarios.

As alternative approach we created GAN-like model with generator and discriminator network. The discriminator was tasked with detection of real fraudulent transactions from the training set versus "fake" fraudulent transactions, produced by the generator.

Training of the generator to produce data with similar statistical properties as the fraudulent transactions proved to be difficult task. In terms of Euclidean distance, which was chosen as metric of the performance of the generator, the model saw long periods of convergence to 0, followed by periods of steady divergence. To deal with the problem the state of our model was saved each 10 epochs and then a state with minimal Euclidean distance was chosen to be used at the next step.

At the next step we trained the classifier by using batches of 200 legit transactions, combined with 100 fake transactions and 100 "synthetic" fake transactions, generated by the generator part of the GAN from the previous step.

Figure 5 shows AUPRC values of both methods over 1000 epochs, 1000 batches each.

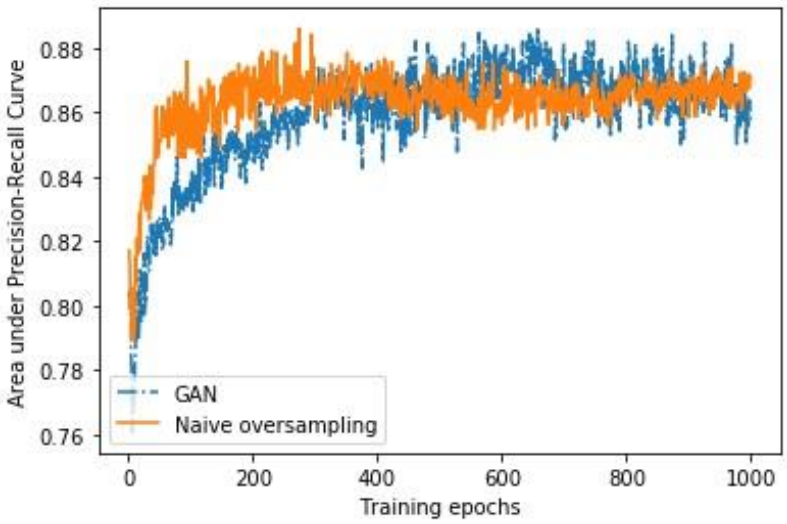


Fig. 5. Naïve oversampling vs. GAN-inspired approach

In terms of AUPRC on the test set, best achievement of the GAN-inspired approach was AUPRC of 0.885. The naive approach achieved AUPRC of 0.886. In terms of convergence, the GAN-inspired classifier turned out to be more stable though, which is observable on figure 5.

5. Conclusions

Detection of credit card fraud is non-trivial task and there is absolute certainty that perfect classification will never be achieved. The reasons for this are that many fraudulent transactions will look, in terms of the data gathered by the credit card companies, indistinguishable from the legit ones.

The achieved results on the dataset seem to indicate that classification models that produce AUPRC of beyond 0.88 under similar conditions (i.e. 1/4 of the total rows being test

set) on the same dataset will be result of precise calibration and non-trivial breakthroughs in the methods for binary classification.

In addition, it cannot be confirmed with absolute certainty that our chosen methods for oversampling lead to marked improvement of the classifier. Due lack of data, we cannot answer another very important question - how would oversampling techniques perform, given they work on bigger (in absolute terms) set of positive class members, that is - fraudulent transactions.

What it can be stated with certainty is that DNNs can be used successfully to augment standard statistical techniques and to improve detection of credit card fraud. In addition, our results seem to indicate that the problem domain does not hide any particular surprises in terms of unusual results about dropout rates, width of the layers and chosen activation functions.

References

1. Chintala, S. & LeCun, Y., 2016. *A path to unsupervised learning through adversarial networks.* (https://engineering.fb.com/ml-applications/a-path-to-unsupervised-learningthrough-adversarial-networks/)
2. Goodfellow, I. J. et al., 2014. *Generative Adversarial Nets.* s.l.:arXiv:1406.2661v1.

Application of Wireless Multimedia Sensor Networks and Finite-State Automatons for Medical Services Delivery and Personal Data Protection

Veselina Aleksandrova, Pencho Popov , Doroteya Smilenova

G. S. Rakovski National Defense College, Sofia, Bulgaria
alexandv@yahoo.com, penchopopov@yahoo.com, dsmilenova@gmail.com

Abstract. The report considers the possibility for specific decisions in collecting the primary database in the delivery of medical services. The goal is to show that Wireless Multimedia Sensor Networks (WMSN) can not only improve existing applications, such as monitoring, automation of processing and transmission of medical indicators, health parameters of different groups of users / patients, but could also allow the use of modern technologies in the delivery of health databases. In the context of an existing COVID-19 pandemic, this is especially important for such areas in Telemedicine as: Teleradiology, Telesurgery, Telepathology, Telemonitoring, Teledermatology, Teleconsultation etc. Wireless networks allow the transmission of video and audio streams, still images and data from scalar sensors. With significant hardware improvements and miniaturization, a single sensor can be equipped with modules for collecting audio and visual information. In addition to receiving multimedia information, (WMSN) can also store and process in real time, multiplex data from disparate sources. In clinical practice, various devices as Finite State Automatons (FSA) are used with the help of which the information is collected, registered, transmitted, processed and stored. The fully automated process on time during which documents, information or tasks are transferred by one participant to another can also be performed by wireless sensor networks equipped with different types of microcontroller board such as the Arduino board for the purpose of Telesurgery. Arduino can be particularly well configured with surgical cameras in Telesurgery to perform remote surgical operations with a robot. The need for the protection of the personal data of the processed subjects (GDPR) is also discussed.

Keywords. Wireless multimedia sensor networks (WMSN), Microcontroller circuit board, Arduino circuit board, Finite-State Automatons (FSA), GDPR, cybersecurity, Health and medical data transfer.

1. Introduction

For specific solutions to collect data database using in the individual sections of Telemedicine, such as Teleradiology, Telesurgery, Telepathology, Telemonitoring, Teledermatology, Teleconsultation and Teleophthalmology can be used different interesting applications related to Wireless Sensor Multimedia Networks (WSMN). In our opinion, the use of the application capabilities of large-scale networks of small devices, which can collect information about their environment, carry out simple processing and transmit it to remote stations, should be expanded. This is especially true in the implementation of electronic medical services through wired sensor networks as WSMN that measure scalar physical quantities such as temperature, skin moisture, blood pressure, pulse, blood sugar and more. Most of these applications use low transfer speeds and are tolerant of delays on non-specialized hardware, such as CMOS (Complementary Metal–Oxide–Semiconductor) cameras and

microphones. They record multimedia data from the environment using wireless multimedia sensor networks (WSMN), i.e. networks of wirelessly connected devices that allow the transmission of video and audio streams, still images, and data from scalar sensors. With significant hardware improvements and miniaturization, a single sensor can be equipped with modules for collecting audio and visual information. In addition in order to receive multimedia information, WSMN can also store, process in real time, and multiplex data from disparate sources. Wireless multimedia sensor networks will not only improve existing applications, such as monitoring, automation of processing and transmission of laboratory parameters, as well as monitoring health parameters of different groups of patients, but will also allow several new applications of WSMN.

Wireless multimedia sensor networks are essentially wireless sensors composed of miniature video cameras powered by batteries, a low-power wireless receiver and a transmitter that is capable of processing, sending and receiving data. Video and audio sensors could be used to improve and complement existing monitoring systems, helping Teleradiology, Telesurgery, Telepathology, Telemonitoring, Teledermatology, Teleconsultation and Teleophthalmology to solve medical problems at the site of origin. Wide Area Networks can increase the ability of qualified healthcare professionals to register and monitor relevant patients grouped by disease, then make appropriate expert decisions and provide the necessary treatment. We propose that sensor networks in telemedicine be integrated with 4G and 5G networks in order to be used to deliver medical services everywhere [6]. In this way patients will wear medical sensors to monitor parameters such as body temperature, blood pressure, pulse, ECG, respiration, etc [7]. In addition, remote medical centers will perform remote monitoring of their patients through video and audio sensors, as motion sensors that can be built into the wrist bracelet. This will enable multimedia data such as video streams or static images together with advanced signal processing methods and techniques to be used for the detection and early diagnosis of oncological diseases or for the identification of existing malformations. In connection with the mention above, we offer a possible configuration between the Arduino board and the operating chamber in the operating room. The Arduino board can be freely upgraded and tuned to an operating camera according to the individual requirements of the operator. In this way, the general configuration of the two Finite State Automatons (Arduino board and camera) becomes an electronic eye. This allows the remote surgeon, with the help of a robot, to determine exactly how, for example, a tumor is oriented, its color, size and possibly growth to the surrounding tissues. In other words, the camera not only registers an image, but also processes that image using the Arduino board, allowing the surgeon to work perfectly and safely from a distance [3, 8].

2. Using Microcontroller Arduino as Finite State Machines (FSA)

In the context of the above, we consider it appropriate in the collection and transfer of medical data in the initial phase to use Arduino microcontroller upgradable boards consisting of 8-bit Atmel AVR microcontroller with complementary components that facilitate programming and integration into other boards or circuit boards [1].

The architecture of the Arduino Uno board is shown in fig. 1.

In our opinion, for the construction of sensor wireless networks for medical services delivery the following types of upgradable boards are suitable, as Arduino boards and modules that are for initial registration with electronics and coding as:

A) Arduino Uno is a microcontroller upgradable board with AVR microcontroller. The connection to a computer is made via a USB cable USB A - USB B. This model of boards is the most used.

B) Arduino Leonardo is a modification of Arduino Uno, microcontroller chip - ATmega32u4. The board is recognized as a keyboard or mouse by the computer.
C) Arduino Mega 2560. The board is bigger than the Uno variant. It is designed for users who require: more inputs, more outputs and more processing power.
D) Arduino Nano. Arduino Nano copies Uno, but measures 1.8cmx4.3cm. This size is suitable for making smaller projects. The Nano has the full specification of an Arduino Uno, uses the same ATMega328 microcontroller, but is more optimized in size and in this way is very suitable for real prototypes.
E) Arduino Mini R5. The mini version of the Arduino is smaller than the Nano. This board also uses the ATMega328 microcontroller chip, but is even more limited in terms of design. Removed all collector pins as well as the Mini-USB connector of the Nano variant. The board is the ideal choice if space is a priority, but on the other hand it is possible to create connection problems. One wrong connection can destroy the board.
F) Arduino Ethernet - is like the Arduino Uno, but is specially designed to connect to the Internet. It is set for automatic access regardless of data traffic. Web browsers interpret the specified text exactly as it is displayed on the screen (for example: alignment, formatting, using images, etc.). If the entered commands are known, the Arduino Ethernet can access the text directly and be used for other purposes.
G) Arduino BT. This board allows connection to Bluetooth devices in its range and allows interaction with mobile phones, tablets and other devices that have this option to support Bluetooth protocol.

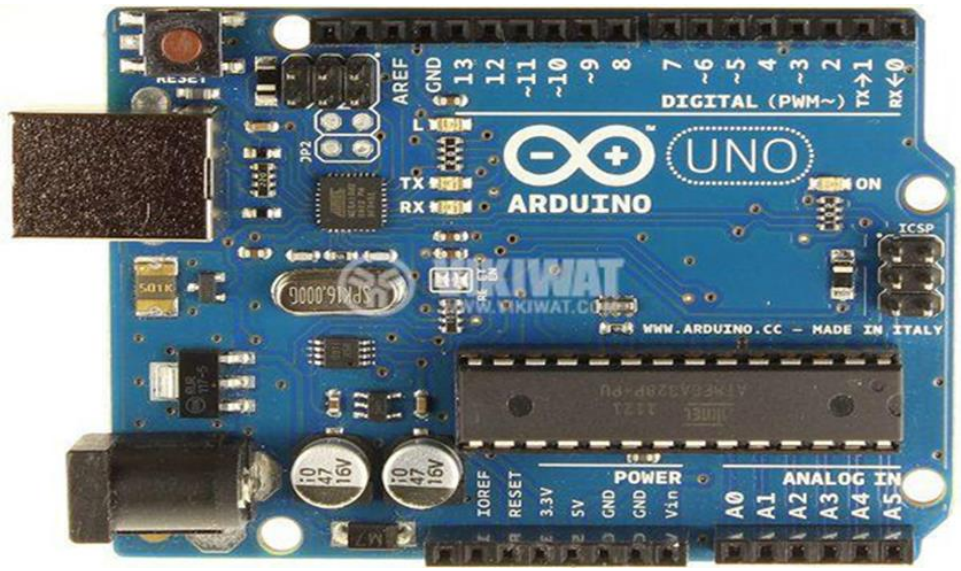


Fig. 1. Arduino Uno Board [1]

The steps in programming Arduino boards are related to the most commonly used Arduino development environment, which is the Arduino integrated development environment (IDE). It is a programming environment that allows to user to write Arduino programs in a simple Processing-based language. The methodology is as follows:

- Start IDE;
- Writing code and logic;

- Check with the verify / compile button for errors and correctness;
- Connect the USB board to the computer;
- Driver installation (PC once only);
- Select and set the serial port (once, if necessary);
- Board selection (once, if necessary);
- Programming the controller via the Upload button and testing.

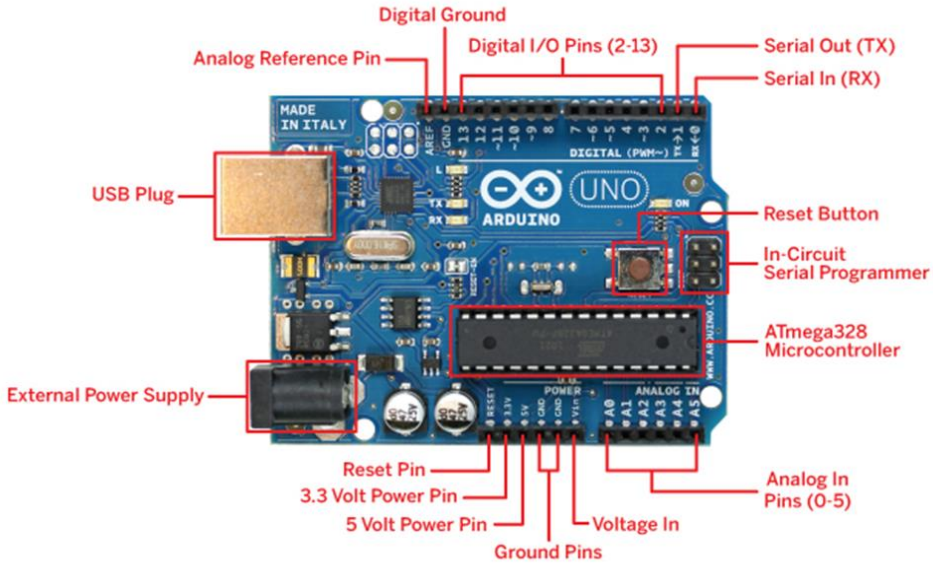


Fig. 2. Architecture of Arduino Uno Board [1].

Another device that in combination with Arduino platforms can be used, for example, in Telesurgery to achieve a good image from the operating room. It is a lamp above the operating field which is equipped with a surgical camera from the EMAICAM System. In this way it is possible to record and transmit operations in SD or HD quality [2].

Fig. 3. Operating LED lamps with camera [2]

With this surgical camera from EMAICAM-System it is possible to record and transmit operations in SD or HD quality. The central focus of the camera in the body of the lamp ensures largely undisturbed recording of the procedure and ensures the best possible quality of film recording without shadows. EMAICAM OR (operating room) camera fits

perfectly in the operating lamps of 560 and 500 series. In combination with the Arduino platform, the EMAICAM surgical chamber is ideal for [2]:

- Documentation of procedures;
- Teleconsultation in surgery;
- Transfers video stream from operating rooms to classrooms for educational purposes;

The advantage of most Arduino platforms is that they have an indicator light and a specific resistor connected between (pin 13 and ground), which is quite convenient for working with simple tests. The code used cannot be seen by a standard C ++ compiler as a valid program, so when the user clicks the "Upload to I / O board" button in the development environment, a copy of the code is saved in a temporary file in which the main function is added to validate the code. The Arduino development environment uses the GNU toolchain and AVR Libc to compile programs and uses AVRdude (graphical user interface) to upload programs to the device. AVR Libc is a Free Software that provide tool for programming Atmel AVR microcontrollers. Because Arduino uses Atmel microcontrollers, the Atmel AVR development environment can also be used to write Arduino software that uses Arduino platforms. The following option for building an internal network with IP addresses configuration for medical data transmission using Arduino platforms are shown on Fig. 4., Fig.5., Fig. 6., and Fig. 7. The figures show exemplary configurations of wireless sensor multimedia network using Aruindo boards. For this purpose, Packet Tracer simulation software for network design was used. The figures show sample settings and address configurations for switches (S4 and S1) and routers (R2) to establish network connectivity.

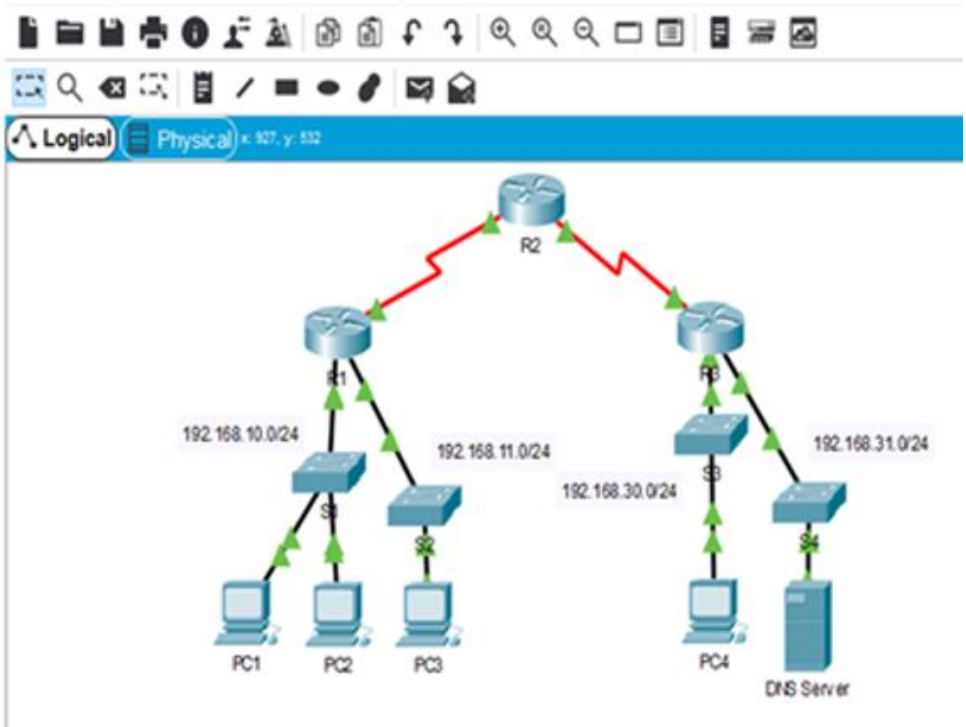


Fig. 4. Example of Packet Tracer settings configuration for wireless sensor multimedia network using Arduino boards

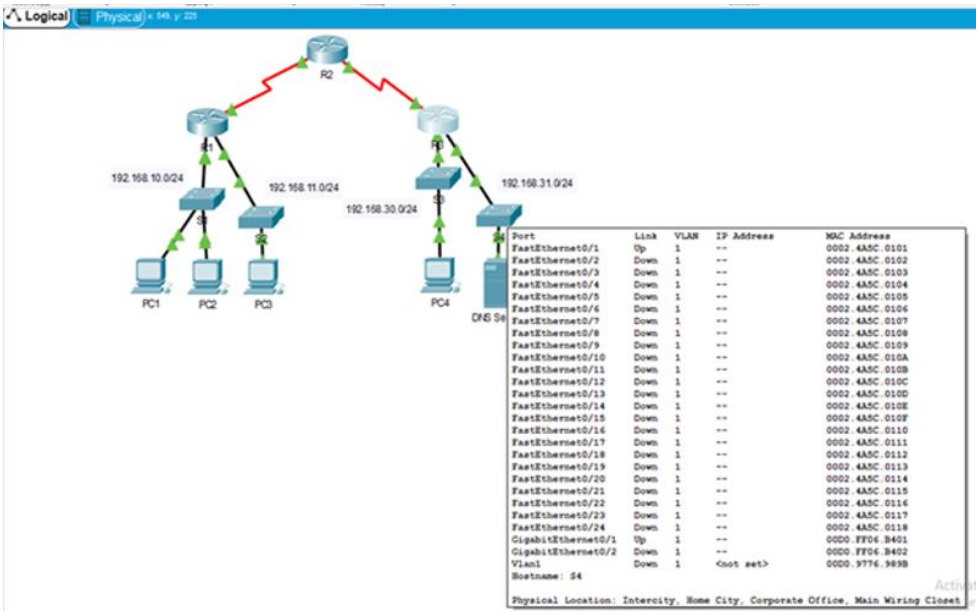


Fig. 5. Example of Packet Tracer S4 Switch configuration for wireless sensor multimedia network using Arduino boards

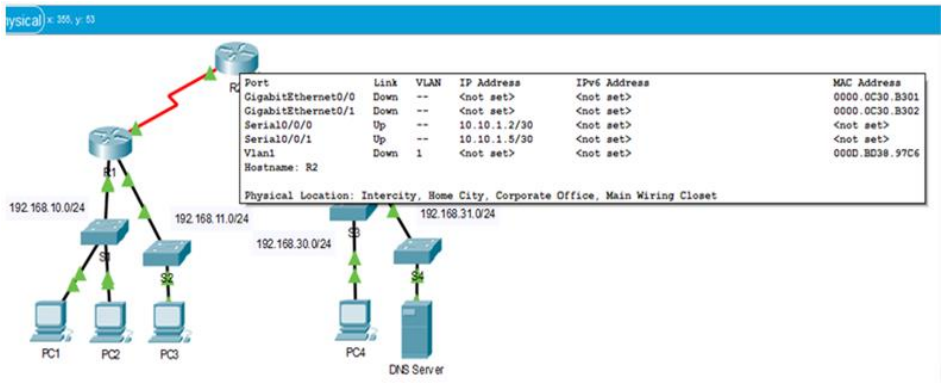


Fig. 6. Example of Packet Tracer Router R2 configuration for wireless sensor multimedia network using Arduino boards

3. Protection of Personal Health Data – GDPR

In the process of using wireless sensor networks, especially when it comes to health data processing, it is extremely important to ensure the securitization of medical information [4]. In practice, this is achieved extremely successfully with through various methods of encryption and decryption. The encryption of medical data in the specific case under consideration is a method of protection in which the information is encrypted at the source and can be accessed (or decrypted) only for a user who has the correct encryption key. Encrypted medical data is illegible to users who have gained access without permission.

The protection of transmitted medical information through encryption algorithms can be divided into two main groups - symmetric and asymmetric algorithms. In symmetric

algorithms, the same key is used for both encryption and decryption of messages. Typical symmetric algorithms are DES, 3DES and AES. In asymmetric algorithms, the encryption and decryption of the medical message will be performed with a pair of keys - public and private, the encryption is performed with the public key, and the decryption - with the private key of the recipient.

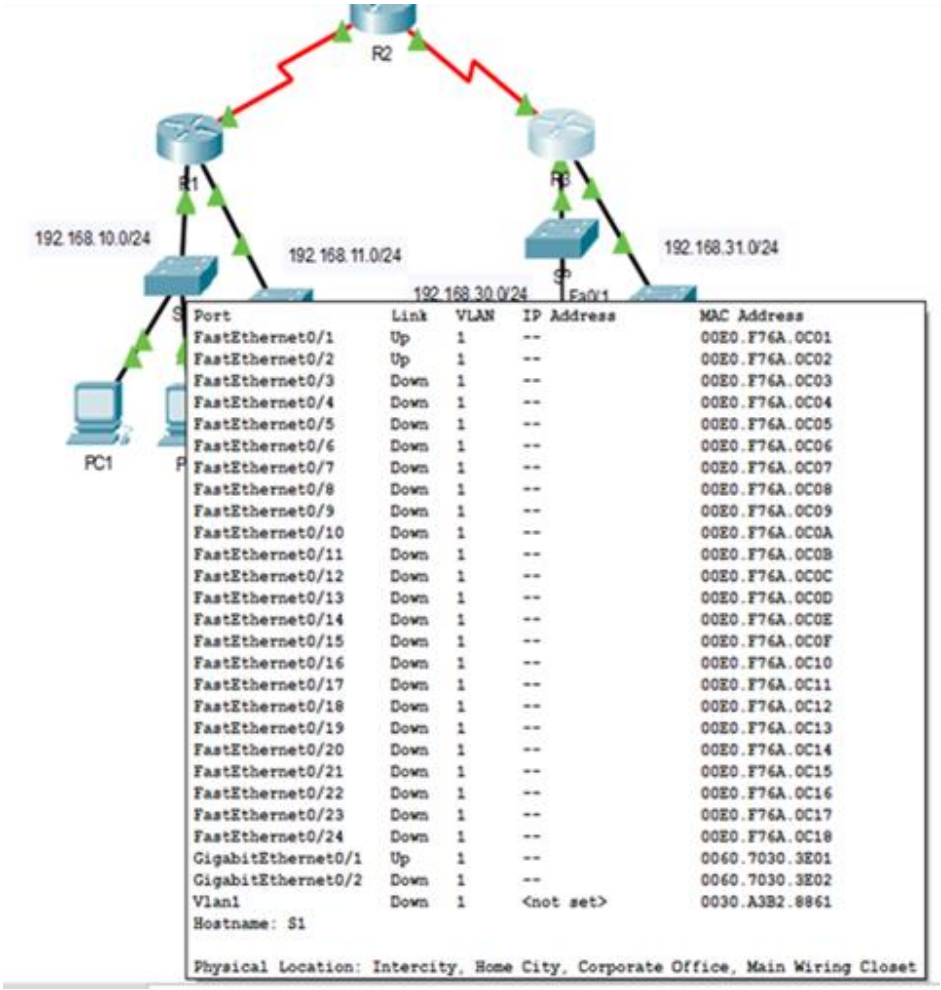


Fig. 7. Example of Packet Tracer S1 Switch configuration for wireless sensor multimedia network using Arduino boards

Typical asymmetric algorithms are: Diffie-Helman, RSA, DSA and others. The following characteristics are important in practice [5]:

1. Symmetric algorithms:
   - Only one key is needed, both for encryption and decryption;
   - The size of the encrypted text is the same or smaller than the original text (before encryption);
   - The encryption process is faster;
   - Can be used when it is necessary to transmit larger volumes of medical information;
   - Symmetric algorithms can only ensure the confidentiality of medical information.

2. Asymmetric algorithms:
- Two keys are required - one for encryption and another for decryption;
- The size of the encrypted text is the same or larger than the original text;
- The encryption process is slower than with symmetric algorithms;
- Can be used when larger volumes of medical information are transmitted;
- Asymmetric algorithms provide not only confidentiality, but also authenticity and the so-called "Non-rejection" / certification.

Common symmetric methods for encrypting information messages that can be used to transmit medical data in the specific in wireless sensor networks can be

- Data Encryption Standard (DES) - uses a 56-bit data encryption key. DES is considered insecure for use in many applications due to the 56-bit key size;
- Triple-DES - uses three consecutive DES operations, thus providing more reliable encryption than DES. The algorithm is considered to be practically secure;
- Advanced Encryption System (AES) Advanced Encryption System. Uses 128, 192, or 256 bits to encrypt and decrypt data in 128-bit blocks;

Asymmetric methods for the proposed communication solution for health data transmission can be encryption via a key pair (public key and private key). Messages will be encrypted using the recipient's public key and can only be decrypted with the recipient's private key. This method eliminates the need for the sender and recipient to share secret keys over a secure channel. All communications use only public keys, and the private key is not transmitted or shared.

4. Conclusion

In conclusion, it can be noted that good database processing in the provision of medical services is very important. This is even more true when we are in a pandemic situation, especially when wireless sensor networks are used. The quality of the medical service depends on the quality of the image or the signal. In this regard, Aruindo microprocessor boards are invaluable. In practice, they act as a microcomputer that can be placed anywhere and programmed for the full range of medical data.

In this regard, we suggest that Multimedia data such as images, together with measurements of temperature, blood pressure, blood sugar or heart rate, be used in case of critical emergencies. Computer vision techniques should be applied to images obtained from wireless multimedia sensor networks in order to assist the process of registration and transmission of physiological and anatomical parameters in the diagnosis of the respective group of patients / users. For example, in image quality control, in a surgical procedure, the Arduino programmable microprocessor board is automatically checked for image defects. Also, Arduino allows you to tune in with the recording camera in the operating room to act as a computer eye. If the operation involves the removal of a tumor or other formation, the position and orientation of the tumor can be understood and it can be treated remotely with the help of robotic hands to perform the surgery intervention from a distance.

References

1. Arduino Uno (https://datasheet.octopart.com/A000066-Arduino-datasheet-38879526.pdf).
2. IMD-International Medical Distribution, Surgical LED lamps, (https://www.imd-bg.com/index.php?lang=bg&go=view_product&id=87&cid=56&scid=).

3. Mazadjiev G., Mobile information systems for personal health telemonitoring, V. Levski National University, Shumen SEC sources: Staff Legal Bulletin No. 14E (CF), 27 Oct 2009; press release, 27 Jan 2010.
4. General Data Protection Regulation GDPR, URL: https://gdpr-info.eu/.
5. Menezes A., P. van Oorschot, and S. Vanstone, Handbook of Applied Cryptography, CRC Press, New York, 1997, p. 81-83.
6. Global Data Systems, I. (2018). Internet of Things Solutions from Global Data Systems. [online] Getgds.com. Available at: http://www.getgds.com/internetofthings.
7. Delgado, R. (2018). The Impact of Wearable Technology on Small Business. [online] TechCo. Available at: https://tech.co/impact-wearable-technology-small-business-2015-04.
8. Garden, H. (2018). How the Internet of Things Works. [online] HowStuffWorks. Available at: https://computer.howstuffworks.com/internet-of-things.htm.

Architectural Approach Usage in the Public Sector and Personal Data Protection

Pencho Popov [1], Doroteya Smilenova [1], Violeta Vasileva[2]

[1] G. S. Rakovski National Defense College, Sofia, Bulgaria
[2] Tech Tour Global, Sofia, Bulgaria
penchopopov@yahoo.com, dsmilenova@gmail.com, violetta.ziv@gmail.com

Abstract. The report considers application of architectural approach to the algorithms applicable in the process of delivery of medical services and personal data protection**.** For this purpose algorithms from the National Health Insurance Fund are shown. They could be represented using the architectural approach in the processing of medical services delivery in clinical laboratories. Processes are represented as a systemic sequence of related steps, in which each stage follows without interruption or delay the previous one and ends just before the next one begins. In clinical practice, various devices as finite state machines (FSA) are used with the help of which the information is collollected, registered, transmitted, processed and stored. The fully automated process on time during which documents, information or tasks are transferred by one participant to another is carried out through the chosen algorithms and represented by the architectural approach. For that purpose UML activity diagram for processes of Medical Services through EMDD and UML deployment diagram for Functional Configuration of the Information Infrastructure of Medical Services Delivery System through EMDD in the public sector are developed. The need for the protection of the personal data of the processed subjects (GDPR) is also discussed.

Keywords. Public Sector, Architecture, Configurations, Algorithms, Finite-State Automatons (FSA), GDPR, Cybersecurity, Clinical Laboratory.

1. Introduction

The ability to extract data and use it to generate information serving different purposes in the delivery of medical services in clinical laboratories is related to the development of appropriate algorithms. The automated execution of the procedures to process the information or its carriers aims to satisfy the needs of the healthcare to provide information of the various clinical laboratories that participate in the process of processing, storage and transmission of a medical database. In the human machine system, in practice, various devices from finite state machines (FSA) are used as a set of technical means including all machines, sensors and devices, with the help of which the information is collected, registered, transmitted, processed and stored. Thus, the activities separated in the information system can be subject to automation. On the other hand, the application of the architectural approach and its compatible algorithms will achieve the separation of data from the applications that use them. This will allow providing a logical view of the data, regardless of the physical details of their storage, providing access only to the data necessary for individual users of medical services and applications. In our opinion, in medical practice this can be achieved by technical means with the characteristics of finite state machines through appropriately developed algorithms. They are integrated with the help of an architecture model for a database on three levels:

- Conceptual schema that gives a logical view of the entire database;

- External schemas provide application or user-specific views;
- Internal schemas that gives information on data storage details and that have very little to do with the logical content of the data.

The good compatibility between algorithms and health database architecture will implement an automated system for managing the different architectural views and how these views are related to each other. For example, through the appropriate algorithms, the components of the internal schemes will be represented at the logical level by the components of the conceptual scheme. It is necessary to be able through the appropriate algorithms to reorganize the physical database without changing its logical content (conceptual scheme), or to be able to change the conceptual scheme without changing the existing external schemes. The use of architectural views in the practical solution of health database management should provide opportunities for the development of appropriate compatible algorithms. These algorithms should be linked to the need to add additional fields containing health details in the records where there were none data. In this sense, the organization of clinical examinations and the application of algorithms approved by National Health Insurance Fund (NHIF) are compatible and do not change the conceptual or internal scheme as well as do not affect existing external schemes. When choosing an appropriate algorithm and management system for a database, it must be determined to which of the three types it belongs, depending on the restrictions that are applied. Algorithms for these types must correspond to three different health structures data models: network, relational, and hierarchical. It means that, the health database model is the type of architectural structure that is most suitable for the purpose. In this case, considered by us, by applying the algorithms proposed by the NHIF for processing of the respective biological materials aims to obtain the following results:

- The conceptual scheme is the definition of this model through the language for defining the data of the respective database management system;
- Network and hierarchical models express relationships through connections;
- The relational model expresses a connection through a relation;
- Algorithms compatible with the indicated architectural approach are proposed.

2. Architectural approach – configurations and views

The Architectural approach is used to define information infrastructure of organizations. For this purpose various models and architectural frameworks [11, 12, 13] and configurations are used. Eight functional configurations possess such capabilities, for example, devices, workstations and servers, which are practically the main elements of information infrastructure. These configurations must have available functional capabilities at different levels: networking, infrastructural and others belonging to basic services and applications [5]. The architectural description includes all the products defining the architecture. The architectural development requires modeling which helps us understand the connection between the requirements on the one hand and the architectural concept on the other hand. The requirements must be presented in a chart format for easier perception. To facilitate this requirements transformation the different parts in the chart must provide a connection and transition from functional to technical and finally to application concepts. These charts or schemes are called: Functional Configuration (FC), Technical Configuration (TC) and Software Configuration (SC) [4].

The Functional Configuration is actually a logical unity of functional components which as a group can carry out a number of services. For example, the components of the informational infrastructure could be user workstations, database servers, e-mail servers, web-servers different type of sensors or FSA.

When Architecture is being developed, modeling is required that can help to understand the relationship between the requirements on the one hand and the required architectural concepts on the other. The goals will have to be transformed into architectural "building blocks" that will make easier to describe the Comprehensive, Reference and Target Architectures. So that there are no problems during this transformation process, these building blocks must provide the way for interaction from functional, through standard, technical and even to the final software concepts. Thus, it turns out that the building blocks of these phases are Functional Configuration, Technical Configuration and Software Configuration.

- Functional configurations are composed of applications, basic services and functional interfaces between them.
- Technical configurations are a collection of components and standards that interact with other technical configurations through interoperability profiles.
- Software configurations are a physical collection of software products and segments that provide software interfaces with other software configurations.

All configurations are models and should be considered as a common template, which can be presented with different functionality in terms of functional configuration, different technical solutions for the Technical configuration, as well as different product solutions for the Software configuration. The decomposition of the architectural approach leads to the separation of three interconnected architectures: operational, system and technical.

Operational architecture

The Operational architecture is a set of functionally related elements, describing: the tasks, activities, structures and information flows required for the maintenance of the system (for example, the medical services delivery system as part of public sector administration).

The operational architecture (OA) shall determine the nature of the information exchange to a sufficient extent to ensure interoperability requirements [1, 2, 3].

Functional architecture

By definition, functional architecture involves defining participants and their tasks in an operation. In order to be able to perform or maintain the defined tasks, the information flows are determined (by type and direction). The functional connections between the sources and the recipients of information are also revealed. They are defined in great detail: all participants, their tasks and the information they need to use (organizational structures, type and amount of information they need), as well as the necessary information outside their structures (global GPS, other public structures, etc.). The information flows for both use and exchange of information are also determined, as for each information, flow the time of existence, the need for it as well as the need for direct or indirect exchange of information are identified. The functional architecture, as such, allows to determine the relationships between the individual elements and objects involved in the public administrative structures, the mode of exchange and the type of information. It is important to note that the functional structure and information flows do not always correspond to the functional and administrative structures as such. In principle, the functional architecture does not depend on the technical capabilities of the CIS. However, it is a leader in the technological development of the system. In practice, there are a number of cases in which certain functions performed by elements and objects in the public administrative structures can be taken over by technology. In these cases, the technology has a significant impact on the functional architecture.

System architecture

The system architecture is determined by public administrative structures internal architecture. In essence, it is a description of the systems and connections that provide or

support the functions of the OA. System Architecture (SA) shows the interaction between the systems. It contains the specific set of services, physical connections and location of its nodes and networks, as well as such parameters of the individual components as average time between failures, maintainability and survivability. According to DoDAF [1], the system architecture includes system views.

The system architecture itself, in the whole range of all phases of the medical service delivery process as part of public sector processes, must remain relatively unchanged. The system itself is presented as a complex system with the characteristic features of communication - information system such as: multifunctionality, the presence of a large number of subsystems and elements, probabilistic nature of the processes in it; large spatial parameters; availability of management system, etc.

The specific architecture always includes a representation of the physical connections, the location of the elements, nodes, networks, etc. located in the public sector information infrastructure.

Technical architecture

The technical architecture defines the rules for the construction of SA. It includes a set of technical standards, conventions, rules and criteria that help to regulate the system services, interfaces and interconnections for SA by connecting to a specific OA. From the point of view of system-system compatibility, TA (technical architecture) outlines the criteria or "rules" for technical implementation, which the system must adhere to as set out in the SA. TA in practice ensures the interoperability of the systems horizontally and vertically. The important goal of TA is to determine the standards and rules that govern SA and the functioning of CIS.

This includes the minimum number of rules, algorithms, norms, standards and interfaces allowing the subsystems and elements to exchange information in accordance with the requirements. The technical architecture is the basic list of specifications, standards, rules and criteria that govern the services in the system, the interfaces and the interaction between systems, as well as the form in which the information is presented. The technical architecture defines the set of standards, rules, algorithms and norms allowing the functioning of CIS.

The profile of the technical architecture is determined, which should correspond to the type and capabilities of the platforms building the individual systems. The technical architecture should allow for embedding new elements in the system with new standards, gradually excluding the old ones, while it should approach to the commercial standards. The technical architecture is the main limiting factor in the planning of CIS. Requirements for the adopted standards allow unimpeded interaction between the systems, as well as act restrictively on the possibilities for development and installation of new systems. TA also influences the functional architecture in the course of immediate planning.

3. Architectural Approach Implementation in Finite-State Automatons (FSA)

The automated execution of procedures over information or its carriers aim to meet the information needs of decision-makers. During this process we propose to apply a systematic architectural approach and to automate not the solution of the individual task but the procedures and operations of the objects in the information system.

In practice, this can be successfully achieved by using different devices in the human machine system such as state machines through which the activities separated in the information system should be subjected to automation. In this way, when determining the object of automation, the data will be separated from the applications that use them and providing a logical view of the data, regardless of the physical details of their storage,

providing access only to the data necessary for individual users and applications. [9, 10]

In our opinion, in practice this can be achieved with the help of finite state machines, which are integrated with the help of a database architecture model on three levels: conceptual schema, external schemes and internal schemas.

4. Algorithms

Algorithm 1: For Registration of a patient and organization of a medical service through the Electronic Medical-Diagnostic Direction (EMDD) an ongoing procedure for clinical examination in a clinical laboratory is approved and suggested by the NHIF [7, 8, 17, 18]. The algorithm can be represented as a processes in public sector structures consisting of relevant medical services delivery activities according to the Architectural approach, as follows:

1 Physician issues *EMDD* in his health information system, enters *NHIF* codes and prints it out;
2 The patient takes the *EMDD* form and goes to the Clinical Laboratory Registry (CLR);
3 Gives the form and his ID card and re-enters in the laboratory registry the data from the printed *EMDD*;
4 The clinical laboratory takes blood or other biological material from the patient;
5 The patient is examined according to a certain methodology, the results of the examination are entered manually or automatically in the Laboratory Information System (LIS);
6 An ongoing procedure marked as *EMDD-2* (test results form) follows and the results are given to the patient on paper and / or on a website with an username and password for data access;
7 Physician can get the results on paper and / or via the Internet (LIS) if the patient gives him his name and password;
8 Physician must manually enter the test results into his or her Health Information System (HIS).

Algorithm 2: a clinical laboratory current procedure for processing Electronic medical-diagnostic direction (EMDD) is proposed and approved by the NHIF:

1 Family physicians and specialists enter EMDD in their health information systems and when printed, the information is stored in a cloud EMDD server;
2 The patient goes with the printed EMDD to the Clinical laboratory registry / reception;
3 The laboratory system looks for information on EMDD on various indications.
4 The data in the Laboratory Information System (LIS) are automatically imported from the EMDD server as EMDD-2;
5 LIS automatically transmits tasks to the respective devices on the basis of entered barcodes when taking blood;
6 The devices complete the tests and transmit the results to the LIS;
7 The laboratory doctor validates the results and they are recorded on the Laboratory's Internet Server (LIS). LIS includes the Laboratory's Internet Server;
8 Physician who issued the EMDD checks the EMDD server periodically, through his health information system (HIS);
9 When the results are ready, they are automatically imported from the laboratory's Internet server (LIS);
10 Results are stored in the cloud server;
11 Physician can now review the results and make a chronological and clinical analysis.

As examples for defining informationaл objects, respectively with their names and parameters, for the architectural model can be indicated the codes approved by the NHIF for erythrocytes in a blood test, as follows:

Example 1. Informational and software objects with their names according to the electronic 01-03 EMDD using codes proposed by the NHIF for ESR (Erythrocyte sedimentation rate):

Logical Observation Identifiers Names and Codes (LOINC) – **Names**;
30341-2 Erythrocyte segmentation rate;
43402-7 Erythrocyte segmentation rate by 15 minute reading;
82477-1 Erythrocyte segmentation rate by Photometric method;
4537-7 Erythrocyte segmentation rate by Westergen method;
18184-2 Erythrocyte segmentation rate by 2H Westergen method;
4238-5 Erythrocyte segmentation rate by Wintrobe method;
4539-3 Erythrocyte segmentation rate Zeta by Zetafuge;

Example 2. For each name of the object there are respectively up to seven parameters according to the electronic 01-03 EMDD adopted by the NHIF for ESR (Erythrocyte sedimentation rate):

Logical Observation Identifiers Names and Codes (LOINC) – **Parameters:**

1 Component (Component) to be measured: In this case it is the erythrocyte sedimentation rate.
2 Measured property (Property) - in this case speed.
3 Timing (Timing) - whether the measurement is an observation at the current moment or is a set of measurements for a long period of time.
4 System (System) - the type of system - urine, blood.
5 Scale (Scale) - whether the measurement is quantitative, nominal or descriptive (free text).
6 Method (Method) - the way in which a given result is obtained or a certain observation is performed.
7 Units of measurement (exUnits) - sample units of measurement (mm / h, mm / 15min, mm / 2H).

Activity diagram of Medical Services Delivery Processes through EMDD is part of the Operational architecture of the system and is shown on Fig. 1. According to the NATO architectural framework, such a diagram is called an Operational Activity Model - NOV-5, and respectively according to the DoDAF framework, the diagram is called Operational Activity Model - OV-5. Operational Activity Model. It was developed by usage of the Enterprise Architect software program [6, 14, 15].

As it was already mentioned, Operational architecture (OA) represents the Operational activity model of the organizational processes. This OA formally describes the system processes, activities and events and consists of UML's activity (behavioral) diagrams, which present the basic main system processes, and their separation by structures/ partitions and consequences.

These activity diagrams contain notations (graphic images) depicting main organizational structures with conducting activities of the Processes of Medical Services Delivery through EMDD, as they follow:

- Partitions (Physician cabinet, Clinical Laboratory Registry, Clinical Laboratory, etc);
- Start and End notations;
- Activities for each partition (patient registration, electronic register etc.);
- Connectors.

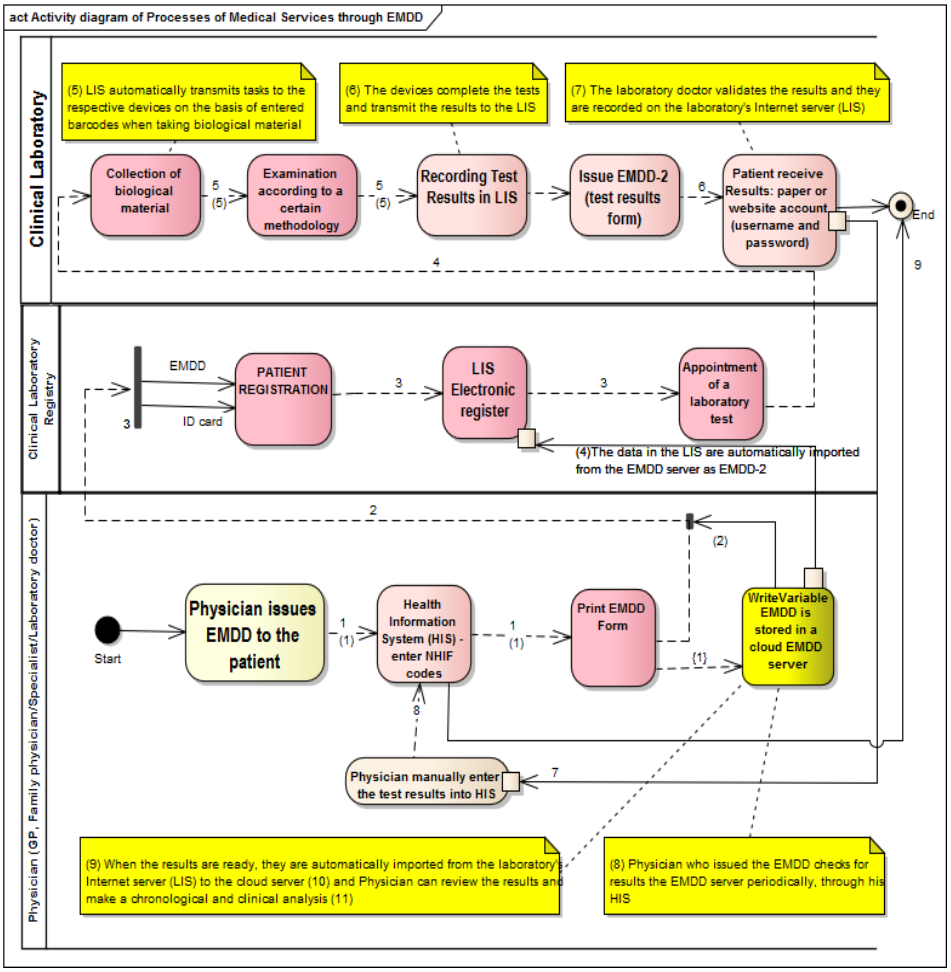


Fig. 1. Activity diagram of Processes of Medical Services Delivery through EMDD

System view

The Medical Services Delivery through EMDD in the Public Sector System view (SV-1), which was developed by usage of the Enterprise Architect (EA) software program, is shown on Fig. 2.

As it was already mentioned, SV-1 represents the Description of system interfaces of the Medical Services Delivery System informational infrastructure. This SV-1 formally describes the system interfaces and connections and mainly consists of UML's deployment (structural) diagrams, which present the basic main system of the CIS and their interfaces.

These deployment diagrams contain notations (graphic images) depicting main system interfaces of the Medical Services Delivery System through EMDD, as they follow:

- Nodes (software and hardware components)
- Connectors between nodes or systems/subsystems interfaces.

Additionally, according to the technical reference model system services necessary for the information infrastructure of organizations should be defined. Functionalities needed by users and based on system services are also identified using reference models like specialized

workstations, servers- e health server, Database Server, terminals, printers and others which are connected in an integrated information infrastructure of the Medical Services Delivery system through EMDD. A specification of the technical and software configuration of the automated information systems and networks is offered as part of the system design of the information infrastructure which guarantees interoperability in cyber space. This also can be seen as a prerequisite for improving human-machine interfaces in Cyber security.

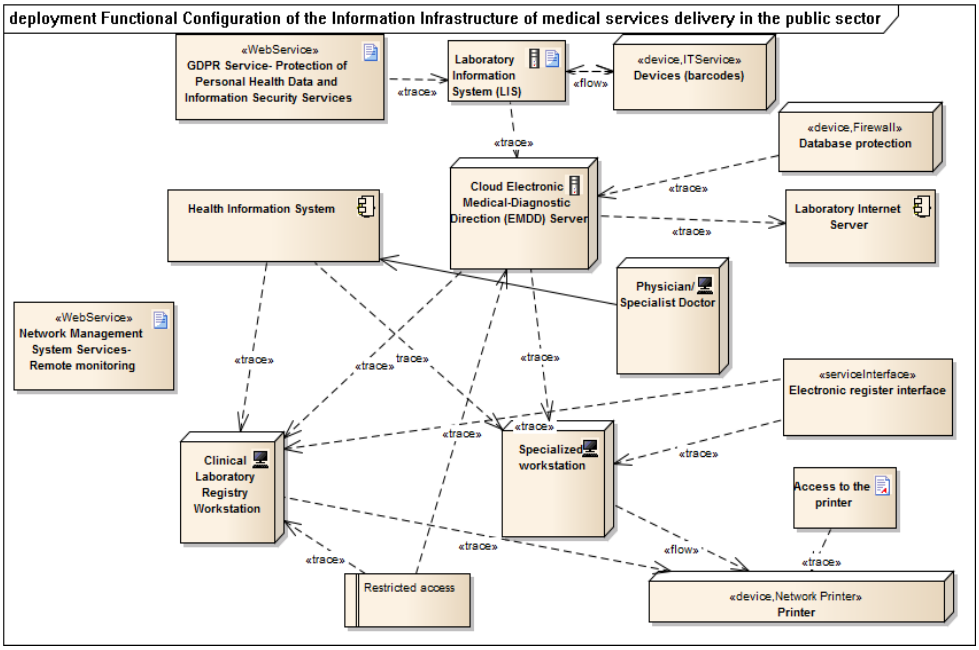


Fig. 2. Functional Configuration of Information Infrastructure of Medical Services Delivery in the Public Sector- System view (SV-1)

5. GDPR- Protection of Personal Health Data

Designing and building complete, integrated security solutions to prevent and monitor unauthorized access, abuse, modification, or denial of service is critical to the processing of medical data. Solutions that help ensure business continuity and safety must maintain user confidentiality, reduce operating costs by offering multifunctional approaches, be affordable and contain integrated components. They should focus on the core business of the healthcare provider, devoting more time to strategic growth and less time to network security issues. The protection of personal data in medicine in the proposed algorithms for the delivery of medical services is an absolutely mandatory action, both by law and from a moral point of view and can be done through:

Privacy by Design - healthcare service providers are required to include data protection measures at the outset of the introduction of new systems and algorithms. The personal data that are collected must not exceed the data needed to perform their duties.

Data Protection Officer - introduction of a new role in the organizations dealing with medical data, whose main activities require work with large amounts of personal data and which is responsible for internal rules, security policies, data storage, compliance and development of secure algorithms.

There must also be explicit consent to the provision of personal data to patients or other users of the health service, and the algorithms used for this must be developed only in connection with the objectives of the clinical investigation. In order to protect health personal data, anyone who processes such data and makes the appropriate algorithms for it must perform a GAP analysis. GAP analysis is an audit of the current situation, which establishes whether existing technological and organizational measures meet the requirements of GDRP. Specific analyzes and assessments in order to implement the information security must take into account:

- The organization and accountability in the processing and use of personal data;
- The degree of centralization of data protection;
- Data protection levels;
- Levels of data consistency;
- Data compromise notification mechanisms;
- Actions in international data transfers;
- Data protection roles and responsibilities;
- Overall level of compliance with GDPR.

In order to realize the real protection of medical data and the use of established and protected algorithms, the analysis is not sufficient, but it is necessary to introduce and implement organizational and technical measures and processes [16]. Depending on the needs of the respective organization, this stage may include development and implementation of processes, procedures and algorithms for elimination of discrepancies in:

- Defining the necessary changes in the medical processes of data processing and security policies;
- Preparing to make changes in the IT environment for data processing;
- Defining appropriate control mechanisms when working with data;
- Development of advanced reporting and notification mechanisms;

In practice, a number of automated tools as Finite State Machines (FSA) are used in the delivery of health data, for the implementation and maintenance of process continuity solutions, such as Next Generation Firewall, DLP, PKI, Two-factor authentication, encryption, Application Control, Access Control, File & Data Transfer and many others, both on premise and cloud-based. Healthcare data administrators and algorithm developers responsible for this implementation must protect the system by:

- Define - create and / or update internal rules and security policies;
- Carry out development and verification of the changes made in the IT environment and data sets, medical processes, internal controls, work organization and reporting;
- Implement and test the created rules and the made developments;
- Conducting an internal audit - general verification of compliance with the requirements for cybersecurity and GDPR.

Performing penetration testing is mandatory for the security and protection of health data, as well as the implementation of monitoring systems, SIEM and Vulnerability Management.

6. Tools for Monitoring, Control and Diagnostics of Networks Using the Proposed Algorithms in the Medical Services Delivery

The standardized approach for managing the network environment in view of the proposed algorithms and its protection is that all its elements - protocols, bridges, routers, gateways to be managed as network objects. The management of a network object by

algorithms can be considered in several aspects:

- Direct fault management (fault management);
- Direct performance management (performance management;
- Direct layer management (layer management);
- Direct security management (security management).

In the interests of reliability and security, the network should be monitored regularly to correct and rectify problems. Various tools are available to help the administrator monitor, manage, and troubleshoot LAN and WAN environments. They range from simple TCP / IP programs to complex intelligent software packages and hardware devices. Key concepts in analyzing and optimizing network performance and security are:

- Proven bottlenecks (botlenecks) - this is a point in the system that limits performance and can be important for its safety;
- Established baselines (baselines) - comparing different measurements, such as bytes or packets, with those previously measured, is important for safety;
- Proven best practices (best practices) - performing administrative and other network tasks in the most efficient and cost-effective way of safety.

Network monitoring software can also be used to protect personal data. Network monitoring applications are called protocol analyzers or sniffers. They are designed to intercept packets / frames that are transmitted between two or more network devices. A sniffer is any program that eavesdrops on network traffic. The proposed algorithms for processing, delivery and archiving of biological materials can be compliant with network monitoring tools such as: Wireshark, Sniffer Pro LAN; Sniffer Pro WAN; Sniffer Pro High Speed; Gigabit Sniffer Pro; Sniffer Distributed Analysis Suite, etc.

7. Conclusion

In order to increase the efficiency in the delivery of health services, it may be decided to reorganize the storage characteristics in the database, but combined with established and imposed by practice algorithms. Although this could affect the internal schemes in an architecture, it will not affect either the conceptual or the external schemes. The main principles in building system architecture are: maximum compliance of the qualities of the structure with its goals and objectives, compliance between the functional capabilities of the Communication and Information Systems (CIS) structure with the functional interactions between the participants in the processes and their tasks, determination of the interaction points and interfaces between different CIS systems based on their structure, as well as the system structure corresponding to the technical characteristics and capabilities of the system components.

It can be summarized that the Architectural approach and the selected algorithms for clinical processing of a medical database are strict sequence of steps in the planning process. It is above all a way of thinking and a philosophy in understanding the complex interactions between human-machine systems in implementation of automated processing of clinical and laboratory indicators in the delivery of health databases. The architectural approach allows us to use appropriate algorithm, to define processes and to look at CIS as a whole, of organically connected elements, functions and technology.

Protection of the personal data is also considered. Tools such as UML diagrams have been used to describe informational infrastructure and processes. We hope that information infrastructure and its architectural views combined with appropriate algorithms and processes could successfully be used to improve the system design, finite state machines implementation, human-machine interfaces and GDPR- protection of personal health data in order to meet the

challenges of cybersecurity in the public sector administration processes. From that perspective the described architectural approach usage is a promising area for future research.

References

1. The DoDAF Architecture Framework, U.S. Department of Defense, , (https://dodcio.defense.gov/Library/DoD-Architecture-Framework/).
2. NATO Architecture Framework v3, NATO Consultation, Command and Control Board, AC/322-D(2007)0048
3. NATO Architecture Framework Version 4, Architecture Capability Team, Consultation, Command & Control Board, January 2018, (https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2018_08/20180801_180801-ac322-d_2018_0002_naf_final.pdf)
4. NATO C3 Technical Architecture Implementation Handbook (NC3TA-IHB), Version1, ISSC NATO Open Systems Working Group, AC/322(Sc/5)N268, 2002.
5. Allied Data Publication 34(ADatP-34) NATO C3 Technical Architecture, Volume 2 Architectural Descriptions and Models, Version 6.0, ISSC NATO Open Systems Working Group, NC3TA-Vol2-v6.xml, 2004.
6. Kobryn C. & Sibbald C. (2004), Modeling DoDAF Compliant Architectures. The Telelogic Approach for Complying with the DoD Architectural Framework (White_Paper_Modeling_DoDAF_UML2.pdf, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.304.6686&rep=rep1&type=pdf).
7. National Health Insurance Fund, Appendix №19, (http://www.en.nhif.bg/).
8. National Health Insurance Fund - Instructions for Using the Electronic Medical File Review Service from Health Insured Persons, (https://www.president.bg/docs/1352303612.pdf).
9. Hopscroft J.E., Motwani R., Ullman J. D., "Introduction to Automata Theory, Languages, and Computation" Second Edition, Adison-Wesley Publishing Company New York, 2002
10. Karpov U.G., " Theory of Automata " Piter Publishing Moscow, 2003
11. DoD Architecture Framework, Volume I: Definitions and Guidelines, 23 April 2007.
12. DoD Architecture Framework, Volume II: Product Descriptions, 23 April 2007.
13. DoD Architecture Framework, Volume III: Architecture Data Description, 23 April 2007.
14. Enterprise Architect 9.3 Reviewer's Guide, Sparx Systems, 2012.
15. Enterprise Architect User Guide, Sparx Systems Pty Ltd, 2012.
16. Aleksandrova V., Nickolova-Kotseva T., Zakonovata ramka, svarzana sas sigurnostta na informaciata, Nauchni trudove 1, Sttrategii I politiki na sigurnost, MNK, Sofia, 2010, izd. "Balkanski forum po sigurnostta", s.264-273, ISBN: 978-954-91632-4-7 / The Legal Framework Related to the Information Security, Balkan Security Forum, Sofia, 2010, pp. 264-273, ISBN: 978-954-91632-4-7
17. National Health Insurance Fund, Appendix №4
18. National Health Insurance Fund, Appendix №18

Information Sharing Processes in Public Sector Cybersecurity Structures

Violeta Vasileva

Tech Tour Global, Sofia, Bulgaria
violetta.ziv@gmail.com

Abstract. In this report the author analyses interaction processes in the E-business and Supply Chain Network concepts in order to propose a model of data sharing among information security organizational structures in the public sector.

Keywords. Information Sharing, Processes, E-Business Model, Public Administration, Information Security Structures.

1. Introduction

The model for cybersecurity of the Republic of Bulgaria is established by the newly adopted Law on Cybersecurity of October 31, 2018 and the National Cyber Security Strategy "Cyber Resilient Bulgaria 2020" [1]. The strategy envisages improving the interaction and information sharing between the state, business and society, as well as establishing effective mechanisms for information sharing and engagement of all stakeholders, which includes public sector organizations.

It is assumed that cybersecurity structures are built in public administrative organizations. Achieving information security in the administrative structures is ensured by building an information security system operating within a Computer Incident Response Centers. These centers are considered as organizational structures for information security in the public sector.

At different levels of government, there are different structures related to cybersecurity. They are presented in the diagram as model of the national cyber security and resilience system.

Gaining insights from B2C industry which has a direct impact on processes in B2B it is possible to use E-business concept and its implementation in B2B industry in order to develop a model of information sharing processes in public sector organizations and their Cyber Incident Response Team, (or Computer Incident Response Teams) (CIRT, CSIRT, CERT - Computer Emergency Response Team) or their other information security structures.

From business perspective Internet technologies open the opportunity to do business online – electronic business (E-business) and to implement Security as a Service concept as cloud technology in order to support information security in public sector.

Globalization and technological progress are pushing large international corporations as well as public administration structures into moving towards change. Latest advances in business analytics and Information Technology (IT) also allow for better performance and monitoring than in the past. Information and Communication Technologies (ICT) are considered one of the key drivers for global economic transformation. They are shaping the new economic landscape of the 21st century - the digital economy.

Public organizations which are successfully applying E-business can convert data, share information and conduct electronic transactions with their business partners - partner

organizations through the information and communication technologies. Fundamental changes in world economy, combined with rapid growth in technology advancements irreversibly transform the security environment and Information sharing processes as part of it.

Another trend is the amount of information and the enormous number of interactions which users start making around the world [11].

To build a model of information sharing processes between the structures for information security in the public sector, the concept of E-business can be applied. Electronic business (E-business) is the exchange of information via electronic media such as the Internet and private communication networks, enabling the processes of buying and selling goods and services, as well as the transfer of funds and data, over an electronic network. It also involves the adoption of innovative business concepts, such as dynamic pricing through online auctions and direct online sales to users. These business transactions usually represent a business model or a side of it and appear in business-to-business (B2B), business-to-consumer (B2C), consumer-to-consumer (C2C) or consumer-to-business (C2B).

Structures for information security in the public sector are located in a hierarchical organization. According to the Cybersecurity Act in the model of the cybersecurity system the different structures are divided into three levels - strategic, operational and tactical.

In order to build the model of the processes in the cybersecurity structures, applying the Ebusiness model, it is assumed that C (consumers) are users or Computer Security Incident Response Teams (CSIRT) at the tactical level.

Business-to-Business (B2B) processes refer to the relationship between computer incident response centers of public sector organizations. They refer to the information exchanged between organizations or businesses in the form of products, services, and information.

Typical for B2B is that it indicates public organizations as business entities; it also could cover a large number of applications which enable the organization to form relationships with their personnel, users, partner public organizations etc.

A classical B2B model is illustrated in the following graph (Fig. 1.):

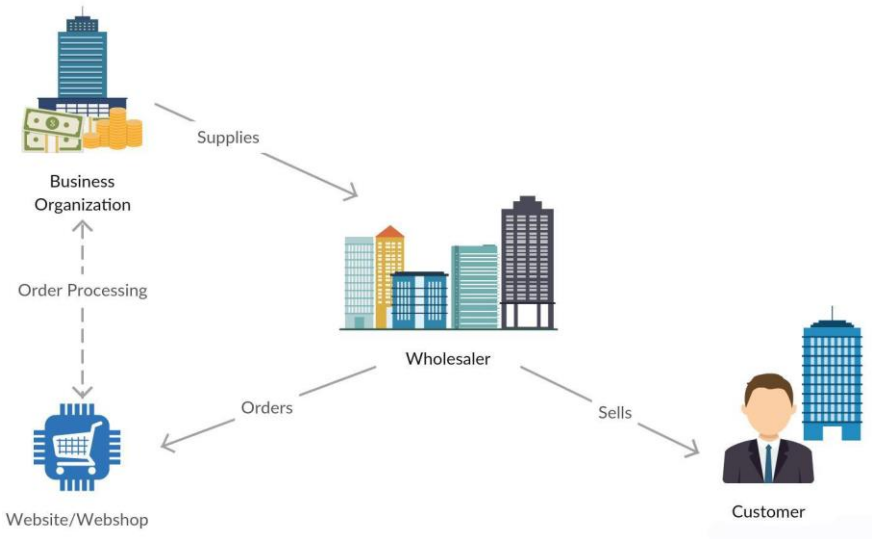


Fig. 1. Classical B2B business model

As we can see the website for example stands as an intermediator between the two business entities - the organization and wholesaler which could be security service provider. In a lot of cases, B2B represents the B2B2B connection – meaning organizations that operate on the very background level of processes [10].

The strive for innovation and improvement emerges from the B2C market and has a direct impact on the B2B. Mostly because the final users in the B2C are also taking part of in the B2B operations. This fact makes the user expectations higher regarding the digital service.

For end users, the emphasis is on the importance of speed and ease of use, self-service demand for more personalized service. They also require the recognition of loyalty and value greater self-reliance [8]. Their expectations evolve around business which operates 24 hours a day, seven days a week, with constant visibility, easy access, manageability, quality, and effectiveness [3].

Types of B2B interactions

As mentioned before, there are several ways which enable the public organization to form a relationship with its business partners. Here are listed some of the most general ways (Table 1.).

Table 1. Types of B2B interactions

Type of B2B interactions	Description
Company Websites (Website)	The target audience is employees and other organizations. In some cases, the website is also an entrance to an extranet available to registered users. Some websites are used as a direct sell platform to other public organizations. In many cases, B2B websites contribute to building the company presence and trust through blogs, vlogs, discussion forums etc.
(Security) Product Supply & Procurement Exchanges	Also known as EProcurement sites, where an enterprise purchasing agent can place orders for supplies from vendors, request proposals for security products or services and even sometimes to make a bid for a desired price. The platform can be industry specific or with focus on a niche market. In some cases, besides from buying and selling, the site can also provide broader services like information, product listings, discussion etc.
Brokering Sites	An intermediary between the buyer and seller of security products and services – for both products and services.
B2B Cybersecurity Platform Providers	A SaaS provider for B2B platforms, including site-building tools and templates, database, methods and transaction software too.

Electronic Data Interchange

The evolution of technology has played a prominent role in addressing the complexities of global supply chain. The first signs of E-business transactions were noticed more than 40 years ago with the introduction of Electronic Data Interchange (EDI). It started in the 1960s by the railroad industry. The reason was the need for faster and more efficient way to communicate information about goods being transported. Initially the companies managed the file transfer and transformation on their own. In the 1980s many companies adopted EDI, which enabled atomized documents to exchange between enterprises in the public sector by bringing computer-to-computer communication into place. The communication process

between public organizations became easier and quicker. The greatest benefit for the public organization was that it could now exchange invoices, orders and other documents to one another's computers [9].

Today, all types of business documents for different industries such as retail, automotive, logistics, and banking can be exchanged using EDI. The reason is that there are several standards along with industry specific standards in use today:

- ANSI X12, commonly used in North America
- EDIFACT commonly used worldwide
- ODETTE, specific for the automotive industry
- EANCOM, subset of EDIFACT which incorporates a system of product codes in order to specify products around the world, also known as European Article Number (EAN)
- HIPPA, USA standard for electronic healthcare transactions and national identifiers for providers, health insurance plans and employers
- Rosetta Nett, based on XML facilitates industry-wide global supply chain processes
- SWIFT, The Society of Worldwide Interbank Financial Telecommunications (SWIFT) standard for management of financial activities such as payments, trade services, and securities
- VDA, standard developed by the German Automobil Association (Verband der Automobil Industrie) and used by vehicle companies.

EDI standards have been developed in order to provide a common data format, enabling computers to communicate and exchange documents without people or paperwork involved. Processing can be initiated on the receiver's computer, where processing can be initiated immediately.

There are also two general ways to transmit an EDI document. The first one is by sending it directly to the business partner, usually via the Internet. The second one is by using the services of an EDI Network Service Provider, where the document is sent first to the provider. The latter makes it available to the business partner. EDI is sufficient for orders or requests for information frequently placed regarding standard and repetitive products. B2B website will not have the capacity to handle these orders as efficient as EDI does. On the contrary, if a company works only with EDI, it will have to implement traditional and manual methods, like a phone service, to cope with non-standard orders. However, the manual work is typically characterized by time inefficiency and error frequency. Therefore, an automated B2B website could process non-standard order more efficiently.

Compared to EDI, B2B E-commerce creates a way of giving richer information about products, for example when a company wants to inform their user about the products offered [2].

Industry 4.0 Impacts

Industry 4.0 stands for the current trend of automation and exchanging data in the public sector, which could be involved in B2B or more particularly B2B2B interactions.

Two of the fundamental design principles of Industry 4.0 are:

Interoperability - The ability of machines, devices, sensors, and people to be able to connect and communicate with one another through the Internet of Things (IoT)

Decentralized decisions - The cyber physical systems capability to make decisions on their own and perform the tasks as autonomously as possible

These innovations have a direct impact on collection, production, distribution, and delivering activities. They are the core components of the supply chain and respectively to information sharing processes [5].

2. Using Supply Chain Management Model in B2B Environment for Developing Information Sharing Process Management Model in Cybersecurity Organizational Structures

Supply Chain Management (SCM) comprises both of information and physical flow. Information flow allows the various partners in the supply chain to coordinate their long-term plans, and to have control over the day-to-day movement of information and respective documents throughout the supply chain.

The information exchange system and related activities is to manage the production and distribution processes quickly and efficiently.

The information sharing processes can be rather complex, especially for B2B businesses that do not own the whole information sharing process. When multiple organizations participate in these processes, it is necessary for all of them to have the same information and to be interoperable. That could be quite challenging. If a misunderstanding or miscommunication at some point and at any time in the information sharing occurs, it will bring about a slowdown in the whole process.

Although B2B and B2C (business to consumer) models have some similar issues concerning the supply chain management anal respectively information management, the B2B activities faces the unique challenges. The end user in the B2C information process expects to receive his data or service as soon as possible. In contrast, the B2B model requires the delivery of the product service and information) by a specific date and time. B2B processes are associated with tight deadlines.

Nowadays, the Internet and communication technologies (ICT) in the face of E-business have made SCM much easier and accessible for the B2B environment. Companies can reach more available suppliers and create partnerships that may not have been possible before. The internet also allows for B2B E-business activities such as the use of information available online to compare various prices and members of their supply chain, thus enabling them to reduce the operating costs. It also provides access to information on all parts of the supply chain.

E-business Architectural Models in Supply Chains and Information Sharing Management

In the information management model B2B environment can be associated with connections among the security operation centers of different public organizations. As mentioned before conducting E-business transactions in the B2B environment is possible in several ways. When a company exchanges business or cybersecurity data electronically with its partners (users, public organizations, providers of cybersecurity services, suppliers, and financial institutions), one of the critical decisions is regarding the communications options to connect to each other.

There could be defined four main approaches for connecting public organizations which are as follows:

• Direct Connection Model

The model provides for direct communication with the public organization, where the business can directly exchange business documents. The company itself is responsible for mapping, translation, technical support and tracking documents. This model is suitable for a relatively small community, less than 100 public organizations. In case the community grows, more resources need to be implemented for the support of new partners - such as monitor communications, manage public organization calls and resolve issues quickly. Quick issue resolution is essential since the documents being exchanged (e.g. orders, invoices, ship notices) are frequently the core element of your business. Additionally, public organizations often insist

on using different protocols, especially when they are trading with other enterprises too. Meaning supporting multiple protocols, which require additional resources.

The figure below represents the direct B2B model. (Fig. 2.) The business is shown as the "Enterprise" connecting with four partner organizations who are sharing information with other partner organizations as well. However, nowadays this model is not preferred by organizations because of the supporting complexity.

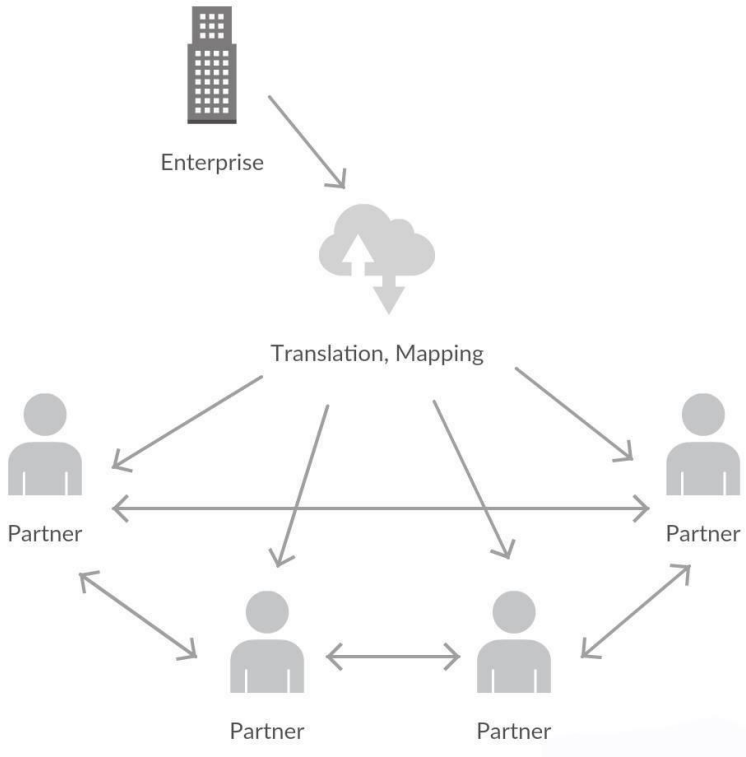


Fig. 2. B2B Direct E-business Model for information sharing among public organizations

• Network Model

When more security operation centers or partner organizations are involved, many public organizations have the option to work exclusively through a B2B Security Service Provider, where there is a single connection to the Service Provider of security services nevertheless of the protocol used. Likewise, the public organizations are also connected to the Service Provider, each being able to choose their protocol requirements. Thus, each public organization Security Operation Center (SOC) makes an independent decision regarding the preferred connectivity protocol and depends on the Service Provider to mediate the connectivity between protocols. The Service Provider facilitates the exchange of electronic documents via its network. The provider also supports the community members regarding communications issues; ensures data security and non-repudiation; and provides audit information, reporting, backup, and recovery. Transaction fees are charged for the services. The public organization still takes care of mapping and translation as well as some reporting and translation-related technical support.

This model solves the challenges regarding interoperability between different systems.

The figure below illustrates the network model of B2B (Fig. 3.). The public organization is shown as the "Enterprise" connected with the Service Provider through a single

communications protocol. Similarly, each partner organization is linked to the Service Provider while implementing their preferred protocols.

This model is used by many companies today.

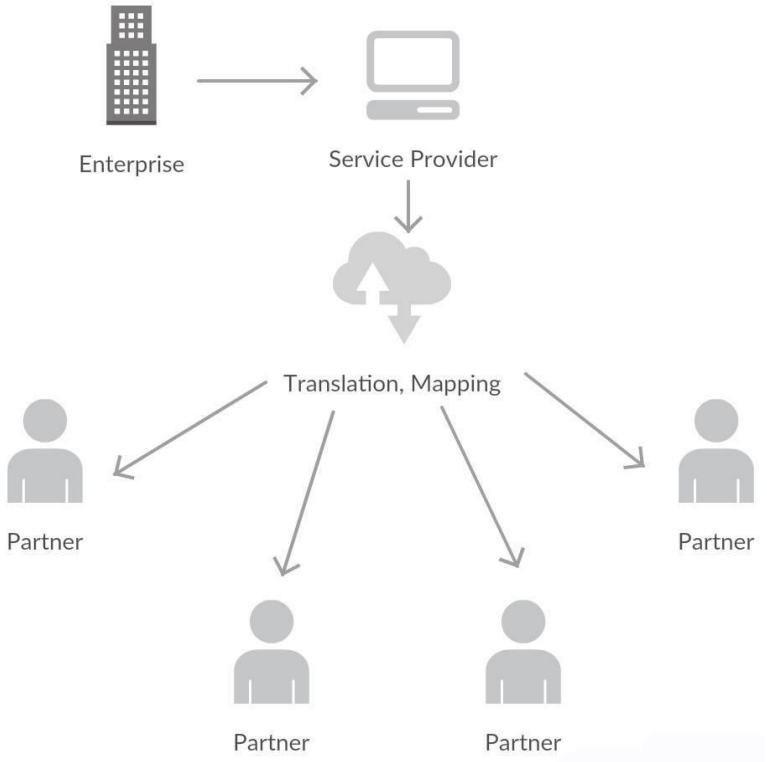


Fig. 3. Network E-business Model for Information Sharing among Public Organizations

• Hybrid Model

This model is a combination of the direct and network models. Typically, public organization will connect directly through the internet to their partner organizations with whom highest volume of transactions is conducted, using one or two preferred protocol. Also for those partner organizations that require a protocol different than the ones used to connect directly. The figure below shows the hybrid model of B2B. (Fig. 4.) For large communities, this model is much more commonly used today.

• Managed Model

In this model, the organization has outsourced the entire B2B process to an external provider. Outsourcing brings for the reduction in resource requirements, expenses, and complexity. The Provider receives the public organization documents directly from the ERP or the document management system (for example SAP, Oracle, etc.) and then takes action for all the mapping, translation, technical support, data center operations and document tracking. Once the documents are ready for transfer to the partner organizations, the service provider transmits them either with a direct connection to the partner or through the network, according to the particular partner organization requirements.

The figure below represents the managed model. (Fig. 5.) Nowadays, enterprises are increasingly outsourcing their entire B2B process to avoid the purchase and management of complicated and expensive B2B mapping, translation, and communications software.

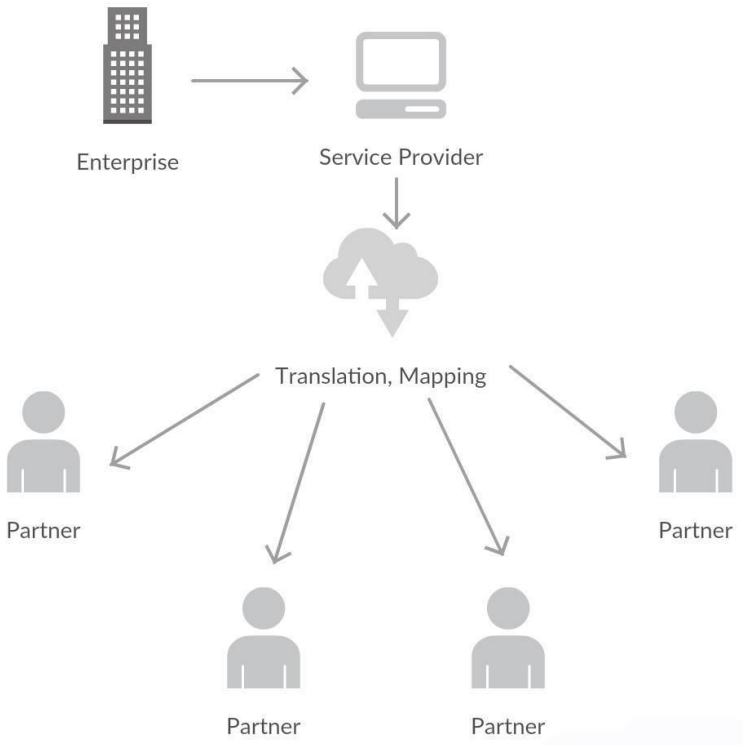


Fig. 4. B2B Hybrid E-business Model for information sharing among public organizations

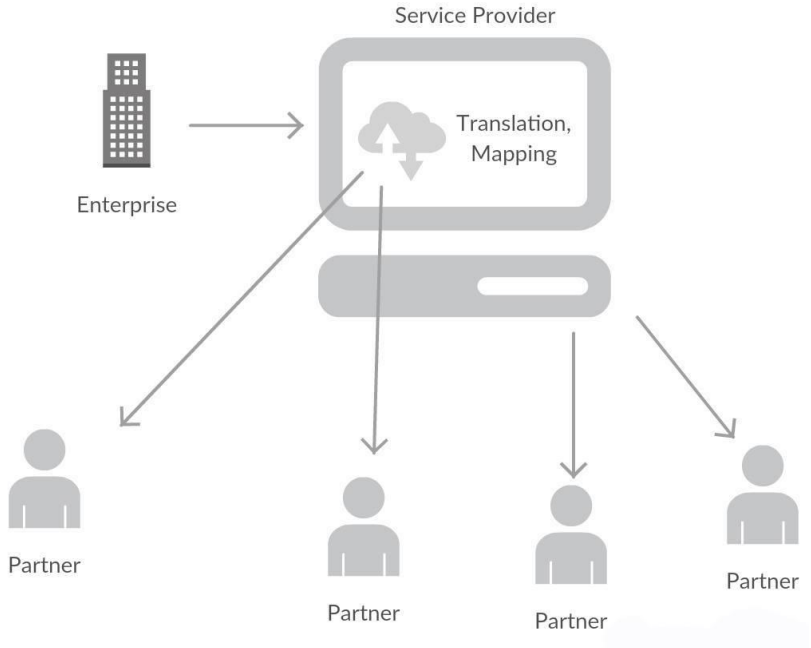


Fig. 5. B2B Managed E-business Model for information sharing among public organizations

Benefits of B2B E-business for the Supply Chain and Information Sharing Management

With the emergence of Cloud technologies and the following integration capabilities, a change in the B2B E-business can be noticed. A change that has a direct impact on the information sharing processes – enabling public organizations to connect easier, to maintain and manage relationships with their partner organizations at each point of the process. These solutions provide for cost savings, efficiency improvements and better user experience across the network.

Following are some of the benefits by applying E-business solutions in the information sharing processes:

- Automation of manual and repetitive processes

In public organizations with complex information channels, the manual processes part of B2B businesses can be automated, such as – information request approvals, updating request status, sending request shipping and tracking information. Automating these processes helps reduce order error rates, decrease order-to-cash time, provide quicker delivery to users, improve inventory management, enhance communication, and contribute to overall user satisfaction, loyalty and trust.

- Real-Time Analytics

The application of real-time analytics in B2B Ecommerce solutions creates a digital environment for public organizations to control the processes within the supply chain and information sharing respectively. For example, learn when and what the users are ordering, the frequency of requests from a particular user, predict the next type and amount of orders before they are placed, etc. Real-time analytics in sales has a direct impact on other processes involved in the supply chain – such as procurement and logistics.

- Order-to-Delivery Times Cutback

According to latest economic studies, Millennials now surpass Baby Boomers in workplace layout. Millennials don't always distinguish work life and non-work life; therefore, it is important for the security service to be running 24/7. With B2B E-business functionalities, users can place orders 24/7/365. This accessibility simplifies the information exchange across channels, and significantly reducing the order-to-delivery time.

- Fewer order errors

An online B2B Ecommerce platform can greatly decrease the errors that commonly occur during the order entry process. Orders captured through an online system eliminate many of the problems associated with manual order capturing. Besides, online ordering systems can enforce business rules for users as they are entering and editing orders, thus ensuring the entry of correct information.

As we can see combining the functionalities of B2B E-business with supply chain activities has significant benefits for the business – like improving operational efficiencies, reduce costs and improving user satisfaction across the supply chain. However, the solutions provided by this collaboration are covering only parts or specific modules of the supply chain. With so many possibilities coming from the Solution Providers, there is a need for an integrated and centralized platform where users, suppliers and logistics service providers can connect and handle various management activities associated with the supply chain processes. A network, which brings value to everyone, is dynamic, reflects the current technology trends and provides for the development and implementation of Industry 4.0 in manufacturing.

3. Supply Chain Network and Information Sharing Management

In the last decades, the technology development has stressed the importance of network collaboration. Especially, the rise of platforms such as Facebook and LinkedIn gave a significant push over the power to be connected over a network, from a communication and collaboration point of view. And by following the examples of these game-changing platforms, the larger the network grows, the more capabilities are added to its platform, which brings more value to its members.

Something similar could be noticed in the business world with the development of supply chain networks and information sharing network respectively. Information management has already become one of the central focus points in many organizational structures. With the global information environment growth, rising user expectations, rising costs and more intense and diverse competitions are boosting complexities in global information management. These trends pressure for developing new information sharing strategies and elaborate network designs.

Using Supply Chain Network Concept for Information Sharing among Cybersecurity Structures in Public Organizations

A Supply Chain Network (SCN) evolves from the core supply chain. Both supply chain and supply chain network represent the flow of goods (materials) and information, by connecting enterprises to serve and bring value to the end-user. While the concept of "chain" stands for a straightforward and subsequent set of links with a primary focus on response, "network" represents a more complex construction where companies are cross-linked with a two-way exchange flow between them [6].

The need for a supply chain and information sharing network development emerged due to the rapid technological enhancements. The latter pushed enterprises with basic supply chain model to develop into a more sophisticated structure, comprising a higher level of interdependence and connectivity between more parties involved in the production and distribution process [7]. The supply chain network represents the links between organizations and how are the information and goods flow between those links (Fig. 6.). In the model for Sharing Information among Cybersecurity Structures in the public sector, Security as a Service (SECaaS or SaaS) services can be defined as goods. These services are provided by SaaS providers to relevant organizations and users.

Organizations are realizing the significance of information sharing networks but are still not able to take a full advantage of its benefits. The challenge lies in selecting the right approach when designing an information sharing network.

4. Supply Chain Network and Information Sharing Network Design

A supply chain and information sharing network design for each business or public sector involves creating a framework that incorporates all elements such as organizational structures, information flows and infrastructure, technology, transportation and warehousing assets owned by the organization and those not owned by it, costs and external environment, etc.

The network design for each industry and business relates to certain information domain and varies, depending mainly on factors such as location and whether the activities are performed on a national, regional or global scale.

There are some main areas which should be considered as: determining the design processes of different activities; Defining the core elements of public sector organizations;

Other external factors that have a direct impact on the information sharing process such as government policies depending on countries, culture, etc.

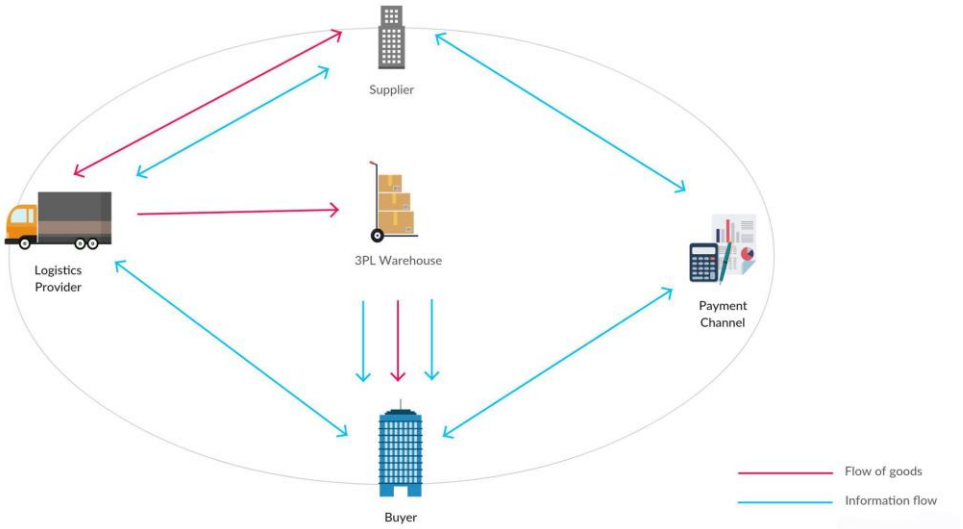


Fig. 6. Supply Chain Network Information and services / goods flow between organizations

Supply Chain Network and Information Sharing Network Capabilities

Analytics which is a predictive type analytics – to take a look at the broad ecosystem what is happening in the public sector and combine that with even more comprehensive unstructured data - things that could be potential reasons for cyberattacks like weather, like news, events, political and social factors etc. A combination of actionable public sector specific transactions data and the broader unstructured big data actually can drive useful insights for the organizations. All of these is managed through a business process dashboard. Not just a technical one but a business process dashboard that can tell if there is a cyberattack and why that cyberattack is just at that moment. That cyberattack could be exactly on that time because of a lot of reasons.

According to a study conducted by Supply Chain Insights [4], most companies think there is room for improvement in their supply chain processes because they are reactive, slow, traditional, inside-out, cautious. What companies want is to have agility, flexibility, and visibility. These are all process improvements which can be achieved through real community type approach, such as a network.

When it all comes together, it could help the companies and public sector organizations as well to finally build a real end-to-end information sharing processes. It would enable organizations to take the processes across the four walls of the organization, to their partners and users, to their suppliers and to their security service providers.

5. Conclusion

In conclusion it could be summarized that in the area of information sharing management, there is a need of more resources regarding guidelines, technologies, and methodologies for cybersecurity information sharing network development and its implementation in the public sector. To some extent, there is a lack of adequate data and information since the concept is relatively new and the need for it emerged with the rapid

technological and public sector-specific developments from the last decade which become a business-shaping phenomenon. Besides, there is no defined way to design a cybersecurity information sharing network as a network footprint. Moreover, the capability, capacity, and information flow - entwined and at the same time independent, are specific for different information domains. Therefore, there is no single optimal information sharing network design. However, there should be a cross-reference between responsiveness, risk tolerance, and efficiency.

In this sense of thoughts, as future challenges for research could be identified issues regarding define key information security technologies as evaluation criteria and weightings of the criteria about the critical information sharing processes, which involve multiple public sector organizations within the same context of connectivity and process design.

References

1. National Cyber Security Strategy "Cyber Resilient Bulgaria 2020", Sofia, 2016, (http://:www.strategy.bg).
2. Pavlov, G., Aleksandrova, V. (2003) "E-Tender - An Approach for Ensuring Transparency in Defence Budget Management", Information and Security, An International Journal, Vol. 11, 2003, ProCon Ltd., Sofia, p.148 – 162, ISSN 1311-1493.
3. Becker, A S 2007, Electronic Commerce: concepts, methodologies, tools, and applications, IGI Global.
4. Bolsinger, M 2015, 'Bringing value-based business process management to the operational process level', Springer-Verlag Berlin Heidelberg, InfSyst E-Bus Manage 13:355–398.
5. Capgemini Consulting, GT Nexus, Infor 2016, The Current and Future State of Digital Supply Chain Transformation.
6. Harland, C 1996, 'Supply network strategies. The case of health supplies', European Journal of Purchasing & Supply Management, vol. 2, no. 4, p. 183-192.
7. Harland, C, Lamming, R C & Zheng (eds) 2001, 'A Taxonomy of Supply Networks' The Journal of Supply Chain Management, Fall, 21-27.
8. Insurance Governance Leadership Network, 2015, Risk and opportunity in an increasingly digital world, viewed 10 May 2017, (http://www.ey.com/Publication/VW Luas sets/EY_-_Risk_and_opportunity_in_a_digital_world/$FILE/ey-risk-and-opportunity-in-an-increasingly-digital-world.pdf)
9. Miva, October 26, 2011, The History of E-commerce: How Did It All Begin?, viewed 30 June 2017,
10. Supply Chain Management Encyclopedia, 2011, Vendor managed inventory, viewed 19 June 2017, (http://scm.gsom.spbu.ru/index.php?title=Vendor_managed_inventory&oldid=3527)
11. Woollaston, V, Revealed, *What happens in just ONE minute on the internet: 216,000 photos posted, 278,000 Tweets and 1.8m Face book likes*, Daily Mail online, 30 July 2013.

Using Microsoft SQL Server 2019 Big Data Clusters

Mitko Radoev

University of National and World Economy, Sofia, Bulgaria
mradoev@unwe.bg

Abstract. The development of the modern database management systems is leading to providing users and developers of information systems with new and powerful tools for storing and processing large volumes of data. Microsoft SQL Server 2019 provides extremely interesting new features with Big Data Clusters. This paper aims to present the capacity of this tool and to suggest options for using it to store and process large volumes of heterogeneous data.

Keywords. Big Data, Microsoft SQL Server, Big Data Clusters, Machine Learning, Artificial Intelligence.

1. Introduction

Microsoft SQL Server is a contemporary relational database management system. The development of information technology in recent decades has led to a dramatic increase in the amount of stored and processed data, while also increasing the variety of data types. As a typical relational DBMS, Microsoft SQL Server is not designed to store Big Data, nor to store and process unstructured data, such as media files.

Microsoft SQL Server 2019 is expanding its data platform to cover big and unstructured data by integrating Apache Spark and HDFS into the Big Data Cluster [1].

Apache Spark is a platform for large-scale distributed data processing. Spark combines SQL, machine learning, graph computation, and stream processing. It provides high-level APIs in Java, Scala, Python and R, and an optimized engine that supports general execution graphs. It also supports a rich set of higher-level tools including Spark SQL for SQL and structured data processing, MLlib for machine learning, GraphX for graph processing, and Spark Streaming [2].

Hadoop Distributed File System (HDFS) is a highly fault-tolerant distributed file system designed to run on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets. HDFS was originally built as infrastructure for the Apache Nutch web search engine project [3].

Big Data Cluster uses a scalable storage layer that integrates SQL Server and HDFS to scale to petabytes of data storage. Integrated with SQL Server Spark enables the use of open source data processing libraries and large-scale processing and analyze high-volume data in a distributed, in-memory compute layer.

2. Big Data Clusters Architecture

The SQL Server Big Data Cluster is a group of Linux containers organized by Kubernetes. Kubernetes is responsible for the status of the SQL Server Big Data Clusters. Kubernetes builds and configures the cluster nodes, assigns pods to nodes, and monitors the state of the cluster.

The following figure shows the components of a big data cluster for SQL Server.

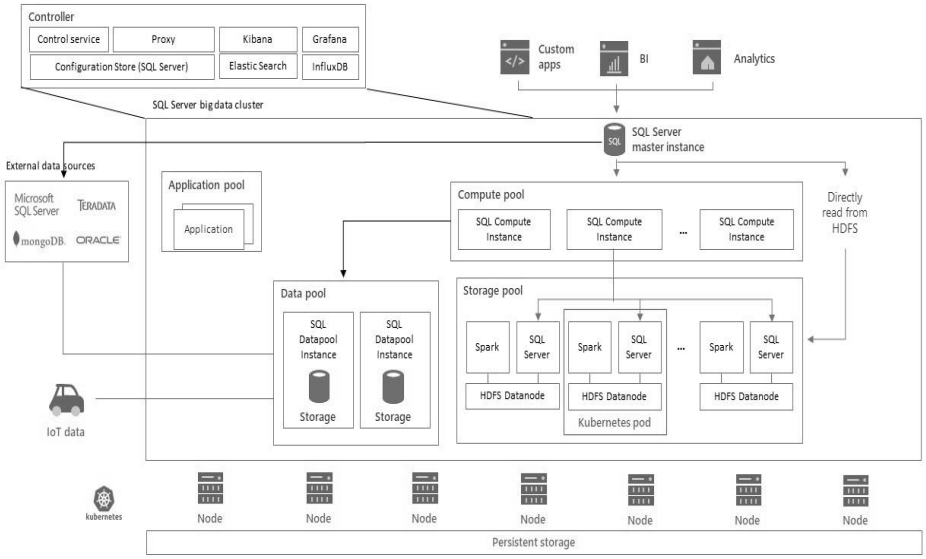


Fig. 1. The components of a big data cluster in SQL Server 2019 [4]

Controller

The controller is responsible for the management and security of the cluster. It includes the control service, the configuration store, and other services such as Kibana, Elastic Search, and InfluxDB.

Compute pool

The compute pool supplies computational resources to the cluster. The compute pool contains several SQL Server pods (group of one or more containers). They can use data from SQL Server or external sources – Oracle, HDFS, or others. The pods are divided into SQL Compute instances for different processing tasks.

Data pool

The data pool is a group of SQL Server engine pods that is used either to cache data from an external source or to store incoming streaming data. Data can be delivered from SQL queries or Spark jobs. In either case, the data is partitioned and distributed across the SQL Server instances in the pool.

Storage pool

The storage pool is a group of pods comprised of SQL Server engine, HDFS data node, and Spark containers. All the storage nodes in a SQL Server big data cluster are members of an HDFS cluster. This provides the scalable storage tier along with the collocated compute for SQL Server and Spark right next to the data.

3. Scenarios for using Big Data Clusters

SQL Server Big Data Clusters has capabilities to store and process big data. There are many options to use data from external data sources, store big data in HDFS, and query data from multiple different sources. Then data can be used for data analysis, machine learning, and artificial intelligence. The following section provide more information about some of these scenarios.

Querying data from different sources

By using PolyBase, SQL Server Big Data Clusters can query external data sources without moving or copying data. PolyBase was introduced in SQL Server 2016 and it was used to process Transact-SQL queries that read data from external data sources. In SQL Server 2019 Polybase was enhanced to provide different data source connectors including Oracle, SQL Server, MongoDB (CosmosDB), and Teradata. These data sources can be accessed without installing any other software.

We can write T-SQL query to create external data source using CREATE EXTERNAL DATA SOURCE command:

```
CREATE EXTERNAL DATA SOURCE <data_source_name>
WITH
 ( [ LOCATION = '<prefix>://<path>[:<port>]' ]
   [ [ , ] CONNECTION_OPTIONS = '<name_value_pairs>']
   [ [ , ] CREDENTIAL = <credential_name> ]
   [ [ , ] PUSHDOWN = { ON | OFF } ]
   [ [ , ] TYPE = { HADOOP | BLOB_STORAGE } ]
   [ [ , ] RESOURCE_MANAGER_LOCATION = '<resource_manager>[:<port>]' )
```

There are seven different prefixes and several different ways to specify the path for each supported external data source.

Table 1**.** Prefixes and paths for external data sources

Data Source	Prefix	Path
Cloudera or Hortonworks	hdfs	<Namenode>[:port]
Azure Blob Storage	wasb[s]	<container>@<storage_account>.blob.core.windows.net
SQL Server	sqlserver	<server_name>[\<instance_name>][:port]
Oracle	oracle	<server_name>[:port]
Teradata	teradata	<server_name>[:port]
MongoDB or CosmosDB	mongodb	<server_name>[:port]
ODBC	odbc	<server_name>[:port]

The next step is to create the necessary external file formats, defining external data, by command CREATE EXTERNAL FILE FORMAT.

Then, in the external data source we can create external tables using CREATE EXTERNAL TABLE command:

```
CREATE EXTERNAL TABLE <table_name>
   ( <column_definition> [ ,...n ] )
   WITH (
      LOCATION = 'folder_or_filepath',
      DATA_SOURCE = external_data_source_name,
      FILE_FORMAT = external_file_format_name
      [ , <reject_options> [ ,...n ] ]
   )
```

There are the two main methods to ingest data into the data pool of a SQL Server 2019 Big Data Clusters:

- Using Transact-SQL. We can load data with Transact-SQL command INSERT;
- Using Spark jobs. We can create a Spark streaming job that loads data from the storage pool (HDFS) into the external table you created in the data pool.

Finally, we can query data from created external tables into data pool along with data from local tables. This option can be useful when we need to integrate data from different systems, for example to integrate data from heterogeneous systems [5], to combine dynamic data with data from relational database [6], and in many other cases [7], [8].

The big data cluster includes also a scalable HDFS storage pool. This storage pool can be used to store big data, potentially ingested from multiple external sources. When the big data is stored in this pool, there are many options to query or analyze the big data along with relational data from SQL Server database.

Artificial Intelligence and Machine Learning

SQL Server 2019 Big Data Clusters provides a comprehensive machine learning and artificial intelligence platform with all the tools and services required to ingest, store, prepare, and analyze data [9].

Big Data Clusters enable artificial intelligence and machine learning tasks on the data stored in HDFS storage pools and the data pools. We can use Spark as well as built-in artificial intelligence tools in SQL Server, using R, Python, Scala, or Java.

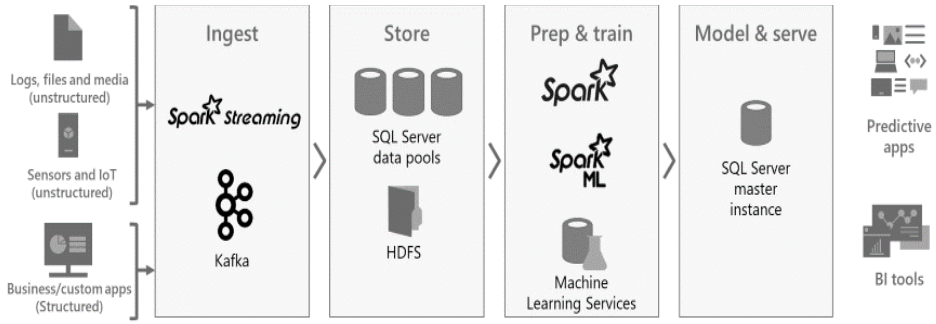


Fig. 2. A complete platform for artificial intelligence and machine learning [4]

SQL Server 2019 Big Data Clusters provide all the tools to prepare data for analysis and to train the machine learning models. We can train a machine learning model using Spark, SparkML, and SQL Server Machine Learning Services. Then we can expose the model as a

machine learning application using SQL Server ML with Transact-SQL or as an application with a REST interface in the application pool [10].

There are the two main group of tools for machine learning in SQL Server 2019 Big Data Clusters:

- SQL Server ML Services;
- Spark and Spark ML.

We can use SQL Server Machine Learning Services to prepare data, do feature engineering, train and deploy machine learning models inside the database. Machine Learning Services is a feature in SQL Server that gives the capability to execute Python and R scripts with data from relational database. We can use the Microsoft Python and R packages for predictive analytics and machine learning. The scripts are executed where the data resides, without moving data outside SQL Server. In addition to the Microsoft packages revoscalepy and microsoftml for Python, and RevoScaleR, MicrosoftML, olapR, and sqlrutils for R, which are included in Machine Learning Services, we can use also the open-source packages and frameworks, such as PyTorch, TensorFlow, and scikit-learn.

MLlib is Spark's machine learning library. At a high level, it provides the following features:

- ML Algorithms - classification, regression, clustering, and collaborative filtering;
- Featurization - feature extraction, transformation, dimensionality reduction, and selection;
- Pipelines - constructing, evaluating, and tuning ML Pipelines;
- Persistence - saving and load algorithms, models, and Pipelines;
- Utilities - linear algebra, statistics, data handling, etc.

Spark ML is a new package introduced in Spark 1.2 that aims to provide a uniform set of high-level APIs that help users create and set up machine learning pipelines.

Data scientists can choose either to use SQL Server Machine Learning Services in the master instance to run R, Python, or Java model training scripts or to use batch scoring over the data in HDFS with Spark.

The important thing is that the entire sequence of actions takes place in the environment of the SQL Server Big Data Cluster. There is no need to transfer the data to an external machine learning server or anywhere outside the big data cluster. The available power of the hardware underlying the big data cluster is accessible for data processing and computing resources can be scaled up and down according to needs.

4. Conclusion

SQL Server 2019 big data clusters offers interesting new possibilities to use SQL Server to bring the highest value enterprise data from a relational database and the high-volume big data together on one common, scalable data platform. Companies can use the power of PolyBase to virtualize their data stores, create data lakes, and create scalable data marts in a secure environment without needing to implement ETL tools. This makes data-driven applications and analysis much more efficient and productive.

References

1. Microsoft Corporation, "Microsoft SQL Server 2019 Big Data Clusters Technical white paper," 9 2018. [Online]. Available: https://info.microsoft.com/ww-landing-SQLDB-Microsoft-SQL-Server-WhitePaper.html.

2. Apache Software Foundation, "Apache Spark 2.4.5," 2018. [Online]. Available: https://spark.apache.org/docs/latest/.
3. Apache Software Foundation, "HDFS Architecture Guide," 22 8 2019. [Online]. Available: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html.
4. Microsoft Corporation, "What are SQL Server Big Data Clusters?," 2019. [Online]. Available: https://docs.microsoft.com/en-us/sql/big-data-cluster/big-data-cluster-overview?view=sql-server-ver15.
5. Tsaneva M., "A Practical Approach For Integrating Heterogeneous Systems," Business management, no. 2, p. 11, 2019.
6. Milev P., "Technological Issues of Storing Dynamic Data in a Relational Database on Research Projects," Trakia Journal of Sciences, vol. 13, pp. 22-25, 2015.
7. Lazarova V., "The Expansion of the E-business Information Systems in the Cloud- Problems and Solutions," Ikonomiceski i Sotsialni Alternativi, no. 4, pp. 34-47, December 2015.
8. Mihova V., "Common Architecture Design of a Business Information System for Performance Management of the Business Applications," in 3rd International conference on application of information and communication technology and statistics in economy and education ICAICTSEE–2013, Sofia, Bulgaria, 2013.
9. Blendea M., "Build an intelligent analytics platform with SQL Server 2019 Big Data Clusters," Microsoft, 2019. [Online]. Available: https://cloudblogs.microsoft.com/sqlserver/2019/11/11/build-an-intelligent-analytics-platform-with-sql-server-2019-big-data-clusters/.
10. Ward B., SQL Server 2019 Revealed Including Big Data Clusters and Machine Learning, North Richland Hills, Texas: Apress, 2019.

General Data-driven File Export Integration Solution

Monika Tsaneva

University of National and World Economy, Sofia, Bulgaria
mtzaneva@unwe.bg

Abstract. The paper presents some options for development of a data export integration software module that can be plugged in different types of business applications. Business and Technological Requirements to a File Export Integration Solution are discussed. Architecture of a Data-driven File Export Integration Module that comprises of universal data base structures and stored procedures is proposed and developed. Promising directions for future optimizations of this Integration solution are outlined.

Keywords. Data-driven software, Integration solution, File export, Universal Database Structures, Heterogeneous System Integration

1. Introduction

Recipients of information within an organization can be different and their needs can vary. This influences the way, the time, the format and the representation of similar core data delivered to different addressees [6]. In core business and customer support, data needs to be processed in a real time manner, thus the organization will be more flexible, it will deliver more value to customers and new business opportunities will be enabled. On the other hand, there is also data that is not needed in a real time or near real time fashion. Such information can be provided after collecting and analyzing of large volumes of data, that can be transferred to some reporting system in the so-called batch window, when the core business is not fully loaded and has free resource for synchronizing. This is the common scenario for ETL process into the company's data warehouse [13] and for data exports for National or European institutions required by low. In integration frameworks based on batch processing, a large group of transactions are collected and then the whole amount of data is processed by a software agent during a single execution. Data integration frameworks are better applied in business scopes with batch-oriented problems, because of their lower network and processing overload and better scalability for large data volumes [5, 6]. In data integration approaches, the integration adapter or interfaces may be implemented in two different manners - to wrap directly the data repository or data-access layer of both interconnected systems [2, 3], or to be encapsulated in the data repository or data-access layer of the system that should provide data for one or more other (including third party) systems. This last approach by-passes the business logic layer and the business rules are not involved in the processing, so it is usually applied when the main aim of the integration is synchronizing or replicating data between systems without direct access to each other's data.

2. Business and Technological Requirements to a File Export Integration Solution

The main purpose of an integration solution for exporting data to files is to synchronize or to replicate data between third-party systems without direct access to each other's data repositories. The choice of an architecture for such an integration module is mainly determined

by two groups of factors – business and technological.

By abstracting the semantics of data, the most important business requirement refers to the file formats that must be provided by individual export procedures. This requirement may be imposed by some national regulatory authorities like National Credit Register (NCR) [1], Nacional Revenue Agency (NRA) [15], National Security Institute (NSI), Bulgarian National Bank (BNB) etc. In other cases, the file format may be chosen based on some internal design considerations like estimated volume of data, future data usage, effort, time and cost of implementation, frequency of data export (periodic or as needed) etc. Among the most commonly used file formats are the following (see 0):

•Fixed width / Delimited	•Larde data volumes
	•Hard to read by humans
•Tag based (xml, json)	•Small data volumes
	•Man and Machine Readable
	•Selfdescriptive
Application specific	On explicit demand only
	Online "save as" operations

Fig. 1. Commonly used file formats

Fixed width file format means that the content of each field starts at a predefined position and has a constant length within which it can be left or right aligned and depending on its type, may be padded with some character (usually zero for numbers and space for strings). All the fields in delimited file formats are separated by a predefined character (separator like tab, comma, semicolon etc.), and they have predefined order and type, but not length thus usually values are not padded. Both fixed width and delimited file formats are suitable for large data volumes export (delimited files are smaller), they may or may not begin with a header row, but in both cases, they are difficult to read by humans. Tag based file formats, are wasteful in size (especially XML, because of duplicated tags) but are both man and machine readable. These characteristics make them suitable for exports of small amount of data with a use that is not predefined, e.g. creating a copy of customers personal data required by GDPR. Basically, application specific file formats (e.g. doc, docx, xls, xlsx, rtf etc.) are not used for general purpose data export, but when some additional processing on reports content like reformatting, restructuring or supplementary calculations and/or transformations must be performed. In the most common scenario data exports in application specific formats is implemented as standard functionality of system reports and is outside of the aim and the scope of this paper.

The technological requirements for the file export module aim to ensure the smooth running and future development of the system:

- Flexibility – easy change of file formats and file structures when business need or legal demand occurs
- Extendibility – new data exports to be easily implemented which is based on high degree of reusability of data structures, software components and metadata that have already been implemented as a part of file export module
- Scalability – the capability of a system to handle a growing amount of work and volume of information by adding resources to the system.

Flexibility and extendibility of the solution may be achieved by designing and developing abstract metadata that describe properties of each individual export, universal data structures to store this metadata and database procedures for necessary information retrieval and file creation. Such a solution is relatively easy to implement in a traditional database, and scalability may be achieved by providing for the possibility of future migration to the cloud [7], both of its constitutive software components and of the files [4, 8], obtained as a result of their work.

3. Architecture of a Data-driven File Export Integration Module

Modern DBMS provide a wide range of tools for creating stored procedures including embedding different languages and programming technologies and tools to provide existing stored procedures as services. Based on the fact that this procedural element of a database is good enough to implement the complete necessary functionality and can easily be made available for use, it can be considered as one of the easiest and cheapest possible options.

The proposed approach includes two main components - database stored procedures to perform the export (0) and abstract data structures implemented as a scheme within the same database (0), which describe the parameters of each individual export and the structure of the resulting file, define the queries to extract the exported data and the processing flow control.

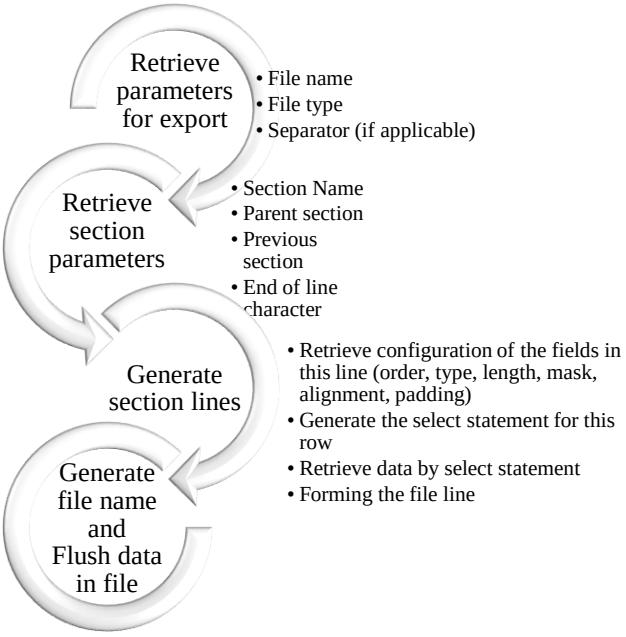


Fig. 2. Procedures implementing file export

The first step of the processing is to retrieve the parameters for export, that needs to be performed. Each individual export is described with a single record of EXP_CONFIG table. It is identified by a unique key (EXP_ID) and has a user-friendly name, that is used for display and logging purposes. The other columns of EXP_CONFIG table are used to set the name and the type of the file which will be generated and the separator (if any) which must be places between each two data fields. If the separator is not set for a particular export, a file with fixed-width records will be produced.

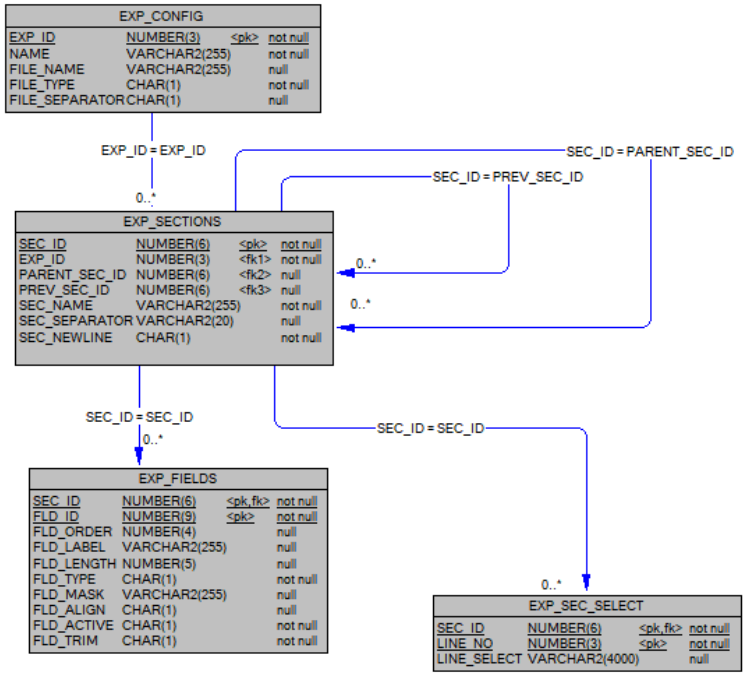


Fig. 3. Physical Model of a Universal Data Base scheme for file export

As a second step of the processing, first section (or single section in the simplest possible case) parameters are retrieved. In the context of this development, the term "section" is used to distinguish a set of lines (rows) with one and the same structure. Basically, each file comprises of one or more sections and even complicated master-detail relations between records in a single file may be described. Each section has a unique identifier (SEC_ID) and a name (SEC_NAME) used to facilitate development, tracking and logging. First section of a file must not have either a parent (PARENT_SEC_ID), or a previous section (PREV_SEC_ID). All other sections must have a previous section to ensure the correct order of records in the resulting file. If needed, any section, except the first one, may be a part of some upper level structure thus belonging to a container with all other sections with the same master (parent) section. Each section may start with some specific separator if such has been defined by SEC_SEPARATOR field value or may start on a new line (depending on the value of SEC_NEWLINE field).

The third step of the export procedure deals with individual rows of the section being currently processed. Each individual line is formed as follows:

1. The configuration metadata for the fields of this line is retrieved. Configuration metadata is stored in EXP_FIELDS table, and it describes each field that must be placed in this section lines. There are some minor differences in the interpretation of meta data depending on the file type – fixed width or delimited. The fields of fixed width files are placed as set by its order (FLD_ORDER column) and the sum of length of all previous fields (their FLD_LENGTH properties), while the length attribute does not affect the position of delimited file fields. Values exported are formatted depending on the setting for their type (FLD_TYPE), mask (FLD_MASK), alignment (FLD_ALIGN) and trimming (FLD_TRIM). Data types currently supported are strings, numbers, dates and datetimes. The specific format for dates,

datetimes and numbers may be set by field's mask (e.g. 'DD.MM.YYYY' or 'dd-mm-yy', '#09.99' etc.), and string values may be trimmed from the left, right or both sides. If set, the alignment property (left or right) is applied in case the length of a fixed width field exceeds the actual data length, otherwise, by default, numeric values are right-aligned, and all the other are left-aligned. Some fields may be temporary excluded from the record by setting their "is this field currently active" flag (FLD_ACTIVE) to false.

2. The select statement for row content retrieval is generated. Select statements that are used to retrieve data for export stored in EXP_SEC_SELECT table. They are generated by combining the ordered lines (LINE_NO) containing clauses and parameter placeholders (LINE_SELECT) which need to be replaced by actual arguments. Partial formatting aimed to ensure compliance with the settings defined as metadata (p. 1) is embedded directly into some "select" statements.

3. The values to be exported are retrieved via batch execution (e.g. using Oracle EXECUTE IMMEDIATE statement) of previously generated select statements (p. 2). All the results of these run time queries are temporarily stored in bulk collect memory structures.

4. File lines comprised of fields containing the values previously retrieved (p. 33) and additionally formatted to comply the settings from metadata (p. 1) are composed.

As a final step a file name based on the configuration data retrieved on the first step (EXP_CONFIG table) is generated and the data lines that were stored as bulk collect memory structures are flushed into this file. In current implementation all files are stored in system folders on the database server but using a binary object as an option for intermediate persistence is also possible.

In order to avoid any possible performance issues, some suitable optimized views may be created in advance, so that queries for export do not include sophisticated and error prone syntax. Such kind of views provide great options for continuous optimisations and rapid business logic adjustments, so they are commonly used by the suggested integration approach implementation, but they are not defined as a part of it because they cannot be abstract but depend on the semantics of the exported data. Additional security and maintainability is achieved by adding some logging procedures with corresponding data structures, but these also are not an intrinsic part of this solution, because of their general purpose design and usage by other subsystems.

4. Conclusion

The proposed integration approach offers several advantages to businesses: it is a general-purpose solution that does not depend on the business process and it's data structures specifics; the solution does not require substantial financial or time resources since it employs the existing IT infrastructure. The designed file-export module is highly efficient, flexible, and scalable. It has been implemented for more than five years now without any substantial delays in the operational system or the business process being registered so far. Based on these results same general idea using database procedures and abstract data structures will be used to design and implement universal integration solution for file import, and this way bidirectional batch integration with third party systems will be ensured. Migrating the universal database scheme to a cloud data base may be planned for some point in the future.

References

1. Bulgarian National Bank, Guidelines for Submitting and Receiving Information from the Information System of the Central Credit Register (BG), 2019, (https://www.bnb.bg/bnbweb/groups/public/documents/bnb_law/guidelines_ccr_part_new_bg.pdf)

2. Denchev, E. Problems and solutions for information provision of clusters by small and medium enterprises (SMEs), International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2018), (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2018.pdf).,
3. Denchev, E. Building a model for consolidating information in clusters of small and medium-sized enterprises (SMEs), International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2017), (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2017.pdf).
4. Denchev, E. Swot analysis of cloud backup, International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2013), (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2013.pdf).
5. Kouzmanov, S., Analysis and Classification of Business Software Systems Integration Approaches, Economic Alternatives, 2015, 2, (https://www.unwe.bg/eajournal/en/journalissues/article/7770)
6. Kouzmanov S. , M. Tsaneva, Impact Of Business Problem Characteristics On The Architecture And Specification Of Integration Framework, 3-rd International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2013), (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2013.pdf).,
7. Lazarova, V., Managing User Access to Cloud Services by Company Administrators. 2016 TEM Journal, 5(3), p.289-293
8. Lazarova, V., Principles of Transactional Approach in the Classical Web-based Systems and the Cloud Computing Systems - Comparative Analysis, Economic Alternatives, 1/2013, (https://www.unwe.bg/uploads/Alternatives/Article9_01.2013.pdf)
9. Lazarova, V., The Expansion of the E-business Information Systems in the Cloud-Problems and Solutions Ikonomiceski i Sotsialni Alternativi, 4/2015, (http://www.unwe.bg/uploads/Alternatives/3-br4-2015.pdf (application/pdf))
10. Marzovanova M., Building Multi-Touch User Interface, 4TH International Conference On Application Of Information And Communication Technology And Statistics In Economy And Education (ICAICTSEE-2014), (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2014.pdf).
11. Marzovanova, M., Creating Annotations and Indexes in System for Intelligent Tagging and Search in Unstructured Data, International Conference on Application of Information and Communication Technology and Statistics and Economy and Education (ICAICTSEE-2013), (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2013.pdf).
12. Mihova V., Murdjeva A. Metadata for generating a specific data warehouse. International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2012), Sofia, Bulgaria, 2012. (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2012.pdf)
13. Mihova, V., Stefanov, G., Marzovanova, M., Cognos Mobile -dashboards design and implementation technology, International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2016), (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2016.pdf).
14. Radoev M., A Comparison Between Characteristics of NoSQL Databases and Traditional Databases, Computer Science and Information Technology, vol. 5(5), 2017, San Jose, CA, USA, (http://www.hrpub.org/download/20171130/CSIT1-13510351.pdf).
15. Radoev M., Specific Requirements for the Accounting Software in Budget Organizations, Ikonomiceski i Sotsialni Alternativi, University of National and World Economy, Sofia, Bulgaria, 2016, issue 3, pages 93-98.
16. Tsaneva, M., A practical approach for integrating heterogeneous systems. Business Management 2/2019 (http://bm.uni-svishtov.bg/title.asp?title=1383).
17. Tzaneva, M., Kouzmanov, S., Impact of Business Problem Characteristics on the Architecture and Specification of Integration Framework, International Conference on Application of Information and Communication Technology and Statistics in Economy and Education. Sofia (ICAICTSEE-2013),. 2013, pp. 628-635, (http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2013.pdf).

Security Certificates in Public Web Sites of Banks from Balkan States

Pavel Petrov, Petar Dimitrov

University of Economics - Varna, Varna, Bulgaria
petrov@ue-varna.bg, p.d.dimitrov@ue-varna.bg

Abstract. In the recent years a trend is formed to use the HTTPS protocol as the default protocol for accessing web pages and to be used by default by web applications. In order for this to be done, a valid certificate issued by authority body should be used. In the scope of the study in the spring of 2019 we examined the web sites of 20 banks licensed in Bulgaria, 24 banks licensed in Romania and 27 banks licensed in Serbia. The survey excludes the foreign bank branches, because we try to outline the "good practices" used by domestic administrators of banking websites.

Keywords. Security Certificates; Bulgarian banks; Romanian banks; Serbian banks; HTTPS; bank web site.

1. Introduction

In this comparative study we choose banks which are regulated by local central banks and are from three neighboring European countries, located on Balkan Peninsula - Bulgaria, Romania and Serbia. These countries have a lot of similarities in demographics and economics characteristics. However, from political point of view, they are not similar. Bulgaria and Romania are members of EU and NATO, while Serbia is not. From financial point of view, the countries are not using Euro as currency and their bank systems are formally independent from higher level supervision. In Table 1 are summarized some large-scale similarities which have meaning in the context of the current study.

Table 1. Some large-scale similarities between Bulgaria, Romania and Serbia

Feature	Bulgaria	Romania	Serbia
Population (circa 2019)	7.000 mil.	19.524 mil.	7.021 mil
GNI per capita (Atlas method) in 2017	7 860$	10 000$	5 180$
Monetary policy	Not members of Eurozone		
Currency	BGN	RON	RSD
Political orientation	EU, NATO		-
Credit ratings	Long term Rating, Foreign currency		
S&P Global Ratings	BBB- (2018)	BBB- (2014)	BB (2018)
Moody's Investors Service	Baa2 (2017)	Baa3 (2018)	Ba3 (2017)
Fitch Ratings	BBB (2019)	BBB- (2018)	BB (2018)

The governments have high long-term credit ratings from credit rating agencies such as Standard & Poor's, Moody's, Fitch Ratings and so on. It should be noted that Gross National Income (GNI) represents the total domestic and foreign output of an economy. The GNI per

capita, developed by the World Bank, is widely used as an indicator for the overall level of economic development.

Before the comparison, we raise the hypothesis that in such very similar countries, with similar level of living standard and penetration of information and communication technologies, it should be expected that the web technologies used in local banks' web sites should also be similar. In these institutions there is no funding issues, and there are even opportunities to use expensive proprietary software. We believe that web technologies are a key component in the banking IT infrastructure. These technologies become an integral part of every aspect of the online banking business, therefore security and performance cannot be compromised for any reason.

2. Methodology and Computational Details

In our study home pages of 20 Bulgarian, 24 Romanian and 27 Serbian banks were inspected in April 2019. The main method used in the survey includes analysis of the responses given by the web servers. Google Chrome ver.73, working under typical desktop PC - Windows 10 Professional Edition x64, was used as a web client with "Developer tools" module activated. It should be noted that this module has not been intended specifically for such kind of studies, but it is a very useful tool in such cases. The process of inspection was done manually by expert estimation. Other approaches to do the same research could include using command line tools such as "curl". In our opinion, using real web browser is more straightforward. The methodology of the study is based partially on methodology used in previous studies on web technologies used in banks, and other studies [6], [7], [8], [9].

Table 2. Usage of the HTTPS protocol in public web sites of banks in Bulgaria

№	Bank Name	Bank' domain	HTTPS
1	Allianz Bank Bulgaria AD	allianz.bg	yes
2	Bulgarian Development Bank AD	bdbank.bg, bbr.bg	problem
3	Bulgarian-American Credit Bank AD	bacb.bg	yes
4	Central Cooperative Bank AD	ccbank.bg	yes
5	D Commerce Bank AD	dbank.bg	yes
6	DSK Bank EAD	dskbank.bg	yes
7	Eurobank Bulgaria AD	postbank.bg	yes
8	First Investment Bank AD	fibank.bg	yes
9	International Asset Bank AD	iabank.bg	**NO**
10	Investbank AD	ibank.bg	yes
11	Municipal Bank AD	municipalbank.bg	yes
12	Piraeus Bank Bulgaria AD	piraeusbank.bg	yes
13	ProCredit Bank (Bulgaria) EAD	procreditbank.bg	yes
14	Raiffeisenbank (Bulgaria) EAD	rbb.bg	yes
15	Societe Generale Expressbank AD	sgeb.bg, expressbank.bg	yes
16	TBI Bank EAD	tbibank.bg	yes
17	Texim Bank AD	teximbank.bg	yes
18	Tokuda Bank AD	tcebank.com, tokudabank.bg	yes

19	UniCredit Bulbank AD	bulbank.bg, unicreditbulbank.bg	yes
20	United Bulgarian Bank AD	www.ubb.bg	yes

Table 3. Usage of the HTTPS protocol in public web sites of banks in Romania

№	Bank Name	Bank' domain	HTTPS
1	Alpha Bank Romania S.A.	alphabank.ro	yes
2	Banca Comerciala FEROVIARA S.A.	bfer.ro	problem
3	Banca Comerciala Intesa Sanpaolo Romania S.A.	intesasanpaolobank.ro	yes
4	Banca Comerciala Romana S.A.	bcr.ro	yes
5	Banca de Export Import a Romaniei EXIMBANK S.A.	eximbank.ro	yes
6	Banca Romana de Credite si Investitii SA	brci.ro	yes
7	Banca Romaneasca S.A. Membra a Grupului National Bank of Greece	brom.ro, banca-romaneasca.ro	yes
8	Banca Transilvania S.A.	bancatransilvania.ro	yes
9	Bank Leumi Romania S.A.	leumi.ro	yes
10	BRD - Groupe Societe Generale S.A.	brd.ro	yes
11	CEC Bank S.A.	cec.ro	yes
12	Credit Agricole Bank Romania S.A.	credit-agricole.ro	yes
13	Credit Europe Bank (ROMANIA) S.A.	crediteurope.ro	yes
14	First Bank S.A.	firstbank.ro	yes
15	Garanti Bank S.A.	garantibank.ro	yes
16	Idea Bank S.A.	idea-bank.ro	yes
17	Libra Internet Bank S.A.	librabank.ro	yes
18	Marfin Bank (ROMANIA) S.A.	marfinbank.ro	yes
19	OTP Bank Romania S.A.	otpbank.ro	yes
20	Patria Bank S.A.	patriabank.ro	yes
21	Porsche Bank Romania S.A.	porschebank.ro	yes
22	ProCredit Bank S.A.	procreditbank.ro	yes
23	Raiffeisen Bank SA	raiffeisen.ro	yes
24	UniCredit Bank S.A.	unicredit.ro	yes

The lists of banks authorized to operate in Bulgaria, Romania and Serbia were taken from the websites of Bulgarian National Bank [10] (Table 2), Romanian National Bank [11] (Table 3) and National Bank of Serbia [12] (Table 4). In our study websites of foreign bank branches and representative offices of foreign banks operating in Bulgaria, Romania and Serbia are excluded. We surveyed only local ones, which operate under supervision of the domestic national bank. So, the websites of those banks that operate on a branch or on a cross-border basis were omitted.

The summarized results of the studied home web pages are presented in several tables (Tables 5 - 7) based on the following key indicators: presence of automatic redirection to HTTPS, certificate type, certification body and validity period of the SSL certificate.

Table 4. Usage of the HTTPS protocol in public web sites of banks in Serbia

№	Bank Name	Bank' domain	HTTPS
1	Addiko Bank AD Beograd	addiko.rs	yes
2	Agroindustrijsko Komercijalna Banka AD, Beograd	aikbanka.rs	yes
3	Banca Intesa AD Beograd (Novi Beograd)	bancaintesa.rs	yes
4	Banka Postanska Stedionica AD, Beograd (Palilula)	posted.co.rs	yes
5	Credit Agricole Banka Srbija AD Novi Sad	creditagricole.rs	yes
6	Direktna Banka AD Kragujevac	direktnabanka.rs	yes
7	Erste Bank AD, Novi Sad	erstebank.rs	yes
8	Eurobank AD Beograd	eurobank.rs	yes
9	Expobank AD Beograd	expobank.rs	problem
10	Halkbank AD Beograd	halkbank.rs	**NO**
11	Jubmes Banka AD Beograd (Novi Beograd)	jubmes.rs	**NO**
12	Komercijalna Banka AD, Beograd (Vracar)	kombank.com	yes
13	Mirabank AD Beograd-Novi Beograd	mirabankserbia.com	yes
14	MTS Banka AD Beograd	mts-banka.rs	yes
15	NLB Banka AD, Beograd	nlb.rs	yes
16	Opportunity Banka AD, Novi Sad	obs.rs	yes
17	OTP Banka Srbija AD, Novi Sad	otpbanka.rs	yes
18	Piraeus Bank AD Beograd (Novi Beograd)	piraeusbank.rs, direktnabanka.rs	yes
19	Procredit Bank AD, Beograd (Novi Beograd)	procreditbank.rs	yes
20	Raiffeisen Banka AD Beograd	raiffeisenbank.rs	yes
21	Sberbank Srbija A.D. Beograd	sberbank.rs	yes
22	Societe Generale Banka Srbija AD, Beograd	societegenerale.rs	yes
23	Srpska Banka AD Beograd (Savski Venac)	srpskabanka.rs	yes
24	Telenor Banka AD Beograd (Novi Beograd)	telenorbanka.rs	yes
25	Unicredit Bank Srbija A.D., Beograd (Stari Grad)	unicreditbank.rs	yes
26	Vojvodanska Banka AD Novi Sad	voban.rs	yes
27	VTB Banka AD Beograd	vtbbanka.rs, apibank.rs	yes

3. Empirical Results and Discussion

Around 5% of the Bulgarian and 7.4% of the Serbian bank web sites surveyed are not using HTTPS at all. In our opinion this is not a high percentage, but it is strange, because the prices for a simple DV certificate starts at around 30€ annual fee. Also, there is a free alternative. For example, in the recent years, major organizations and companies, such as the Electronic Frontier Foundation, Mozilla, Akamai, Cisco, IdenTrust, and others, have

collaboratively set up a certifying authority, Let's Encrypt, to issue free certificates. These certificates are currently valid for 90 days. Since the beginning of 2018, the so-called "wildcard certificates" covering all subdomains of a domain was introduced.

One Bulgarian, one Romanian and one Serbian bank web sites are using HTTPS with issues. Bulgarian bank №2 (from Table 2) is using expired in 2012 self-signed certificate for "localhost.localdomain". Romanian bank №2 (from Table 3) is using certificate which hostname does not match the main domain but only the subdomain bcfonline.bfer.ro. Serbian bank №9 (from Table 4) is using self-signed certificate for "marfin.gridsrv.net" with validity period of 10 years - until 2027. In these cases, we can give the following recommendation: either to support HTTPS according to good practices or better not to use HTTPS at all, because these problems could weaken the confidence of customers in the financial institution's ability to keep its systems up to date.

Table 5. Main features in usage of the HTTPS in public web sites of Bulgarian banks

Bank №	Automatic redirection to HTTPS	Certificate type	Certification authority body	Validity in
1	Yes	**DV**	RapidSSL RSA CA 2018	9
3	Yes	EV	COMODO RSA Extended Validation Secure Server CA	26
4	Yes	EV	COMODO RSA Extended Validation Secure Server CA	24
5	Yes	EV	GeoTrust EV RSA CA 2018	14
6	Yes	EV	DigiCert SHA2 Extended Validation Server CA	13
7	Yes	EV	GeoTrust EV RSA CA 2018	24
8	Yes	EV	COMODO RSA Extended Validation Secure Server CA	26
10	**NO**	**DV**	COMODO ECC Domain Validation Secure Server CA 2	6
11	Yes	EV	GeoTrust EV RSA CA 2018	24
12	Yes	EV	DigiCert SHA2 Extended Validation Server CA	24
13	**NO**	**DV**	DigiCert SHA2 Secure Server	24
14	yes	EV	COMODO RSA Extended Validation Secure Server CA	24
15	yes	**DV**	COMODO RSA Domain Validation Secure Server CA	24
16	yes	EV	GeoTrust EV RSA CA 2018	24
17	yes	EV	GeoTrust EV RSA CA 2018	24
18	yes	EV	GeoTrust EV RSA CA 2018	9
19	yes	**DV**	Let's Encrypt Authority X3	3
20	yes	EV	DigiCert SHA2 Extended Validation Server CA	24

Around 90% of the Bulgarian, 96% of the Romanian and 89% of the Serbian bank web sites are using HTTPS without major issues. They use certificates from recognized certification authorities (CAs) that certify the connection between the public key used to encrypt the connection and the domain name stored on DNS servers (Tables 5, 6, 7).

Table 6. Main features in usage of the HTTPS in public web sites of Romanian banks

Bank №	Automatic redirection to HTTPS	Certificate type	Certification authority body	Validity in months
1	yes	EV	DigiCert SHA2 Extended Validation Server CA	14
3	yes	EV	GeoTrust EV RSA CA 2018	26
4	yes	EV	DigiCert SHA2 Extended Validation Server CA	12
5	yes	**DV**	GeoTrust RSA CA 2018	26
6	yes	EV	GeoTrust EV RSA CA 2018	14
7	yes	EV	DigiCert SHA2 Extended Validation Server CA	5
8	yes	EV	DigiCert SHA2 Extended Validation Server CA	13
9	yes	EV	DigiCert Global CA G2	13
10	yes	**DV**	Let's Encrypt Authority X3	3
11	yes	EV	DigiCert SHA2 Extended Validation Server CA	14
12	yes	EV	DigiCert Global CA G2	22
13	yes	**DV**	Thawte RSA CA 2018	27
14	yes	**DV**	GeoTrust RSA CA 2018	24
15	yes	**DV**	DigiCert SHA2 Secure Server CA	12
16	yes	**DV**	DigiCert SHA2 Secure Server CA	14
17	yes	EV	DigiCert SHA2 Extended Validation Server CA	24
18	**NO**	EV	DigiCert Global CA G2	13
19	yes	EV	GeoTrust EV RSA CA 2018	14
20	yes	**DV**	COMODO RSA Domain Validation Secure Server CA	24
21	yes	**DV**	Let's Encrypt Authority X3	3
22	yes	**DV**	Go Daddy Secure Certificate Authority - G2	12
23	yes	**DV**	DigiCert SHA2 Secure Server CA	26
24	yes	EV	Actalis Extended Validation Server CA G1	12

Three types of certificates are used: Domain Validation (DV), Organization Validation (OV), and Extended Validation (EV) [13], [14]. When validating a domain (DV), the certification authority checks to see if the applicant can use a specific domain name. No company identity checks are performed, and no other information is displayed in the browser, unless the connection is secure. Upon Validation of Organization (OV), the Certifying Authority additionally conducts a survey of the organization that appears when examining the certificate. In the Extended Validation (EV), the Certification Body carries out an in-depth verification of the organization with regard to the legal form of existence, real address, and right to use a particular domain, where the name of the organization is displayed in the browser along with the information that the link is protected. In general, the DV certificate is cheaper than EV certificate.

Table 7. Main features in usage of the HTTPS in public web sites of Serbian banks

Bank №	Automatic redirection to HTTPS	Certificate type	Certification authority body	Validity in months
1	yes	EV	Thawte EV RSA CA 2018	11
2	yes	**DV**	Go Daddy Secure Certificate Authority - G2	14
3	yes	**DV**	Entrust Certification Authority - L1K	24
4	**NO**	**DV**	Thawte RSA CA 2018	14
5	yes	EV	GeoTrust EV RSA CA 2018	12
6	yes	**DV**	cPanel, Inc. Certification Authority	3
7	yes	**DV**	DigiCert Global CA G2	12
8	yes	**DV**	Go Daddy Secure Certificate Authority - G2	36
12	yes	**DV**	Go Daddy Secure Certificate Authority - G2	37
13	yes	**DV**	Thawte RSA CA 2018	13
14	yes	**DV**	GlobalSign Organization Validation CA - SHA256 - G2	25
15	yes	**DV**	COMODO RSA Domain Validation Secure Server CA	24
16	**NO**	**DV**	Let's Encrypt Authority X3	3
17	yes	EV	GeoTrust EV RSA CA 2018	14
18	yes	**DV**	cPanel, Inc. Certification Authority	3
19	yes	**DV**	Thawte RSA CA 2018	14
20	yes	EV	Thawte EV RSA CA 2018	24
21	yes	EV	GeoTrust EV RSA CA 2018	14
22	yes	**DV**	Thawte RSA CA 2018	24
23	**NO**	**DV**	cPanel, Inc. Certification Authority	3
24	**NO**	**DV**	Let's Encrypt Authority X3	3
25	yes	**DV**	Actalis Organization Validated Server CA G1	12
26	yes	EV	GeoTrust EV RSA CA 2018	14
27	yes	**DV**	cPanel, Inc. Certification Authority	3

Table 8. Certificate type of the SSL certificates

Certificate type	Bulgaria		Romania		Serbia	
	Count	**%**	**Count**	**%**	**Count**	**%**
No certificate or problem	2	10	1	4	3	11
DV	5	25	10	42	18	67
EV	13	65	13	54	6	22
TOTAL	***20***	***100***	***24***	***100***	***27***	***100***

From the data presented in Tables 5, 6 and 7, in the Bulgaria the majority (13 banks) uses EV, and the rest (5 banks) use DV. In Romania the situation is the same - the majority

(13 banks) uses EV, and the rest (10 banks) use DV. But in Serbia is the opposite - the majority (18 banks) uses DV, and the rest (6 banks) use EV (Table 8, Fig. 1).

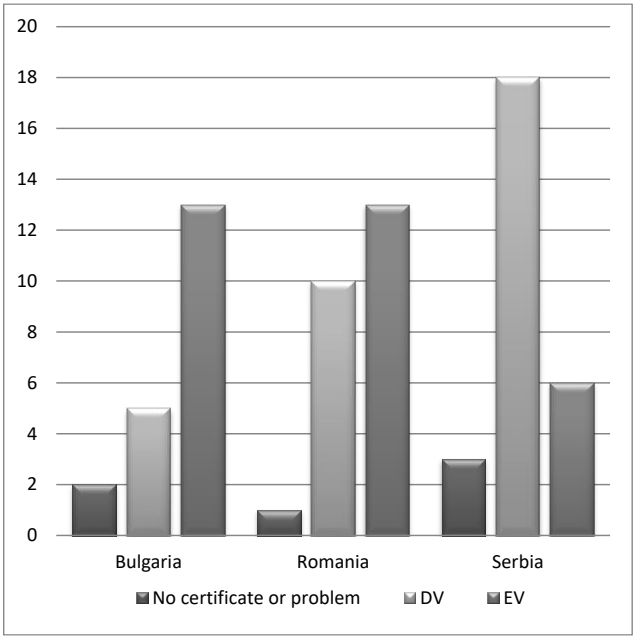


Fig. 1. Shares of certificate types.

In two cases in Bulgaria, one case in Romania and four cases in Serbia of the banking websites, the good practices are not followed, and they do not automatically redirect from unsecure to secure connection.

Table 9. The issuers of the SSL certificates used by Bulgarian and Serbian banks

Certification authority body	Bulgaria		Romania		Serbia	
	Count	%	Count	%	Count	%
Actalis	-	-	1	4	1	4
COMODO	6	33	1	4	1	4
cPanel	-	-	-	-	4	17
DigiCert	4	22	12	52	1	4
Entrust	-	-	-	-	1	4
GeoTrust	6	33	5	22	4	17
GlobalSign	-	-	-	-	1	4
Go Daddy	-	-	1	4	3	13
Let's Encrypt	1	6	2	9	2	8
RapidSSL	1	6	-	-	-	-
Thawte	-	-	1	4	6	25
TOTAL	***18***	***100***	***23***	***100***	***24***	***100***

There is a wide variety of preferences for a certification authority, but the most popular choices in Bulgaria are: Comodo and GeoTrust - 6 banks, DigiCert - 4 banks. It should be noted that RapidSSL is owned by GeoTrust, so actually the real leader is the GeoTrust. In Romania the most popular choice is DigiCert - 12 banks, followed by GeoTrust - 5 banks. In Serbia the most popular choices are other market players, which are not presented in Bulgaria and poorly presented in Romania: Thawte - 6 banks, cPanel - 4 banks, Go Daddy - 3 banks. The only large common player is GeoTrust - 4 banks (Table 9).

One bank in Bulgaria and two banks in Romania and Serbia are using free of charge 3 months-long Let's Encrypt certificates.

4. Conclusion

This research leads to several conclusions. First, the bank sector in Bulgaria is more consolidated that this in Romania and Serbia. Second, as for the use of SSL certificates, in Serbia there is more diversity - ten SSL certificate providers, while in Bulgaria - they are only five, in Romania - seven. In Serbia the most popular SSL certificate provider is Thawte with share of 25% web sites. In Bulgaria the most popular is GeoTrust with 39% share (if taking into account RapidSSL). In Romania the most popular choice is DigiCert with 52% market share. It is interesting that one Bulgarian and two Romanian and Serbian banks are using free certificates from Let's Encrypt Authority. Two banks in Bulgaria, one bank in Romania and four banks in Serbia are not redirecting automatically from unsecure HTTP to secure HTTPS connection. Two banks in Bulgaria, one bank in Romania and three banks in Serbia are not using SSL or have some problem with the certificates. In Bulgaria the average validity of certificates is 19 months (median - 24 months), while in Romania and Serbia - more shorted - around 16 and 15 months (median - 14 months).

The collected data are related to period April 2019. The results of the study could have important practical impact for bank managers and IT specialists when evaluating options which technologies to implement in order to minimize the risk to the financial institution. Also, the results reveal some good practices used in Bulgarian, Romanian and Serbian banks. The research conducted on the use of the HTTPS protocol on the banks' public web sites covered the sites of all 20 Bulgarian, 24 Romanian and 27 Serbian banks licensed to operate on the respective country territory by the domestic National Banks.

From banks that are using HTTPS without major problems, in Bulgaria the majority (13 banks) uses Extended Validation (EV) types of certificates, and the rest (5 banks) use Domain Validation (DV). In Romania the situation is the same - the majority (13 banks) uses EV, and the rest (10 banks) use DV. But in Serbia is the opposite - the majority (18 banks) uses DV, and the rest (6 banks) use EV. We suppose that one of the reasons for this is because DV certificates is usually cheaper than EV certificates.

As it is known, when using text protocols (such as HTTP/0.9/1.0/1.1) the so called "man in the middle" could be listening to all traffic and exchanged queries and responses between clients and servers. Network packets can easily be read, even without the use of complex tools. In many cases confidential information is exchanged, e.g. passwords, bankcard numbers for payment, personal information, etc. In those cases, the best option is to use the specially intended for this purposes protocol HTTPS, also known as "HTTP Secure", "Secure HTTP", "HTTP over SSL", "HTTP over TLS", etc. The use of a secure connection increases the processing load on the client and on the server in terms of busy CPU time and the amount of RAM, but in recent years, this is not considered a serious argument given the great benefits of communication security.

References

1. National Statistical Institute (2018). *Population by districts, municipalities, place of residence and sex as of 31.12.2018* <http://www.nsi.bg/en/content/6704/ population-districts-municipalities-place-residence-and-sex>
2. National Institute of Statistics (2018). *Population* <http://www.insse.ro/cms/sites/default/files/com_presa/com_pdf/poprez_ian2018e.pdf>
3. Statistical Office of the Republic of Serbia (2017). *Estimated population*, <http://www.stat.gov.rs/en-us/vesti/20180629-procene-stanovnistva-2017/?s=1801>
4. The World Bank Group, GNI per capita, Atlas method (current US$), <https://data.worldbank.org/indicator/ NY.GNP.PCAP.CD>
5. Sovereigns Ratings List, <https://countryeconomy.com/ratings> (20.04.2019)
6. Petrov, P., et al. (2018). Web Technologies Used in the Commercial Banks in Finland. *CompSysTech'18*, University of Ruse, Bulgaria, ACM, 94-98.
7. Petrov, P., Dimitrov, G. & Ivanov, S. (2018). A Comparative Study on Web Security Technologies Used in Irish and Finnish Banks. *18 International Multidisciplinary Scientific Geoconference SGEM*, Vol. 18, Issue 2.1, 3-10.
8. Petrov, P. (2013). Trends In The Use Of Web Server Software In Bulgarian Banks. *International Conference On Application Of Information And Communication Technology And Statistics In Economy And Education (ICAICTSEE-2012)*, UNWE, 359-364.
9. Petrov, P. & Hundal, S. (2018) Application of Security Technologies in the Public Sites of Banks in Serbia. Izvestia Journal of the Union of Scientists - Varna. Economic Sciences Series, Union of Scientists - Varna, Vol.7, Iss.2, pp.298-305.
10. BNB, Banks Licensed in the Republic of Bulgaria as of 11.02.2019, <http://www.bnb.bg/ BankSupervision/BSCreditInstitution/BSCIRegisters/BS_CI_REG_BANKSLIST_EN>
11. Banca Naţională a României, REGISTRUL INSTITUTIILOR DE CREDIT LA DATA 18-01-2019, PARTEA I - Sectiunea I - Banci, <http://www.bnr.ro/files/d/RegistreBNR/InstitCredit/ban1_raport.html>
12. National Bank of Serbia (NBS), List of Banks as of 22.03.2019, <https://www.nbs.rs/internet/english/ 50/50_2.html>
13. Cooper, D., et al. (2008) RFC5280, Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile, <https://www.ietf.org/rfc/rfc5280.txt>
14. Saint-Andre, P., Hodges, J. (2011) RFC6125, Representation and Verification of Domain-Based Application Service Identity within Internet Public Key Infrastructure Using X.509 (PKIX) Certificates in the Context of Transport Layer Security (TLS), <https://tools.ietf.org/rfc/rfc6125.txt>

Business Models for Starting Software Companies

Svetoslav Ivanov, Pavel Petrov

University of Economics - Varna, Varna, Bulgaria
svetoslav_ivanov@ue-varna.bg, petrov@ue-varna.bg

Abstract. Companies implement and utilize business models in their activity. The starting companies and software companies have specific ones. The paper aims a theoretical review of the business models appropriate for starting software companies. The subject is a starting software company that develops software products or services to prove their expediency, feasibility and market potential. The products can be sold on the market with one or a combination of available theoretical business models. These models are an essential part of the entrepreneurial process to reach market success and growth.

Keywords. Software development, Business model, Entrepreneurship, Starting company

1. Introduction

Information technologies development and the effect on business they provide drive their implementation in various sectors and functions of the business. Their role moves from supporting one to a key factor in the business. IT allows implementing the digitalization of parts of the business, which changes the processes in companies and even the type of business or the way the business operates. Two key components are in this process – the software and the innovation process. It is well known following the boom in the software industry and entrepreneurship in software development, how many failures happen due to inappropriate management of the innovation process or the starting business. In order these initiatives to be successful and sustainable, they should be accepted by the market and the business should be vital and profitable. The starting software companies are established and managed by entrepreneurs. To achieve the goal for market acceptance and grow, they should answer the fundamental questions - "How we make money in this business? What is the underlying logic that explains how we can deliver value to the customer at an appropriate cost?" [15]. To answer these questions, the entrepreneurs should find and implement appropriate business models, so the new product or service they invent to be successful.

The purpose of this paper is to review the most common theoretical concepts for business models appropriate for starting software companies. To achieve this goal, we will focus on the following: definition of the subject – the starting software company, define the term "business model" and "software product", as well as review the theoretical concepts for business models appropriate for the subject.

2. Definition of Starting Software Company

By the means of the name, the term "starting software company" is related to the terms "starting technology company", "technology start-up", "starting enterprise", "software company" etc. The legal base in the European Union and Bulgaria uses the term "enterprise" and divides the enterprises in four categories – micro, small, medium and big. The distinction is based on the size of the company measured as the number of employed people and annual turnover or the value of assets. Using this classification, we will focus mostly on micro and

small enterprises - less than 50 employees and less than 10 million Euro turnover or assets [7]. Alternative approaches are to measure the annual revenues or the stage/amount of investments. One simple criterion is "before becoming a big company" [1]. All these criteria are equivalent to the definition for the size above.

To specify better the term "starting software company", we add the criterion for age. The definitions in the literature are different for the age – up to 3, up to 5 or 10 years, or "have little or no operating history" [1]. The shorter cycle of innovation and technological development lead us to select a period of fewer than 5 years.

The third criterion defining the company is based on the key goal in the business – to develop a new product, to apply a business model, to prove vitality and potential growth. The definition for a start-up from this point of view is "a company or project initiated by an entrepreneur to seek, effectively develop, and validate a scalable business model" [23]. Similar and related goals are to reach satisfactory product-market fit, to develop customer base where the focus is on searching clients, adapting the product and business model that fits the market (so-called Customer Development Model) [1]. Another formulation of the same goal is that such companies may target new markets. The new markets are characterized by low structure, high unpredictability, and limited understanding of customers, rivals and market segments [12, 24]. "It seems likely that moves that explore the market for opportunities - rather than moves that exploit existing positions – are particularly key" [11]. This leads us to the next statement, that an established company focus on exploiting existing positions, or in other words uses Product Development Model to optimize the processes and structures for established products and markets. This criterion is the most important one for the distinction between a starting and an established company.

Some authors, describing start-up companies give an accent to whether the company wants to become big or more precise "to develop and validate a scalable business model" [1]. In our study, we will skip the starting companies which plan or behave to stay small – for example freelancers, lifestyle, small family or service businesses. Therefore, we add the definition "search for growth or scalable model".

In the literature can be found descriptions of different forms of entrepreneurial activity. Many of them are described as activities that copy other company activity and products, reuse a ready product, business model, and brand at a high level like a franchise, business acquisition, etc. On the fifth criterion, we will focus on a model with own product and business model. Even when it uses third party components, a known business model, the company has at some level unique product, model, brand, etc.

Including the word "software" in the term shows that the company is going to develop technological product software. We will use the term "technology company" in the means to produce a technological product. In the literature, there are many definitions of the term "software", it's content and development, and specific features we review below.

Finally, a starting company is initiated by one or more entrepreneurs. We will use the term "entrepreneur" for a single entrepreneur or a team of entrepreneurs who initiate and lead a starting software company. There are researches [4, 26] that describe and prove that the main factor for SME company success and growth is connected to the entrepreneur and his/her/their personal goals, motivation, orientation to market, capability for growth, and professional management. Therefore, the entrepreneur is a key factor. This leads us to the conclusion that the knowledge and successful implementation of an appropriate business model is a factor for company success.

After all these definitions we could summarize that a starting software company is a new micro, small or medium company up to 5 years, founded by one or team of entrepreneurs that develop a new software product and search business model that proves feasibility, vitality

and product-market fit with growth potential. We should note that there are differences between such companies. The criteria could be different - model, strategy, processes, product, team, entrepreneur, type of product (product, service, platform, component, solution), customer-driven or technology-driven, and others. But in this article, we focus on two common elements – a new software product and the business model.

The next term we need to specify is "product". In our study, we use a more general meaning of product as a result of company activity. In the literature very often instead of terms innovations, entrepreneurship, start-ups, business modeling are used different terms like product [23], solution [14], service, customer value [1, 20], etc. We use the term "product" as a result of the innovation process for a starting software company. A similar definition is: "Anything customers experience from their interaction with a company should be considered part of that company's product" [23 p.38]. The product is technological innovation based on innovation with IT technologies. The term "technology" is defined as "knowledge, skills, and artifacts, which could be used for the development of products and services, as well as their production and delivery systems" [2]. In our paper, we refer mostly to software technologies, but often this includes other information and communication technologies, organizational changes, etc. The term "innovation" we use in the meaning of "a result from the innovation process, which presents all activities, which lead to products, services or their production or delivery systems, which could be successfully realized on the markets" [10]. The important distinction here is that one or more technologies could be considered as discovery, knowledge, the market success in a form of product (or service, solution, value, etc.) done with this technology makes it a technology innovation. In the literature concerning innovations, there are classifications for the type of innovations, but we will mention one more important classification for the innovations [2]:

- Evolutionary – innovations that aim optimization of existing systems, products and structures.
- Radical – innovations that change the existing processes, products and structures.
- Architectural – innovations that change the components of existing processes, products and structures.

The starting software company is defined by the result – the software product. The software has specific features described in the literature as a product - both production and support, and we will mention some of them [8, 19, 20]:

- Fast development in IT and software technologies require short releases.
- Product complexity - various vendors of platforms, components, tools, etc.
- Quality is a complex term – features, security, scalability, performance, user experience, robustness, integration with 3rd-party software, and others.
- High-quality requirements - increasing quality requirements over time.
- Requirements - not clear initially, need discovery, may change over time.
- Serious investment before first working or marketable copy.
- Multi-disciplinary – requires the participation of various specialists.
- Product defects share in all copies of the product.
- The high price of software development specialists.
- There is extensive work with knowledge and a need for constant learning.
- Must work on various devices, platforms, browsers, servers, etc.
- Regular upgrades - because of component changes, technology changes, bugs, security threats, legal, etc., the product needs regular upgrades.
- Available support - the product requires available support service.
- Quality or time of delivery could be a critical problem

- Specific licensing rights to meet and keep.

Therefore, the process of software product development should reflect these changes. Two key concepts in successful software development are user involvement and requirements description and stability. For starting software company this is not the case – it works often under uncertainty. Initially, the user or customer may not be clear, requirements may change during the search of the market niche fit. The feedback is not always direct – asking questions about getting the answer. It might need an analysis of data, experiments to validate a supposed requirement. Also, business analysis is required for the financial and marketing effect of every requirement.

3. Definition of Business Model

The market success of a product can be realized with a specific business model. The term "business" means a sustainable organizational structure whose purpose is the return on investment [25]. The term "model" is defined in the dictionary as "a standard or example of limitation or comparison" or "a representation, generally in miniature, to show the construction or appearance of something" [21]. Another definition is that models represent the real object in an abstract way or phenomenon to show its essence, characteristics, dependencies, interactions by using a language for model description [8 p.31-32]. The combination of both terms is our focus.

The term "business model" is first defined by Peter Drucker in his article "The Theory of the business" as the company's theory of the business. He states "what business will and won't do", as well as "The assumptions on which the organization has been built … the assumptions that shape any organization's behavior, dictate its decisions … the assumptions about markets, identifying customers and competitors, … technology and its dynamics, about a company's strengths and weaknesses, … about what a company gets paid for" [5]. In the literature of management and entrepreneurship can be found many definitions of the term "business model", where they look at the definition from different aspects of the business organization. There is no single definition for it. "At a general level, the business model has been referred to as a statement, description, a representation, an architecture, a conceptual tool or model, a structural template, a method, a framework, a pattern, and as a set..." [31]. In most cases, authors give a set of definitions. We will use definitions more appropriate for a business model related to starting software company as follows:

- "A concise representation of how an interrelated set of decision variables in the areas of venture strategy, architecture, and economics are addressed to create the sustainable competitive advantage in defined markets" [16].
- "Describes the rationale of how an organization creates, delivers, and captures value" [18].
- "Refers to the core architecture of a firm, specifically how it deploys all relevant resources (not just those within its corporate boundaries) to create differentiated value for customers" [27].
- "Describes how a company makes money" [6].
- "The business model depicts the content, structure, and governance of transaction designed so, as to create value through the exploitation of business opportunities" [31].

From these definitions, we understand that the business model is the integration of decisions and representation of the business logic to provide competitive advantage and market realization of the innovation based on relevant resources. At the same time, it must keep the company a sustainable structure and to assure return on investment (often expressed with terms like make money, generate revenue, make a profit, etc.). The business model is connected to

the business strategy, it is dedicated to a product and this determines the component competitive advantage. On the other hand, the strategy is usually for the company and does not determine the other components of the business model. The "business models describe, as a system, how different pieces of a business fit together. But they do not factor in one critical dimension of performance: competition". The "strategy explains how you will do better than your rivals" [15].

The business model differs from the business processes and business plan. The business modeling is not a process modeling. "The central notion in any business model should be the concept of value, to explain the creation and addition of value in a multi-party stakeholder network, as well as the exchange of value between stakeholders". The business process modeling determines how value-creating and value-exchange activities are carried out [9]. The value as a concept is in terms of benefits and revenues [30]. The development of a business model is part of the innovation process. As Croll and Yoskovitz write in their book "One thing we noticed about almost all successful founders we have met is their ability to work at very detailed, and a very abstracted level in their business ...they are not only trying to run a business ... they are also trying to discover the best business model" [3].

In the literature, there are different classifications of business models. Morris defines them as Economical, Operational and Strategic [16]. Others see the business models as a tool and define them as Architectural – a tool for strategic decision modeling, Representational – a diagram to present it, and Example - to illustrate a class of business model. We will skip them and focus on more important details like business models components and those for starting software companies.

4. Business Model Components

As every model, a business model could be described with various tools. It could be described with text mentioning the components of the business model. A standard methodology for business modeling could be used, and this is discussed later in this paper.

Every business model has own components. The components determine various sides of the business and the value it creates and delivers. Some of the cited business models below have a focus on specific aspects of the business and the components do not contain all components. Others try to cover more aspects. We will list the main list or key list of components that are good to address in a specific business model.

The main component of a business model is the Value Proposition [17,18]. It contains the product, services, solution as well as their attributes as quality, price, image, etc. In some models, it contains a competitive advantage and in others, the competitive advantage is a separate component [17]. In customer-driven models we can see separate sub-components about the problem that the company solves, the existing alternative solutions, the company's solution and the reasons that it is better (competitive advantage) [1, 14].

The next component is the market. It contains specific structures of the industry, distribution methods, sales channels, customer segments. Some authors separate the customer segments, distribution channels, and customer relationships in separate components to make the business model more practical [6]. Other authors mention this component directly using the term "customers" or any of its synonyms - clients, users, etc.

The third component or set of components is related to how the company will create and deliver the value. They are about the resources, the activities, the processes and structures, the partner networks (suppliers, distributors, etc.). They provide the answer to the question of how the company will be directly connected to the fourth component – about the costs. And here it relates to the fourth component, the costs of the company.

The fifth component is the revenue – how the company is getting paid for the value it delivers. Sometimes it is presented in a simplified manner as the business model, but it is one of the components. The goal – return on investment is formed from 4th and 5th component – the costs and the revenue which form the model of profit and model of liquidity [17].

The sixth component is the delivery (known also as logistical) component. It determines how the product (value) is delivered to the customer. This means the design of the supply chain for the business, delivery service, logistical network,

The seventh component is the organizational unit, the position of the company which creates the value. It is connected to the core competencies of the organization [6, 18]. Tapscott illustrates the idea of B-webs that utilize the Internet, where many organizations have own position in the B-web and concentrate on their core competencies. This way, the B-webs appears more supple, innovative, cost-efficient, and profitable than traditional vertically integrated competitors [27]. Therefore, each business model should have a component specifying the position of the organization in the B-web and is connected to the component specifying the partnerships with other organizations in the markets.

5. Business Models for Starting Software Company

In the literature, there are many described business models appropriate for software products (also, services, solutions, etc.). We will try to present them based on various criteria. Most of them could be found in the literature for software development, e-business, e-commerce, and others [3, 8, 13, 28]. The known business models could be characterized and classified by different criteria.

First, they could be classified by the type of actors – business, end customer, administration and other. Therefore, there are various combinations between the actors exchanging value – B2B, B2C, C2C, C2B, B2B2C, etc. Second, the organization which does the business model could be in a specific type of industry and network, where each participant has a specific role. For example, later in the paper, we describe Tapscott's business webs (B-web) as a network of business relations and each has a specific role [27]. Therefore, business models could be classified according to the organization's role in the business web or ecosystem.

The third criterion could the level of aggregation used to classify a business model. A model could be more general or concrete and detailed. A more general model will describe the basic structural assumptions of the business model and could be detailed later in sub-types, each specifying more details about the model. For example, in e-commerce, we have a business model for selling products online. These could be detailed to a model where buyers select from a limited list of products proposed by sellers and fixed prices. In more detail, this could be the company's e-shop, third party e-shop or participate in third party business that integrates multiple sellers like e-mall, etc. Different authors use different criteria for segmentation, and we will list them and note some similarities. This leads to the fourth criteria – the model's author and the fifth is the criteria by which segmentation is done. We will list a few authors of business models that could apply for a software company and could be found in the literature for e-business. Those are listed in tables, sorted historically, each for one author. Each model will be named and described briefly, as well as examples from another author will be given if it is a more general one. It is possible, a model from one author to match exactly model from another author or to have great coverage or to be in an abstraction relationship – one is a group of models, the other is an element that fits in the group.

In Table 1 we list the earliest models from Timmers [29]. Those are the key e-business models that exist. Most are well-known electronic shops, marketplaces, and other e-commerce models.

Table 1. E-business models from Timmers [29, 31].

Model	Description
e-Shops	Web marketing and promotion that increases the possibility to order and pay.
e-Procurement	Electronic tendering and procurement.
e-Malls	Collection of e-shops enhanced by a common umbrella (well-known brand).
e-Auctions	Electronic implementation of the bidding mechanism from traditional auctions.
Virtual Communities	Brings together virtual communities that contribute value in a basic environment provided by a virtual community operator. Revenues come from membership and advertising. It could be used for customer loyalty and feedback to other marketing operations.
Collaboration Platforms	A set of tools and information environment provided for collaboration between enterprises.
Third-party Marketplaces	The company leaves web marketing to third-party that offers a user interface to the supplier's product catalog.
Value Chain Integrators	Represents companies that integrate multiple steps of the value chain and may exploit the information flow between steps for further added value.
Value Chain Service Providers	Stands for companies that specialize in a specific function for the value chain, such as electronic payments or logistics.
Information Brokerage	Embraces a whole range of new information services that are emerging to add value to the huge amount of data available on the open networks or coming from integrated business operations.
Trust and Other Third Parties	Stands for trust services, such as certification authorities and electronic notaries and other trusted third parties

The models from Timmers stated above are more detailed. A more abstract view of business models is given from the typology description of b-webs [28]. The classification is based on two criteria – economic control and value integration. Hierarchical b-webs have a leader who controls the value proposition, pricing, evaluation, customer satisfaction, and transactional flow. Other b-webs do not depend on many actors' decisions for the value proposition, pricing, etc. (open-source, stock exchange). Value Integration means how much value/options the company integrates into the product from other suppliers. The low level of Value Integration means to propose the product as is and high integration means to get multiple parts, options, add-ons and propose them as one product. In Table 2 are listed the types of b-webs, which could be assumed as a description of the type of network with specific role for a company. This determines the key type of business model components for company role, value proposition, control over pricing, etc.

Another list of classified at higher level business models is proposed by Applegate. It aggregates in categories many types of models and describes the sub-types in each category. The list of models is [31]:

- Focused Distributors – provide products from specific industry or market niche (retailers, marketplaces, aggregators, exchanges, infomediaries).
- Portals – horizontal, vertical and affinity portals differentiated based on gateway access, affinity group focus, revenue source, and cost structure.
- Infrastructure Distributors – Enable technology buyers and sellers to make transactions. Could be retailers, marketplaces or exchange of infrastructure.
- Infrastructure Portals – Enable consumers and businesses to access online services and information. Horizontal ones are Internet, network or hosting service providers, and vertical ones are producers and distributors of application service providers (ASP).
- Infrastructure Producers – Design, build, market and sell technology hardware, software, solutions and services - equipment manufacturers, software firms, customer software and integration, infrastructure software firms.

Table 2. E-business models from Tapscott [28].

Model	Description	Sample
Agora	Markets where buyers and sellers meet to freely negotiate and assign value to goods. Low integration.	Marketplace Auction Exchange
Aggregation	A leader takes responsibility for products, targeting, fulfillment, and prices. Limited value integration.	Amazon.com
Value Chain	The context provider structures and directs a b-web network to produce a highly integrated value proposition. The seller has the final decision for price.	Cisco systems
Alliance	High level of Value Integration and without control. Actors design products, create knowledge, create dynamic shared experiences. They depend on rules and standards that govern interaction.	Linux Wintel
Distributive Network	Keep the economy alive and mobile. Ensure the healthy balance of the systems they support. The support other b-webs by distributing and allocating goods and services.	Internet Postal Service

As listed in this classification, we could position software companies in the last group as software firms and customer software, integration, and infrastructure firms. But actually, the product of the software firm may fit and be operated in the other categories and the software company to develop a product as an infrastructure provider, but also the product is focused in a niche market or industry and it to operate it, so it fits into another category.

Another classification is from Rappa. It “classifies companies according to the nature of their value proposition and their mode of generating revenues” [31]. The list of models is described in Table 3.

Table 3. E-business models from Rappa [22].

Model	Description	Samples
Brokerage Model	Bring buyers and sellers and facilitates transactions. Revenues: fee or commission per transaction.	Marketplace, Auction, Trading Community, Distributor, Demand or Transaction Broker, Buy/Sell fulfillment, Virtual Mall
Advertising Model	The broadcaster provides content and services mixed with advertising messages in the form of banner ads. Content can be own or else created.	Portal, Personalized or Niche Portal, Classifieds, Registered users, Query-based Paid Placement, Contextual Advertising.
Merchant Model	(Information) intermediaries by collecting consumer data or producer products and sell the data to firms to find meaningful patterns and other useful information for better service. Wholesalers and retailers selling over the Internet.	Advertising Networks, Audience measurement Services, Incentive Marketing, Metamediary. Virtual merchant, Catalog merchant, Click and Mortar, Bit Vendor.
Manufacturer Model	A manufacturer reaches buyers directly through the Internet and compresses the distribution channel.	
Affiliate Model	When people surfing to provide purchase opportunities by offering financial incentives (a percentage of revenue) to affiliated partner sites.	The site contains points for purchase through click-to-the-merchant's site or sale option.
Community Model	Based on user loyalty. Users invest time and emotion in the site. They may contribute content or money.	Voluntary Contributor Model, Knowledge Network
Subscription Model	Users are charged a periodic fee to subscribe to a service.	Content Providers, Person-to-Person Networking Services, Trust Services, Internet Service Providers.
Utility Model	Based on metering actual usage of a service or pay-as-you-go approach.	

A very concise classification is given by Croll and Yoskovitz when they review different kinds of business models from the business aspects connected to marketing. Their focus on the business model description is to components as an acquisition channel, selling tactic, revenue, product type, delivery model. The list of models is described in Table 4 [3]. Interestingly, they list the options for each component appropriate for starting software company.

Table 4. Business models from Croll and Yoskovitz [3].

Model	Description	Revenues
E-commerce	A visitor buys something from the website of a retailer.	Sales
Software as Service	The software is used/delivered on an on-demand basis. It could be done usually through a website. Well-known formats are PaaS, SaaS, API.	Fees for period or charges based on consumption (operations, requests, time spent, etc.)
Free (Mobile) Application	An application can be used free of charges on a smartphone from an online store.	Advertising, purchases of features or in-app purchases.
Media site	Provide content on a web site and mix with advertising messages. The content could be own, from a third-party.	Advertising (views, click-throughs, affiliates, sales), sponsorships, etc.
User-generated content	Support an engaged community that creates content and shares it between users. Social networks are one example of the model.	Advertising, sales
Two-sided Marketplace	The company brings buyers and sellers to complete transactions. Auctions, Exchanges, Supporting sites.	Subscription fees and/or transaction fees.

In the literature could be found other similar classifications and lists of models (for example, Weill and Vitale) that seem common or contain similar models like the above one listed classified according to other approaches. What is common is that they are based on some criteria and each of them contains assumptions and rules, which could be used to construct the specific product or company business model. Of course, the business model has its structure of components and the mentioned models cover theoretically just a few of the components without parameters. The entrepreneurs in starting software companies may select any of the models or combine them to construct their business model and through an appropriate entrepreneurial process to calibrate the business model with more detailed assumptions, definitions, and parameters. Often, in entrepreneurial society, a given business model has a name that comes from the brand name of a very popular company applying it. For example, Table 4 with models could have these samples – Amazon for E-commerce, Azure for SaaS, Game for free mobile app, Medium or Yahoo or newspaper for a media site, Facebook or Twitter for user-generated content, eBay or Airbnb for a two-sided marketplace. In this case, very often the starting software companies could get advantages using such "ready well known" business models when they develop their business and business model:

- In a concise sentence to explain the key idea of the business with similarities – like "This model for this industry and these customers".
- To be understood quickly by partners (customers, investors, employees) for the essence of the business.
- To get a ready model that works in the same or other industry
- Quickly focus on testing and experimenting whether this will work.
- Get fast key metrics and tactics to evaluate and grow the business.

6. Business Models Description

Together with business model selection, creation and validation an entrepreneur could use tools to describe it as a model. Describing a model will help to clarify its components, to see the business model from different views. There are many techniques to present a business model. Among the standard business practices (describe it with structured text or structured presentation), we should note a few other specific ones for the goal. They present diagrams of the whole business model and could be detailed with diagrams of each component or sub-component. Very well-known is Alexander Osterwalder's Business Model Canvas [18]. It is suited to describe any business model, both new and established companies. The business model is described in a diagram on a sheet as a rectangle with different sections as shown in Figure 1.

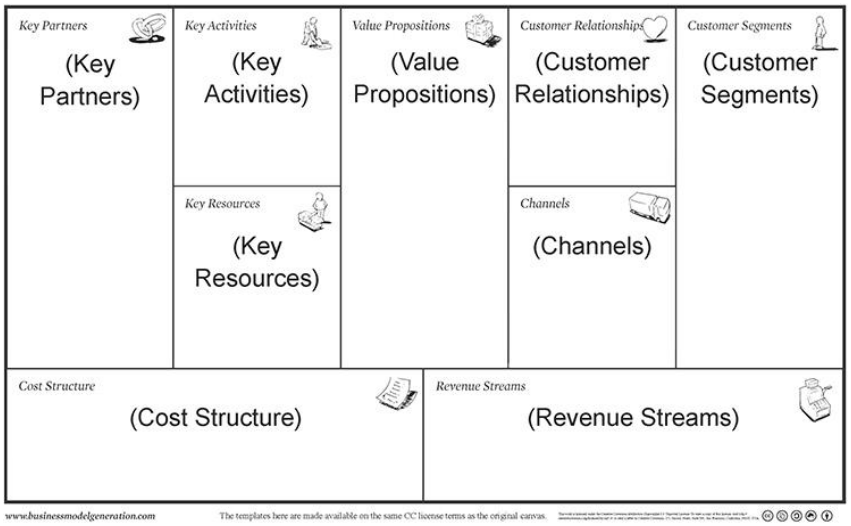


Fig. 1. Business Model Canvas
(source http://alexandercowan.com/business-model-canvas-templates).

A more appropriate form to describe a business model for starting companies is given from Ash Maurya named Lean Canvas [14]. It is a similar diagram in a sheet with a rectangle of business model components but is modified especially for the Lean Development process and optimized for Lean Startups (see Figure 2). It copies half of the components from the general model, but differs in the top-left components and focuses more on the Value Proposition and the related vital question for a startup - "What problem will solve our product and how it does it better for customers than competitors?". Therefore, the different components are the Problem (and existing alternatives), the Solution, and the Unfair Advantage. The Lean Canvas has a new component "Key Metrics" because metrics are vital in business management and the key goal of starting software companies is to prove vitality and growth potential of the solution. The other key sub-component is early adopters into Customer Segments because these will be the first customers to try and test the product. It is a base for the Customer Development process that is vital for the starting company.

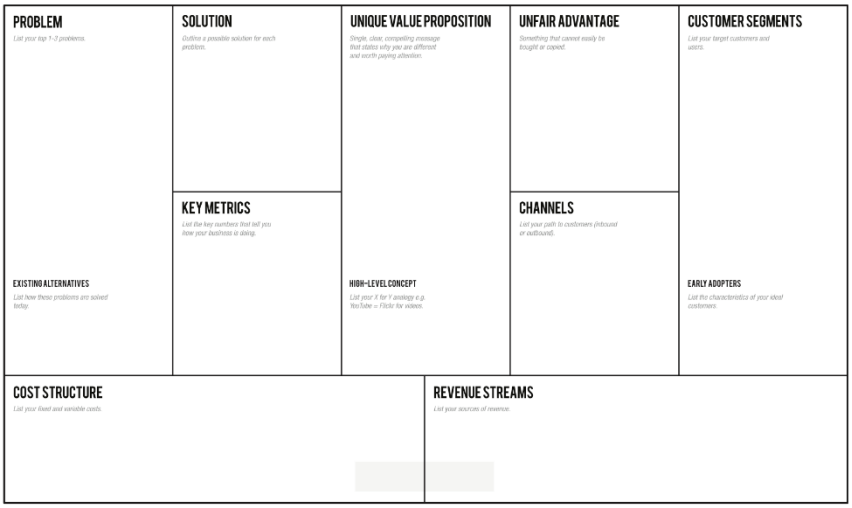


Fig. 2. Lean Canvas (source http://leanstack.com/leancanvas).

7. Conclusion

The current paper defined the starting software company and distinguish it clearly from other kinds of businesses and entrepreneurship. Each starting software company has the main goal to develop a product and prove market vitality and growth potential for it. The definition and classification of the term "business model" are given, as well as the components in a business model. Different theoretical classifications and definitions of business models appropriate for e-business and software companies are presented from Timmers, Tapscott, Ticoll and Lowy, Applegate, Rappa, Croll, and Yoskovitz. A starting software company has different options for a business model based on different options for each component of the business it needs to make decisions for and implement. Moreover, two approaches for describing business models are presented, where, the second one Lean Canvas is more suitable for starting software companies. The stated models show that the aim of the paper is met to make a theoretical review of the business models appropriate for software companies. The details for business models are a promising area for future research on the topic. It could be extended with future research in two directions – optional alternatives for each component in the business model that are appropriate for starting software companies and metrics to evaluate the business model and the business progress in different stages of development.

References

1. Blank, S. *The Four Steps to the Epiphany*, K&S Ranch Publishing, 2013.
2. Burgelman, R., *Strategic management of technology and innovation*, 4th ed., Boston: McGraw-Hill Irwin, 2004.
3. Croll, A., Yoskovitz, B., *Lean Analytics, Sebastopol*, CA: O'Reilly Media, 2013.
4. Davidsson, P., *Continued entrepreneurship and small firm growth*. Stockholm: The Economic Research Institute, 1989.
5. Drucker, P. *The Theory of the Business. Harvard Business Review*, September 1994 issue (https://hbr.org/1994/09/the-theory-of-the-business) [Accessed: 01/10/2019].
6. Dubosson-Torbay, M., Osterwalder, A., Pigneur, Y. *E-business model design, classification and measurements.* Thunderbird International Business Review, 2002, 44(1), pp.5-23.

7. European Commission. *What is an SME*? (https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en/) [Accessed: 01/10/2019].
8. Eskenazi, A., Maneva, N., Softuerni Tehnologii. 2nd ed. Sofia: KLMN, 2006. (in Bulgarian)
9. Gordijn, J., Akkermans, H., Vliet, H. *Business Modelling Is Not Process Modeling.* In Liddle, S., Mayr, H, Thalheim, B. hrsg. *Conceptual Modeling for E-Business and the Web.* Lecture Notes in Computer Science. Berlin Heidelberg, Germany: Springer, 2000, pp.40-51.
10. Hart, M., Stevenson, H., Dial, J., *Entrepreneurship: A Definition Revisited*. In Proceedings of 15th Annual Entrepreneurship Research Conference. Frontiers of Entrepreneurship Research. Boston, p.75-89.
11. Katila, R., Chen, E., Piezunka, H. *All the right moves: How entrepreneurial firms compete effectively.* Strategic Entrepreneurship Journal, Volume 6 Issue 2, June 2012, pp.116-132.
12. Katila, R., Shane, S. *When does lack of resources make new firm innovative?* Academy of Management Journal, 48(5), pp.814-829.
13. Laudon, K., Traver, C., *E-Commerce 2017*. Boston: Pearson, 2017.
14. Maurya, A. *Running lean: iterate from plan A to a plan that works*. The lean series (2nd ed.). Sebastopol: O'Reilly. 2012.
15. Magretta, J. *Why Business Models Matter*. Harvard Business Review. 2002 May Issue. (https://hbr.org/2002/05/why-business-models-matter) [Accessed: 01/10/2019].
16. Morris, L., *Business Model Warfare.* Pensylvania, USA: Ackoff Center for the Advancement of Systems Approaches (A-CASA). 2003.
17. Mullins, J., Komisar, R. *Getting to Plan B: Breaking Through to a Better Business Model*, Harvard Business Press. 2013.
18. Osterwalder, A., Pigneur, Y. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers.* 1. Aufl. Hoboken, NJ: John Wiley and Sons. 2010.
19. Panayotova, G., Dimitrov, G., Petrov, P., Bychkov, O. *Modeling and data processing of the information systems.* 3rd International conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, Poland, IEEE, 2016, pp.154-158.
20. Poppendieck, M., Poppendieck, T. *Lean Software Development: An Agile Toolkit.* Addison-Wesley Professional. 2003.
21. Random House Inc., Model – Define model at Dictionary.com. Random House Dictionary. (http://dictionary.reference.com/browse/model/) [Accessed: 01/10/2019]
22. Rappa, M. *Business models on the web: Managing the digital enterprise.* (http://digitalenterprise.org/models/models.html) [Accessed: 01/10/2019]
23. Ries, E. *The Lean Startup*. Crown Publishing Group. 2011.
24. Santos, F., Eisenhardt, K. *Constructing markets and shaping boundaries: entrepreneurial power in nascent fields.* Academy of Management Journal, 52(4), pp.643-671.
25. Scarborough, N., *Essentials of Entrepreneurship and Small Business Management*, 7.Aufl., New Jersey, USA: Prentice Hall PTR. 2013.
26. Storey, D. *Understanding the small business sector.* London: Routledge. 1994.
27. Tapscott, D. *Rethinking Strategy in a Networked World [or why Michael Porter is wrong about the Internet]. Strategy and Business*, 2001, pp.34-41.
28. Tapscott, D., Ticoll, D., Lowy, A. *Digital Capital: Harnessing the Power of Business Webs.* Ubiquity. 2000.
29. Timmers, P. *Business models for electronic markets.* Electronic Markets, 1998, 8(2), pp.3-8.
30. Timmers, P. *Electronic Commerce: Strategies and Models for Business-to-Business Trading.* John Wiley & Sons Ltd., Chichester, England, 1999.
31. Zott, C., Amit, R., Massa, L. *The Business Model: Theorethical roots, recent developments, and future research.* IESE Business School, WP-862, June, 2010. (http://bmcommunity.sitew.com/fs/Root/8jif5-AmitZottMassa.pdf) [Accessed: 01/10/2019].

A Critical Look at the Metrics for Measuring the Effectiveness of a Cybersecurity System

Nedko Tagarev

University of National and World Economy, Sofia, Bulgaria
ntagarev@e-dnrs.org, ntagarev@gmail.com

Abstract. This publication is part of the material of science project "Economical aspects of security in smart settlements", a University of National and World Economy №15/2018. In this publication are discussed the theoretical aspects of metrics and data that is collected for improving the effectiveness of cybersecurity system. The following text is based on the principle that the security goals have to be in line with business objectives. There are explained the requirements for good metrics, and the usefulness of other ones. There are shown some popular metrics that are not used very often for strengthening security.

Keywords. Cybersecurity, Metrics

1. Introduction

A major problem in assessing the effectiveness of security mechanisms in cybersecurity is the lack of a methodology that shows how the cybersecurity system contributes to the overall economic efficiency of the organization. The lack of such information interferes the organization's management from making informed decisions regarding cybersecurity. The same problem applies to the funds allocated for the construction of protection mechanisms directly connected to the information infrastructure.

The purpose of analysing and choosing the right metrics is important for making the right decisions regarding the cybersecurity of organizations. The result of the right solutions is system resilience and business process continuity. In large organizations, it is difficult to apply a specific metric or one of the metrics for cybersecurity analysis. The many possible points of attack, which increases the scale of the attack and the possible subsequent incident, make this difficult. In this situation, it is important to assess the quality and usefulness of the input information, the analysis process and the output product in the form of a cybersecurity product. In this, in this case, the basic model of process analysis does not change, but its composite factors determine the result.

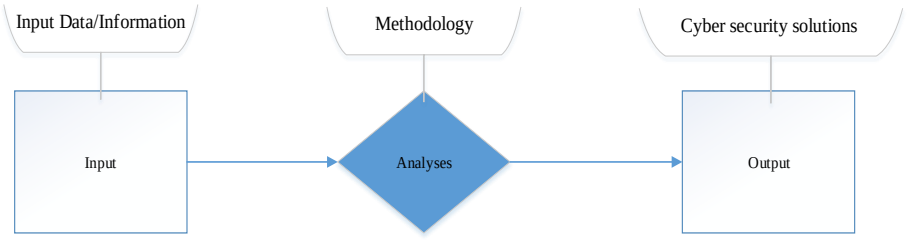


Fig. 1. The basic model for process analyses of information

The importance of the metrics used to measure cybersecurity aims to:

- assess the internal environment of the organisation;
- Communication has to be in business language, as the security program must have measurable business values. To make unification between cybersecurity goals and business goals of the organisation;
- Objective presentation of security issues for assessment and treatment;
- Justification of new investments in cybersecurity;
- Improvement of a system for cybersecurity to increase the value of the business.

2. "Good" versus "bad" metrics

The data collected is determined by the metrics used to measure cybersecurity. The purpose of using the "right" metrics is to properly assess cybersecurity and to be able to able to justify the application of protection measures as well as the investments made for them. The metric evaluation process can be defined in five stages [1] using the principles of a systems approach (Figure 2).

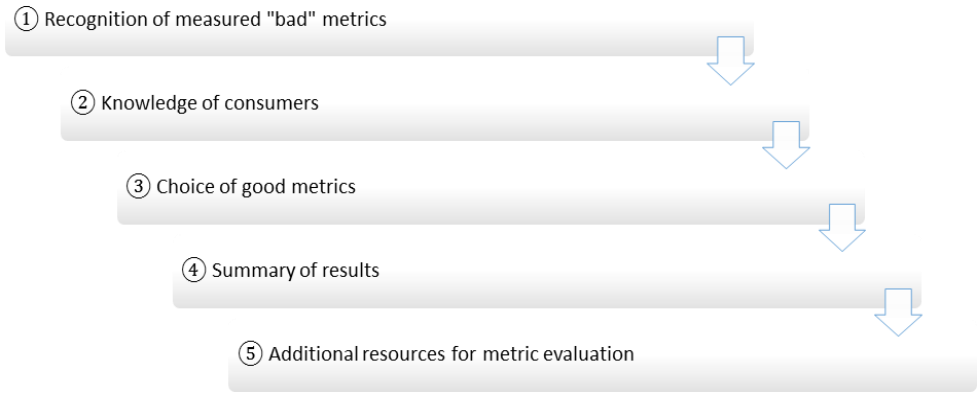


Fig. 2. Stages in the evaluation of metrics

The comparison between "bad" and "good" metrics is shown in Figure 3. A key feature of "bad" metrics is that they are easy to collect but very difficult to apply - they include operational metrics that are different from business-oriented metrics. Examples of what not to be measured are - spam emails we received, blocked attempts and virus infection, vulnerabilities in the technical system we have solved, failed login attempts. This is due to the bad habit of "measuring what we can't control."[1].

Subjectively measurement is what people say. "Subjective data is important but can be challenging to comprehend without the full context of an experience."[2] This data concern reliability and validity [3] of information and can affect the final result. Inconsistently measured is based on the information theory and is discussible as "there are now a number of logics developed to draw non-trivial conclusions from an inconsistent knowledgebase"[4]. Even the speed of cybersecurity incidents doesn't give a lot of room for mistakes. Dear to collect (data) is a challenge, and the collected data can be unrelated to cybersecurity or business goals of the organisation. The data that we collect may not be "indicators" at all, that can provide the information for vulnerabilities. The other aspect is that the model that we use for data collection may not be designed for our organisation. The implementation of benchmarking is risky.

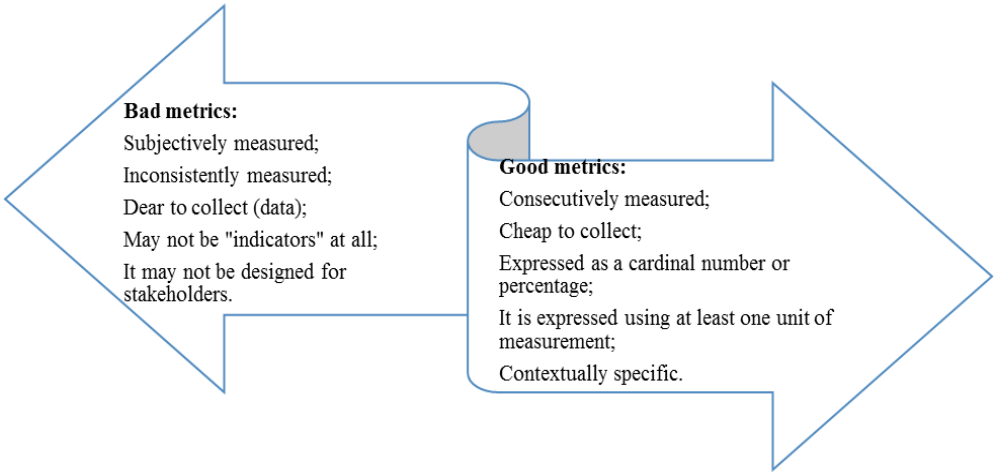


Fig. 3. Comparison between the characteristics of "bad" and "good" metrics.

The following steps can be taken to improve or properly select metrics:

- Getting to know the goals of stakeholders;
- Creating a common policy between business and security;
- The role of Chief Information Security Officer should be business security-oriented;
- The indicators for assessment should reflect the role and benefit of security in an organization.

The choice and application of good metrics to justify the investment are influenced by several factors. It is difficult for a business to understand the benefit to stakeholders that investment in cybersecurity provides. To achieve this goal of mutual understanding the metrics have to be consecutively measured, cheap to collect and to be able to be shown as number or percentage. Cybersecurity policy should be in a business context so the documents that are used have to be standardized. The high-level managers have many obligations so the messages concerning the security have to be connected to the protection of critical information. On the other side, it's very hard for an organisation to limit the risks if the information is not presented in business language. For a better approach, there have to be applied scenarios and case studies.

The critical indicators that have to be collected are:

1) Attacks and incidents -the number of times we were "attacked", number of incidents that harm the critical or sensitive information, number of incidents that caused financial damage;

2) Vulnerabilities - number of unauthorized vulnerabilities, number of critical vulnerabilities attached to critical systems, number of critical systems that cannot be supported;

3) Time - time required to patch critical systems with critical patches, the time required to patch support systems.

3. Some metrics that are often used

One of the most popular metrics are metrics collected through risk analysis. Information security risk management is the overall process that integrates the identification and analysis of the risks to which the organization is exposed, the assessment of potential business impacts and the decision of what actions can be taken to eliminate or reduce the risk to an acceptable level. Comprehensive identification and assessment of the organization's information assets,

the consequences of security incidents, the likelihood of a successful attack on ICT systems and the business costs and benefits of security investment is required. The security risk management applied by an organization usually consists of (Figure 4) [5]:

- identification of business assets;
- identification of threats and assessment of the damage that may be caused by a successful attack;
- security vulnerabilities of the systems that the attack may exploit;
- security risk assessment;
- risk minimization measures with appropriate controls;
- monitoring the effectiveness of the applied controls.

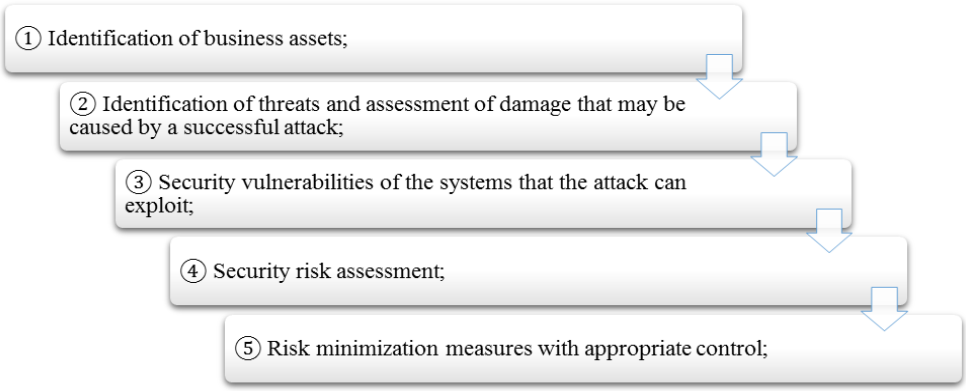


Fig. 4. Risk management

One way risk managers can take a proactive approach to forecast and mitigating cyber risk is to focus on indicators that will help identify their organization's level of exposure and progress in implementation of key programs and cyber risk management.[6]

Some types of indicators that can be considered as part of a cyber risk assessment program and they are divided into the following 5 categories:[6]

The level of threat

Threat level indicators aim to measure how much the organization is exposed to the threat of cyber risk. They can be from both external and internal sources.

The level of risk

Risk level indicators track the organization's current position concerning its cyber risks. This implies that the organization will maintain a risk register for all risks associated with cyber risk, for example:

- Risk of disclosure of user data
- Loss of strategic information

The level of compliance/compliance with rules and procedures

In practice, one of the main drivers of cybersecurity is compliance with established rules and procedures. It is essential to ensure that specific rules, policies and principles are properly maintained and observed to ensure that cyber risks are effectively controlled.

Incident level / Frequency and nature of incidents

Incidents do happen and it is essential to respond to and learn from them. Security incident analysis often provides an additional perspective on risk levels for business stakeholders, allowing them to reassess their current risk assessments.

It is useful for managers to keep track of this information, for example, taking into account whether these projects are on schedule or a budget.

The level of awareness

It is important to have a good broad understanding of the role and importance of staff regarding the threat posed by cyber risk. If employees understand the risks as well as the basic precautions that need to be taken to prevent violations, then a cyber-risk event is much less likely to occur. However, measuring awareness can be a challenge and is something that can best be defined by mixing better assessments.

In short, risk management comes down to a series of decisions and the implementation of those decisions. The main reason for generating indicators is precisely to generate informed risk management solutions. In the world of metrics, the "Goal, Question, Metric (indicator)" method (GQM) (Figure 5) is often used to develop good indicators. This approach helps experts to be focused on the indicator and understand its value.[7]

When using the GQM method to make indicators significant, it should be noted that, in particular, indicators are the main means of comparison, for example:[7]

Current conditions and desired future conditions

- Risk scenarios (prioritization)
- Mitigation options (selection)

Past conditions and current conditions (efficiency of past decisions and actions).

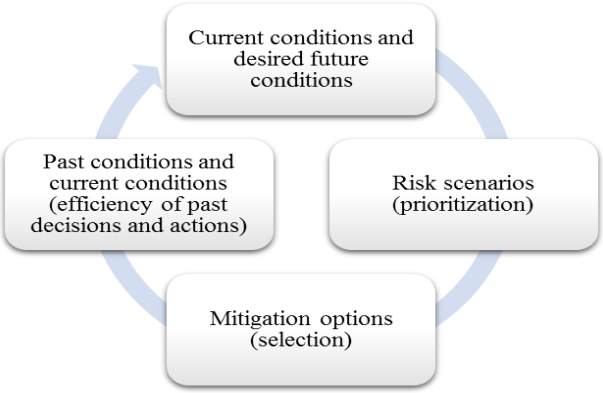


Fig. 5. GQM Cycle

Another important issue is related to the characteristics of security indicators that must be taken into account when formulating and using them. The features are the following:

1. Security indicators are often qualitative rather than quantitative.
2. Security indicators are often subjective rather than objective.
3. Indicators are often defined without a formal model as a basis. Security requirements are determined based on a security model. Security indicators are obtained either by static analysis or by dynamic testing based on specific security requirements and evidence of reliability.
4. There is an absence of time aspect related to the current definitions of the security indicator.
5. Traditional logic with two values not suitable for security analysis - The statement in any traditional logic system is either true or false. When applied to security performance analysis, these logic systems are fragile: even small changes to the system can lead to opposite values of the security specification truth.

At its core, measurement is at the heart of science and engineering.

Indicators (metrics) are quantitative measurements that are important for assessing the effectiveness of the proposed improvements in engineering, regardless of whether they are technologically or process-oriented. Indicators play an equally important role in risk analysis, planning, forecasting, resource allocation and cost estimation.

At its core, measurement is at the heart of science and engineering.

Indicators (metrics) are quantitative measurements that are important for assessing the effectiveness of the proposed improvements in engineering, regardless of whether they are technologically or process-oriented. Indicators play an equally important role in risk analysis, planning, forecasting, resource allocation and cost estimation.

The measurement involves three main steps: (1) data collection, (2) data validation, and (3) data processing. Data collection consists of defining what to collect and how to collect data. The type of data to be collected is directly related to the type of behaviour to be analyzed and the quantitative measures to be evaluated to characterize such behaviour. Data validation consists of an analysis of the collected data for correctness, consistency and completeness. Data processing consists of performing statistical analysis of validated data to identify and analyze trends and assess quantitative measures characterizing security [8].

4. Software quality metrics [9]

Adequacy - shows whether the software provides the necessary functionality, whether all the specified functionality is implemented and whether end-users can perform their work.

Integrity - shows whether the software provides the correct and agreed to results.

Interoperability - indicates whether the software can interact with the designated systems and whether the designated interfaces are correct and implemented.

Correctness - shows whether both legal requirements and corporate standards and regulations are met.

Security - Indicates whether the software can prevent unauthorized access to users and systems. The more secure is software the less vulnerable it is to cyberattacks, and users are protected from breaches and service interruptions due to malicious actions or human errors.

Portability - shows how well the software is suitable for transfer from one environment to another. The software must be able to run on multiple operating systems and in different environments (for example, locally or in the cloud), while dependencies on other software must be minimized and well defined.

Efficiency - shows the time required and the utilization of operational resources (processor, memory, hard disk, etc.) to perform a task.

Reliability - shows the ability of the system to perform the agreed level of performance under certain conditions and time frame. The more reliable is software the fewer bugs users will encounter.

Usability - shows criteria related to user interactions with a given software. The software must be prototyped and tested for usability, and then constantly monitored for potential problems.

Support - shows how easy it is to make changes, to adapt the software for use by other teams, for new purposes or new requirements. The more complex is the code, the harder it is to maintain. The higher the quality of code, the easier it is to work with and the easier it can be adapted to new requirements [10].

5. Economic metrics for security assessment

In addition to the traditional NPV and IRR that we will not cover in this article, there is Annualized loss expectancy. This is the expected loss of a certain asset for more than one year. This indicator is useful in terms of whether a business solution is profitable.

Used in cost-benefit analysis when calculating the profitability of a new project.

The annual duration of losses can be calculated by multiplying the annual occurrence rate (ARO) by the Single-loss expectancy (SLE) [11]. SLE is the expected monetary loss each time a risk arises. This is the term most commonly used during the risk assessment and tries to put a monetary value on each threat.

The equation for the expected duration of a loss is:

$$SLE = AV * EF$$

Asset Value (AV) is the value of a share determined on a specific date or time. The Exposure factor (EF) is measured as a percentage and calculates the potential loss that could occur to an asset if a specific threat is realized [12].

Total Cost of Ownership (TCO). This is the purchase price of an asset plus operating costs. Estimating the total cost of ownership is a greater idea of what a product is and what its value is over time. When choosing from alternatives to a purchase decision, buyers must look not only at the short-term price of an item, known as its purchase price but also at its long-term price, which is the total value of the property. The element with lower total property costs is of better value in the long run.[13] Return on security investments (ROSI) is a key indicator is essential in justifying the commitment of costs and budgets for those companies that regularly deal with security (security departments, CERT, etc.). The valuation of a security investment includes an estimate of the potential loss that can be saved by the investment. Therefore, the monetary value of the investment must be compared with the monetary value of the risk reduction. This monetary value of the risk can be estimated by a quantitative risk assessment – ROSI [14].

Access to information from applications is very important. Packet transfer needs to be explored to make efforts to add security to improve the confidentiality and integrity of the data. In terms of network and access devices have means access to a packet or set of packets in different ways and places using different amounts of effort. For example, to reduce delays during data transfer, local copies or cache values of these packets can also be created from there - written data can be retrieved not only from the network or during transfer but also from a local device. To further illustrate, a sensor connects to a server to identify and authenticate user A and retrieve its permissions. During this process, user B may intercept the packet at various points on the network or the device and receive a set of information from user A.

Information tracking

It is important to have an interoperable and interconnected environment so that the systems can interact with each other. It is also extremely important that, for example, the information used by system B and which was created by system A cannot be returned to its origin. This means that although system A has provided a set of information to B, a user of system B should not be aware that this information is from another part or user.

Tracking citizens

Smart solutions use different sensors (physical or social); these sensors are used to collect data from several urban systems and based on these urban systems may have better city management. To avoid further problems, such sensors must be under the control of a responsible entity to preserve the integrity of its functionality and the data generated. Among the possible problems raised by this feature is that it may be detected or subject to unauthorized

tracking of citizens, detection of traffic patterns and may cause "flooding" of targeted advertising / merchandising.

Loss of user/citizen data

Smart systems in the context of smart cities can use devices such as smartphones, tablets and other gadgets to collect a wide range of data and information. Depending on the type of data processed by such devices, there may be personal and sensitive data, such as messages, photos, appointments, bank account numbers, contact details, and more. This issue deals with the concept that applications store valuable data in the device and if not well processed, this data can be lost or compromised, creating significant problems for the citizens.

Cross-access to information in data centres

For this scenario, we deal with situations involving unauthorized access to information, using flaws on the server. If data security is compromised in any way, for example, while it is under storage, analysis and management procedures, the entire system may be compromised.

For example, when access to information related to students' education systems is available, an application can recover criminal records from a non-specific relationship related to that city, although the solution should only be used with educational services. This situation can arise because both systems share a common area or permissions that must be followed to avoid such behaviour.

6. Conclusions

• Choosing the right metrics when measuring cybersecurity is critical to system security;

• Choosing bad metrics can lead to concentrating resources in the wrong point of defence;

• The first step in building an effective cybersecurity system is to align security objectives with the organization's business objectives.

• It is important to have a good broad understanding of the role and importance of staff regarding the threat posed by cyber risk

• One of the most popular metrics are metrics collected through risk analysis

References

1. B. Smith, ‘Measuring Security - How Do I Know What a Valid Metric Looks Like?’, https://its.ny.gov/sites/default/files/documents/ben_smith_6.9_version.pdf.
2. Cognitive FX Team ‘Objective and Subjective Measurements in Our Research’. https://www.cognitivefxusa.com/blog/objective-and-subjective-measurements-in-our-research
3. Elasy T. A., G. Gaddy, ‘Measuring Subjective Outcomes’, J. Gen. Intern. Med., vol. 13, no. 11, pp. 757–761, Nov. 1998, doi: 10.1046/j.1525-1497.1998.00228.x.
4. Hunter A., S. Konieczny, ‘Approaches to Measuring Inconsistent Information’, in Inconsistency Tolerance, L. Bertossi, A. Hunter, T. Schaub, Eds. Berlin, Heidelberg: Springer, 2005, pp. 191–236.
5. Bojanc R., B. Jerman-Blažič, ‘An economic modelling approach to information security risk management’, Int. J. Inf. Manag., vol. 28, no. 5, pp. 413–422, Oct. 2008, doi: 10.1016/j.ijinfomgt.2008.02.002.
6. ‘White-paper - 6 Key Risk Management Metrics for Controlling Cyber Security’. Xactium Ltd, https://cdn2.hubspot.net/hubfs/91381/2019%20Collateral/White%20Papers/Controlling%20Cyber%20Security/6%20Key%20Risk%20Management%20Metrics%20for%20Controlling%20Cyber%20Security%20.pdf.
7. Freund J., J. Jones, ‘Information Security Metrics’, in Measuring and Managing Information Risk, Elsevier, 2015, pp. 293–333.

8. Wang A. J. A., 'Information security models and metrics', in Proceedings of the 43rd annual Southeast regional conference - Volume 2, Kennesaw, Georgia, Mar. 2005, pp. 178–184, doi: 10.1145/1167253.1167295.
9. 'ISTQB: Software Quality Metrics | FAS-Consulting'. https://cloud.fas-consulting.de/drupal/test-management-istqb/istqb-testing-activity/istqb-software-quality-metrics.
10. 'Software Quality Measurement: 8 Dimensions of Quality', Sealights. https://www.sealights.io/software-quality/software-quality-measurement-8-dimensions-of-quality-and-how-to-measure-them/.
11. 'What is Annualized loss expectancy'. https://capital.com/annualized-loss-expectancy-definition.
12. 'What is Single-loss expectancy'. https://capital.com/single-loss-expectancy-definition.
13. Twin A., 'Total Cost of Ownership (TCO): Asset Costs Over the Long Term', Investopedia. https://www.investopedia.com/terms/t/totalcostofownership.asp
14. 'Return On Security Investment (ROSI) A Practical Quantitative Model', Proceedings of the 3rd International Workshop on Security in Information Systems, Miami, USA, 2005, pp. 239–252, doi:10.5220/0002580202390252.

Cyber Threats for Modern Economy

Georgi Radev

University of National and World Economy, Sofia, Bulgaria
radev1988@gmail.com

Abstract. The paper reviews different types of cybercrimes and evaluate their effect on the modern economy for both – private and public sector. It's showing trends of development for main types of cybercrimes segregated by sector. It's considered a segregation of the types of effect that are caused by cybercrimes. The research is based on official data from independent sources and on multiple examples that are worldwide recognized. Promising directions for future research are outlined.

Keywords. Cyber threats, Cybercrimes, Economy, Cost, Corporate

1. Introduction

In our modern society, the technology is shaping daily life more and more. It's bringing new opportunities, but in the same time entirely new challenges and threads are lying on the horizon. The world economy is based and in high dependence on the technology progress. Almost all sectors are containing digital assets with high value and some are completely based on those assets – like software development business. The financial cost caused by cyberattacks is growing every year and it's causing huge damage to private corporations and whole countries. The need of identifying those threats is more recognizable than ever. Private and public sector are combining effort to stand together against the cybercrimes. To be able to create and organize reasonable cybersecurity strategy the people and organizations engaged with this initiative first need to identify and evaluate the threads for the economy. Estimating the cost caused by cybercrimes is also a difficult task that should be taken from those organization. Based on it the companies and countries are estimating what amount of their budget could be spend for cybersecurity.

When we are talking about threats in the digital world it's considered, as a malicious act that seeks to damage data, steal data, or disrupt digital life in general. Cyber-attacks include threats like computer viruses, data breaches, and Denial of Service (DoS) attacks. Those attacks could be focused against a single user, but also companies or countries. Millions of people could be affected by a single attack organized and executed by few people, or just one person. In 2014, Yahoo! announced it had suffered a cyber-attack, which affected 500 million user accounts constituting the largest massive hacking of individual data directed against a single company. Names, dates of birth, telephone numbers and passwords were stolen. While the company assured users that banking data had not been affected, it nonetheless recommended caution. Prior to this event, in 2012, the hacker "Peace" had sold 200 million usernames and passwords for $1900. Another case for the last years is a marketing analytics firm left an unsecured database online that publicly exposed sensitive information for about 123 million U.S. households. The data included 248 fields of information for each household, ranging from addresses and income to ethnicity and personal interests. Details included contact information, mortgage ownership, financial histories and whether a household contained a dog or cat enthusiast. The private sector is not the only target of the cyber criminals. Whole countries and their interests were attacked either. A recent example is the proved attack again

Bulgarian revenue agency – NAP, where the personal data of millions of Bulgarian citizens and companies have leaked from the agency. All those examples are showing inability of handling and responding to the digital threads in the modern world. In some scenarios the financial or economical aspect of the cyberattack is recognizable and the damage is clear. In other cases the financial damage is huge and it's impossible to be evaluate completed. If we take the example with Yahoo!, if we put aside the all immediate effects, we have a number of side effects which will continue their negative trends for years. Loosing trust in the company will force many of their clients to look for alternatives. This will cause withdrawing of partners or less income from advertising. In August of 2016, Cybersecurity Ventures predicted that cybercrime will cost the world $6 trillion annually by 2021, up from $3 trillion in 2015. This represents the greatest transfer of economic wealth in history, risks the incentives for innovation and investment, and will be more profitable than the global trade of all major illegal drugs combined. This is the reason for fast developing of cybersecurity, as an independent and vital industry. The number of cybersecurity job postings has grown 94% in just six years. That's a 300% increase in demand compared to the overall IT job market. Cybersecurity jobs in 2019 account for 13% of all information technology jobs.

2. Main types of cyberattacks

In next we will review and classify with more details the current cyber threads that are challenging our modern society. According Jeff Melnick (Director, Global Solutions Engineering at Netwrix) "A cyber-attack is any type of offensive action that targets computer information systems, infrastructures, computer networks or personal computer devices, using various methods to steal, alter or destroy data or information systems.". There are hundreds existing types of cyberattacks, but we will review the six most common types, that are considering more than 90% of all cyber-attacks in world.

Denial-of-service (DoS) is a situation where perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting services of a host connected to the Internet. Denial of service is typically accomplished by flooding the targeted machine or resource with superfluous requests in an attempt to overload systems and prevent some or all legitimate requests from being fulfilled. A DoS or DDoS attack is analogous to a group of people crowding the entry door of a shop, making it hard for legitimate customers to enter, thus disrupting trade. Unlike attacks that are designed to enable the attacker to gain or increase access, denial-of-service doesn't provide direct benefits for attackers. For some of them, it's enough to have the satisfaction of service denial. However, if the attacked resource belongs to a business competitor, then the benefit to the attacker may be real enough. Another purpose of a DoS attack can be to take a system offline so that a different kind of attack can be launched. According Department of Homeland Security of USA those types of attacks can't be completely avoided, but the damage could be reduced with installing traffic monitoring and regulations.

Phishing is the most common attack that is affecting single users. This is a type of social engineering attack often used to steal user data, including login credentials and credit card numbers. Typically carried out by email spoofing or instant messaging, it often directs users to enter personal information at a fake website which matches the look and feel of the legitimate site. An attack can have devastating results. For individuals, this includes unauthorized purchases, stealing of funds, or identify theft. Moreover, phishing is often used to gain a foothold in corporate or governmental networks as a part of a larger attack. According APWG (Anti-Phishing Working Group) the world is experiencing more than 1 million unique phishing attacks for 2019. Phishing attacks in 2015 caused 162 million dollars direct losses, but according Microsoft report the total losses are five times higher. Attempts to deal with

phishing incidents include legislation, user training, public awareness, and technical security measures. For users, vigilance is key. A spoofed message often contains subtle mistakes that expose its true identity. These can include spelling mistakes or changes to domain names that can be observed in the URL. Users should also consider why they're even receiving such an email, at a first place. For enterprises the most effective measures are two factor authentication, strict password management policies and regular education campaigns. Those measures should be implemented and used together for maximum efficiency.

Man-in-the-middle (MitM) attack occurs when a hacker inserts itself between the communications of a client and a server. It is a situation when the attacker is able to intercept all messages and send different one to the both communication points. The strategy itself is not new and was developed in the past, much before the digital world. The difference is that with the digital technologies it could be with much bigger scale. The most notable case is when documents leaked from National Security Agency of US which indicated that the agency appears to have put on a Google suit to gather intelligence. Having a strong encryption mechanism on wireless access points prevents unwanted users from joining targeted network and it's the best way of prevention for those type of attacks.

SQL injection attack has become a common issue with database-driven websites. It is a web security vulnerability that allows an attacker to interfere with the queries that an application makes to its database. It allows attackers to spoof identity, tamper with existing data, cause repudiation issues such as voiding transactions or changing balances, allow the complete disclosure of all data on the system, destroy the data or make it otherwise unavailable, and become administrators of the database server. According Imperva Web Application attack report from 2013, it was observed that the average web application received 4 attack campaigns per month, and retailers received twice as many attacks as other industries. The most famous SQL injection attack was performed against one of the largest conglomerate in the world – Sonny. The company confirmed that 77 million users' data was compromised. Other victims of such attacks during the last decade are British Royal Navy, United States Department of Justice and Yahoo. There is a whole set of preventing measures against sql injections, but most common one are do not using dynamic SQL, avoid leaving sensitive credentials in plain text format, filter out user-provided inputs, provide restriction or limitation in database permissions as well as privileges, use web application Firewalls to protect the database and keep the database systems and versions updated to the latest existing patches.

Malware is another very common type of attack. Any software intentionally designed to cause damage to a computer, server, client, or computer network could be reviewed, as malware. It is consolidating a wide variety of types, including computer viruses, worms, Trojan horses, ransomware, spyware, adware, rogue software, and scareware. Computer viruses are one of the first type of digital attacks in the world. In January 1986, two brothers, Basit and Amjad Farooq Alvi, created "Brain A," the first computer virus to attack the MS-DOS operating system, in Lahore, Pakistan. Today malware is used by both black hat hackers and governments, to steal personal, financial or business information. Malware is sometimes used broadly against government or corporate websites to gather guarded information, or to disrupt their operation in general. However, malware can be used against individuals to gain information such as personal identification numbers or details, bank or credit card numbers, and passwords. There are whole strategies developed by companies and governments against malwares. Anti-virus and anti-malware software are big part of it and they are developing and evolving really fast, providing very good level of protection at a good price. While some of the most spread malwares are in decline last years, like Trojan horses and worms, there are others which are making their pike. Such type is "Ransomware". It affects an infected computer system and demands payment to bring it back to its normal state. In some cases it's

nearly impossible to back the system without a key. The most recent example of this attack is "CovidLock, ransomware, 2020". Fear in relation to the Coronavirus (COVID-19) has been widely exploited by cybercriminals. The software infects victims via malicious files promising to offer more information about the disease. Once installed, CovidLock encrypts data from Android devices and denies data access to the victims. To be granted access, you must pay a ransom of USD 100 per device. It's impossible to calculate the total losses caused by this particular crime at the moment, but it will be huge for sure. As malware is the most complex cyber-attack the best prevention is using a software developed to identify and handle Malwares. There are two main types of software. First is providing a real time protection and the second one can be used solely for detection and removal of malware software that has already been installed onto a computer. Some companies are providing both services, as different modules of one system.

3. Main targets of cyberattacks and different types of losses

Targets of cyberattacks could be single users, private corporations, government institutions or whole countries, as well. The attacks are becoming more sophisticated and possess the ability of spreading in a matter of seconds. The World Economic Forum has established cyber-attacks as a global risk within its report from 2013. This theses is confirmed by multiple scientist including "Computer Science Department of University of Memphis, Tennessee, USA". In 2017, cybercrime cost the global economy $600 billion. In 2018, the global financial damage exceeded $1 trillion, a 50% annual increase. In terms of scale, it is equal to the nominal GDP of Australia. The cost of cyber-attacks is expected to grow exponentially in the years to come. As the digital economy expands, so does the threat landscape. Along with Yahoos' breach, WannaCry is another example of a vast and widespread attack. This was caused by focusing single users using a ransomware attack that infected up to 400 thousand computers in over 150 countries and caused global financial losses of up to $4 billion in 2017. WannaCry exploited the Microsoft Windows vulnerabilities to let the ransomware spread fast across computers that didn't have a security patch. Having infected one machine, WannaCry could find and infect other computers in the same network. The malicious software encrypted data and demanded a ransom of $300 worth in Bitcoin to decrypt the files. The ransomware affected governmental entities, health service providers, banks, energy and global companies in Russia, Spain, UK, India, China, Italy, Ukraine, the USA, South America, and more than one hundred other countries worldwide. The creators and the origin of WannaCry are still unknown. Cyberatacks could be reviewed, as focused against private sector or government institutions.

When were are reviewing cyberattacks against the business we can distinguish three types of losses. First is a direct economic loss like theft of corporate information, theft of financial information, theft of money or disruption to trading (eg inability to carry out transactions online). Businesses that suffered a cyber breach will also generally incur costs associated with repairing affected systems, networks and devices. Another aspect is a reputational damage. This is an essential element of customer relationship. Cyber-attacks can damage business reputation and erode the trust that customers have for a particular company or brand. This, in turn, could potentially lead to loss of customers, sales or profits. The effect of reputational damage can even impact on company suppliers, or affect relationships with partners, investors and other third parties. The last aspect is a legal consequences of cyber breach. It's a vital topic, especially after implementing GDPR policy. Data protection and privacy laws require companies and organizations to manage the security of all personal data they hold - whether on their staff or customers. If this data is accidentally or deliberately compromised, and the company have failed to deploy appropriate security measures, it may

lead to fines and regulatory sanctions. In 2017, cybercrime cost the global economy $600 billion. In 2018, the global financial damage exceeded $1 trillion, a 50% annual increase. A research provided by "Ponemon Institute LLC" and "Accenture" is showing separation by the most affected sectors. Banking and Utilities industries continue to have the highest cost of cybercrime across the research with an increase of 11 percent and 16 percent respectively. Other significant sectors are Software, Automotive, Insurance and High tech industries. The same research is also developing an economic model to evaluate the risk globally over the next five years. They are estimating the expected cost of cybercrime as a percentage of revenue for companies in a range of industries and calculated the total industry revenues and multiplied those figures by the expected cost of cybercrime percentage for that industry. Consolidating these findings across industries globally, they found that the total value at risk from cybercrime is US$5.2 trillion over the next five years. Such huge number with some small deviations is confirmed by other authorized sources, as well.

When we focus on the impact on country level we can distinguish two main types of damages. First is pure economical. In this type of attacks there is not much difference, than attacks against private companies or other organization for taking financial benefits. On the chart below we can see the most affected countries in the world for 2017, according "Statista 2020 Security report" (Fig.1).

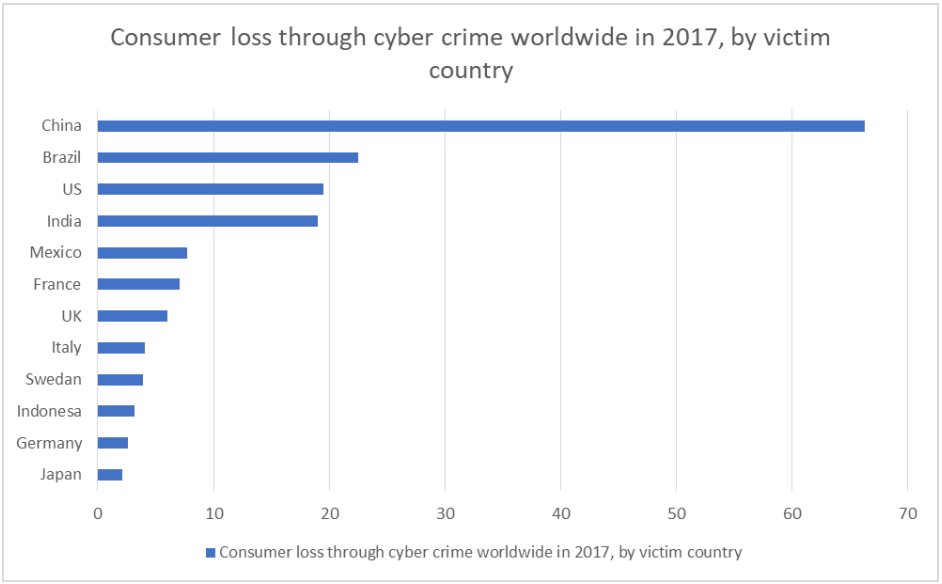


Fig.1.

Second type is political. The main target could be any kind of political manipulation. It could lead to damaging relations between countries or decreasing of prestige of a particular nation. Another reason for the attack could be espionage and sabotage. This approach is not new and in general there is not much differences from the conventional way. The attackers' goal is accessing restricted environment or information that could be used at the right moment for their purposes. Observations are showing more significant development of this type of cybercrimes after 2016. On the table below can be seen a research based on data from Eurostat, International Telecommunication Union and Microsoft. It's calculating index of "cybercrime vulnerability score" for all countries from EU. According it Finland is the least venerable country in EU, following by Estonia and Germany. Bulgaria is in the middle of the table.

4. Conclusion

With growing usage of IT technologies the modern economics are more and more dependent to their digital assets. It's creating new opportunities and new challenges. One of the biggest threads is definitely the cyber-crimes. Their effect is growing every year. Even there is a decrease in some of the business sectors and types of cyberattacks, there are others, which are making significant increase. Such sectors are banking and software. When it comes to types of cyberattacks, such examples are ransomwares and malware. Organizations all over the world are investing more and more resources in protecting their valuable assets. Even there is no organization that is fully protected, there are some good proven practices. Such practice is developing of centralized defensive strategy, specialized to cybercrimes. Such strategy is responsible for evaluating the risk, implementing procedures and steps for prevention, choosing best suitable defensive software, preparing mitigation plans and constantly monitoring for new threads. Using a centralized strategy is assuring that the limited resources for protection will not be spend chaotic, without significant impact. Last few years countries and private organizations are spending more effort on working together against the cyber threads. This is leading to developing of defensive strategy for a whole region or business sector. World Economic Forum is reviewing cyber security, as one of the main topics. On the last session Jürgen Stock, Secretary-General of Interpol is sharing ideas of creating Global Cyber Alliance, which will has vast scale by forming a global public-private cooperation against cybercrime. It could help to meet the current and growing needs. As, it's too early to give a final evaluation of this efforts, still it's looking very promising at the moment and more and more organizations are open for collaborating together against common threads in digital world.

References

1."Cybercrime costs global economy $445 billion a year: report". Reuters. 9 June 2014. Retrieved 17 June 2014.
2. Moore, R. "Cybercrime: Investigating High-Technology Computer Crime," Cleveland, Mississippi: Anderson Publishing. 2005
3. Schatz, Daniel; Bashroush, Rabih; Wall, Julie. "Towards a More Representative Definition of Cyber Security". Journal of Digital Forensics, Security and Law. , 2017
4. Steve Morgan, "Cyber Crime Costs Projected To Reach $2 Trillion by 2019". Forbes. 2017
5. Lepofsky, Ron. "Cyberextortion by Denial-of-Service Attack". 6 July 2011.
6. Ponemon Institute LLC cybersecurity report, 2018
7. Jakub Krousteki. "Avast reports on WanaCrypt0r 2.0 ransomware that infected NHS and Telefonica". Avast Security News. Avast Software, Inc., 12 May 2017
8. "World Economic Forum: Cybersecurity", 2019
9. Ping Wang, Sun-A Park, "Communication in cybersecurity: a public communication model for business data breach incident handling", Robert Morris University 2017
10. Morgan, S. "Cybercrime Damages $6 Trillion by 2021", 2016
11. Mariwan Ahmed Hama Saeed, "Malware in Computer Systems: Problems and Solutions", International journal of informatics for developments, 2020
12. "Statista Security report", 2020

Analysis of Cloud based ETL in the Era of IoT and Big Data

Geno Stefanov

University of National and World Economy, Sofia, Bulgaria
genostefanov@unwe.bg

Abstract. In the era of IoT and Big Data, data volumes and complexity are increasing constantly. Furthermore, data comes in a variety of sources and formats. In this scenario, the success of companies relies on the ability to process, analyze and gain knowledge of data, faster and in a better way. ETL (Extract Transform and Load) tools play a significant role in this process, but it requires novel approaches in order to meet the new requirements in the IoT and Big Data era. The paper aims at analyzing the Cloud based ETL tools and their role in IoT and Big Data environment.

Keywords. ETL, IoT, Big Data, Cloud ETL.

1. Introduction

Big Data and Internet of Things (IoT) are one of the most promising technologies in our time. We can see IoT applications everywhere from our homes to various industries and environments. The rapid development of the IoT generates a huge amount of data and brings many challenges related to managing and using that data in a meaningful way [1]. On the other hand, Big Data technologies are dealing with the management and analysis of such huge amounts of data [2]. Milev [3] states that the increase in the global data volume leads to the need to use modern information technologies for storing and analyze this data. The symbiosis and relationship between IoT and Big Data is obvious. Without the capabilities of Big Data techniques, the IoT data will not bring the knowledge and decision-making information which we all expect.

Business intelligence and data warehousing play a significant role when it comes to data management and analysis [6], [9]. Consolidating the data is often a task for the Extract Transform and Load (ETL) processes. The concept of extracting, transforming and loading data from various heterogeneous sources to a central data repository is relatively old and well-defined in the context of structured and semi-structured data. In the context of IoT and the data it generates, data arrives in enormous amounts and usually in a non-structured format. Along with the rapid development of cloud computing, the cloud-based ETL tools appeared. This has led to a whole new branch of ETL tools – cloud-based ETL tools.

The goal of this paper is to analyze the cloud-based ETL technologies in the context of IoT and Big Data.

The rest of the paper is organized as follows: Section 2 introduces the IoT and Big Data definitions and analysis. In Section 3 an analysis for cloud-based ETL tools is presented and explained. Finally, in section 4, some conclusions and future work are inferred.

2. Internet of Things and Big Data definitions

The British entrepreneur Kevin Ashton proposed the term IoT back in 1999 [1] to identify the importance and the power of connecting RFID tags to the Internet for the Supply Chain Management domain. Since then, many definitions were proposed and there is truly no

universally accepted definition for IoT. The Internet Architecture Board (IAB) defines IoT as a communication service [2]. According to them, the term "Internet of Things" (IoT) represents a set of a large number of embedded devices, which provides communication services based on the Internet Protocols. The Oxford Dictionary gives a very accurate definition and focuses on the use of the Internet as a connecting media between devices [3]. It defines IoT as "the interconnection via the Internet of computing devices embedded in everyday objects, enabling them to send and receive data".

The typical architecture of one IoT solution consists of many devices, which use gateways for data sharing and communication with a back-end server, running IoT platform and are hosted on the cloud or on premise (Figure 1).

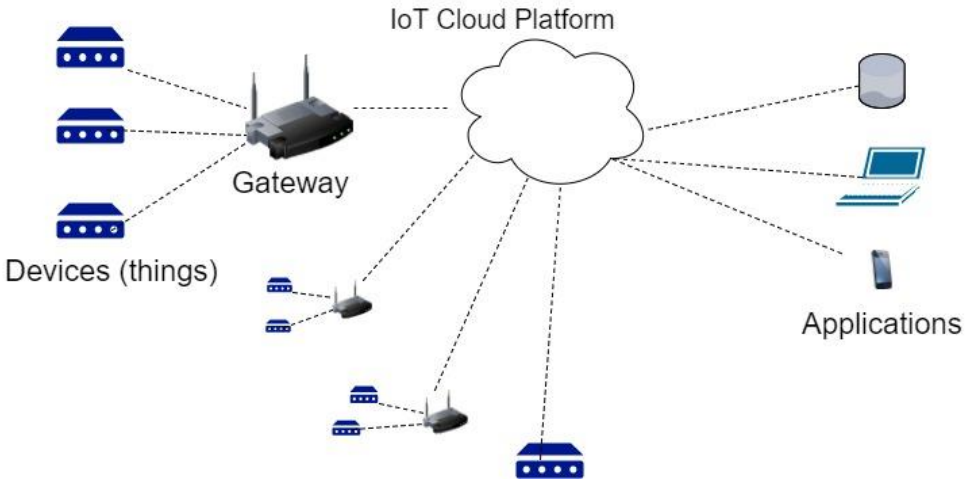


Fig. 1. IoT solution architecture components

According to the authors in [6] Big Data approach comes with respective hardware, instruments and methods developed in order to transform all available data into meaningful information. Big Data has been classified according to five fundamental elements, which are volume (size of data), variety (different types of data from several sources), velocity (data collected in real time), veracity (uncertainty of data) and value (benefits to various industrial and academic fields). Recently, researchers have identified additional Big Data characteristics, such as validity (correct processing of the data), variability (context of data – exact meaning), viscosity (latency data transmission between the source and destination), and virality (speed of the data sent and received from various sources) [10].

3. Cloud-based ETL for Big Data

Prior to the Big Data era, traditional ETL tools were designed to process data sources, such as relational databases, structured and semi-structured data coming from internal information systems or third-party data providers in reasonable data volumes. ETL tools such as Microsoft SSIS, Informatica PowerCenter, IBM DataStage, Pentaho and many others are still very popular and widely used in the Business Intelligence solutions for loading Data Warehouses.

On the other hand, in the IoT and Big Data era, data volumes are huge. Today, around 9 billion IoT devices are connected including about 8 billion mobile phones and home devices [11]. Those devices will generate 79.4 zettabytes (ZB) of data in 2025 [12]. That amount of

data is not suitable for the classical ETL approach. A new approach is required – cloud-based ETL. The main characteristics distinguishing cloud-based ETL from the traditional one are related to:

- Scalability – when data sources, volumes, and complexity increase, scaling and managing this becomes very difficult. The cloud-based ETL scale as the data and complexity grows.
- Maintenance – maintenance is handled automatically, as patches and updates propagate seamlessly and automatically
- Compliance – cloud-based ETL ensure data security and protection.
- All in One Place Simplicity - With cloud-based ETL tools, one tool can be used to manage the entire process, reducing extra layers of dependencies
- Cost effectiveness – pay-per-use.

The Apache Hadoop framework became the road over which cloud-based ETL developed. Hadoop distributes the computing processes, which means that data from divergent sources can be remotely extracted, transformed via a network of computing resources, and then transformed for local analysis. One of the most popular cloud-based ETL tools, Azure Data Factory and Amazon Web Services (AWS) Glue, are based on Apache Hadoop.

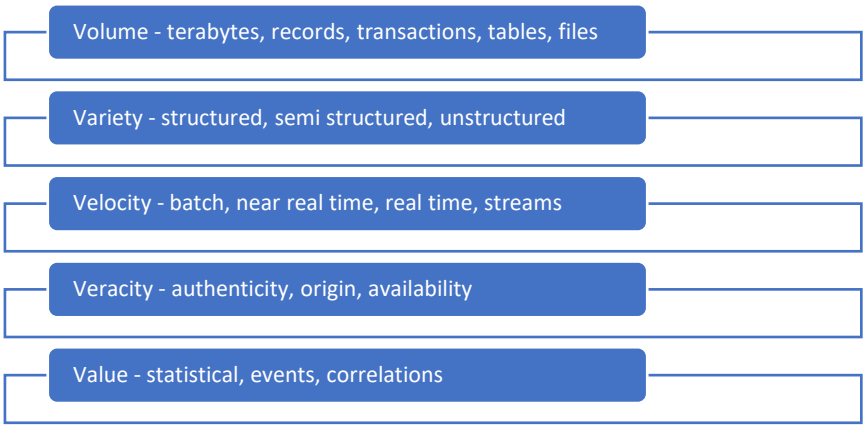


Fig. 2. Big Data Characteristics

Sample cloud ETL workflow process using AWS Glue

AWS Glue is a fully-managed cloud-based ETL tool, that is used to read and load various data source with heterogenous formats. In the context of IoT and Big Data, AWS Glue is quite convenient since it provides event-driven ETL pipelines and automatic execution when new data arrives. It is common for IoT devices to send the generated data to a central storage place in csv, parquet or json formats. AWS S3 is the cloud service used to store and retrieve data. Figure 3 shows a sample ETL process, which consists of the following step:

1. S3 bucket – central place to upload the data files
2. Lambda function – serverless AWS service, which executes programing code
3. AWS Glue Crawler – automatic extraction of source data structure – field and data types.

4. AWS Glue job - the core element in the proposed workflow. Extracts the data from the source, transforms the data, and loads it into target.
5. AWS Glue job – optionally archiving the source files
6. S3 bucket – storing the processed data files

The proposed ETL process does not aim to be complete enough to implement in a real-world scenario. It outlines the main steps for processing IoT data using AWS services.

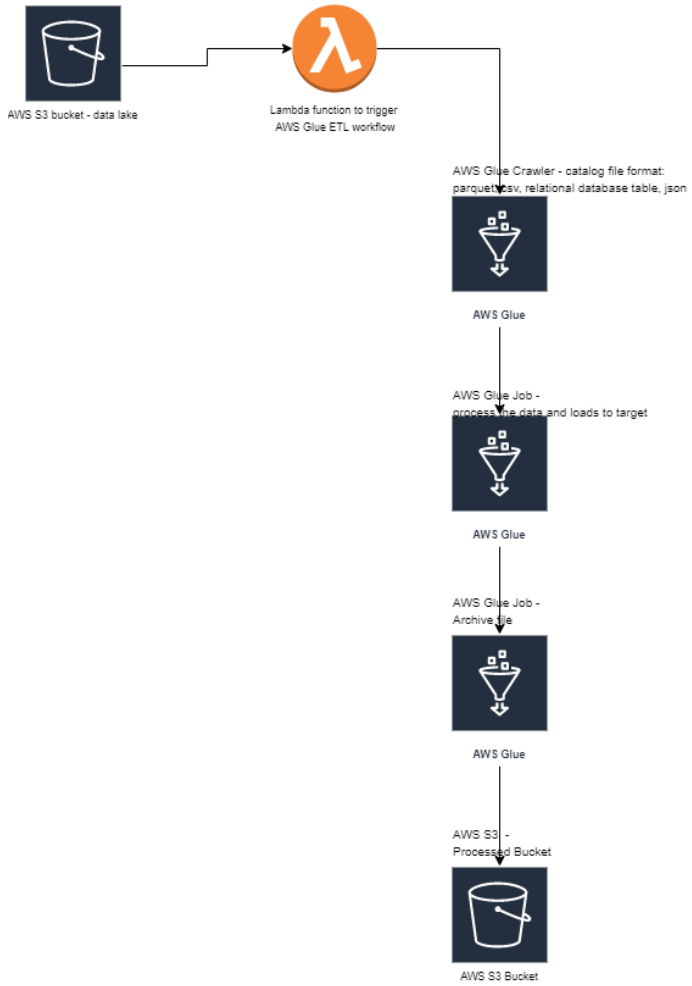


Fig. 3. AWS Glue sample ETL process

4. Conclusion

In the era of IoT and Big Data, data arrives in large volumes, complex format and from a variety of sources. The success of companies hugely depends on how well they can analyze the available data and extract meaningful knowledge from it to attract or retain users. Users, in turn, are becoming more accustomed to personalized services. The Extract-Transform-Load

(ETL) process is essential in accomplishing these goals, but requires significant effort, especially in the Big Data era.

In this paper the main characteristics of cloud-based ETL tools for IoT and Big Data era were briefly explained. A sample ETL process was proposed to manage data generated by IoT devices.

References

1. Boyanov, L., Opportunities and Threats from Implementing Internet of Things, Economic and Social Alternatives, Issue 2, 2018, ISSN 1314-6556.
2. Milev, P., APPLICATION OF REAL-TIME DICTIONARY INSTEAD OF FULL-TEXT INDEX FOR SEARCHING IN WEB PUBLICATIONS, Trakia Journal of Sciences, Vol. 17, Suppl. 1, pp 160-164, ISSN 1313-7069 (print), ISSN 1313-3551 (online), 2019
3. Ashton K., That 'IoT' Thing, in the real world things matter more than ideas, RFID Journal, 1999.
4. RFC 7452: Architectural considerations in smart object networking (March 2015)
5. Internet of Things. Oxford Dictionaries, n.d. Web. 6 Sept. 2015
6. Mihova, V., Methods of Using Business Intelligence Technologies for Dynamic Database Performance Administration, Economic Alternatives, Issue 3, 2015, ISSN 1312-7462.
7. Stefanova, K., Kabakchieva, D., Big Data Approach and Dimensions for Educational Industry, Economic Alternatives, Issue 1, 2019, ISSN 1312-7462.
8. Marzovanova, M., Intelligent Tagging and Search as a Fully Automated System, Economic Alternatives, Issue 2, 2015, ISSN 1312-7462
9. Murdjeva, A, Mihova V, Business intelligent systems for optimization and management of application systems, International conference" Communications, information technologies and statistics. Current issues of theory and practice, Sofia, 2010
10. Mouzhi Ge, Hind Bangui, Barbora Buhnova, Big Data for Internet of Things: A Survey, Future Generation Computer Systems, May 2018
11. Gartner research and Beyond: What's next for IoT, https://www.eseye.com/wp-content/uploads/
12. IDC, https://www.idc.com/getdoc.jsp?containerId=prUS45213219, June 18, 2019

Process Mining Techniques in Business Process Management Systems

Albena Mincheva

University of National and World Economy, Sofia, Bulgaria
a.vicheva@unwe.bg

Abstract. The first wave of BPM software originally aimed to eliminate paper document circulation between relevant parties, and to digitize processes. That was generally where the functionality of BPM software stopped. Process Mining (PM) is something entirely new, with a radically different approach to process optimization. Process mining techniques enable process-centric analytics through automated process discovery, conformance checking, and model enhancement. And since analyses are necessary for optimization of any business process existing in a BPM system, Process Mining is very useful support here. This paper reviews the basic PM techniques that can be used together with the BPM systems in order to achieve process optimization.

Keywords. Process mining, Business process management system, Process optimization, Automated process discovery.

1. Introduction

Nowadays it is all about "Big Data" and here it comes the Data science which involves methods to analyze massive amounts of data and extract the knowledge it contains. It is an interdisciplinary field aiming to turn data into real value. In recent years, data science rises as a new and important discipline. It even can be viewed as a classical discipline like statistics, data mining, databases and distributed systems. Existing approaches need to be combined to turn the available data into value for individuals, organizations, and society. Moreover, new challenges have emerged, not just in terms of size of the "Big Data", but also in terms of the questions to be answered. Data scientists now have the ability to answer questions like:

- (Reporting) What happened?
- (Diagnosis) Why did it happen?
- (Prediction) What will happen?
- (Recommendation) What is the best that can happen?

Data science is an amalgamation of different partially overlapping disciplines. The figure below, shows the main ingredients of data science. The disciplines are overlapping and varying in size. Moreover, the boundaries are not clear-cut and seem to change over time.

In this paper we will focus at the Process Mining as part of the big picture for data science. With that said Process mining techniques could be viewed as data science but based on event data. Process mining techniques use event data to discover processes, check compliance, analyze bottlenecks, compare process variants, and suggest improvements.

Another good point to keep in mind is that most organizations nowadays create business processes which are sometimes difficult to control and comprehend. Understanding these processes is an absolute prerequisite for taking on any improvement initiative. In essence, Process mining is a discipline supported by a powerful analytical techniques which makes it easy and fast to get a complete and objective picture of processes, and as a consequence to control and continuously improve them. This is made possible by analyzing vast quantities of

data available in IT systems. Mainly which activities are performed, when, and by whom. In that sense, process mining sits at the intersection of Business Process Management (BPM) and Data Mining.

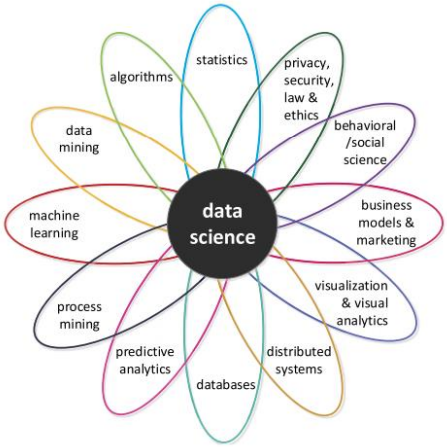


Fig. 1. Components of Data Science [1]

Process mining is a highly valuable solution for insights generation. Wil van der Aalst (Wil van der Aalst is a Dutch computer scientist and full professor at RWTH Aachen University, leading the Process and Data Science group. His research and teaching interests include information systems, workflow management, Petri nets, process mining, specification languages, and simulation.) defines Process Mining as a bridge between traditional process analysis techniques like modeling (which are not based on data) and data-driven techniques like data mining (which are not process-oriented).

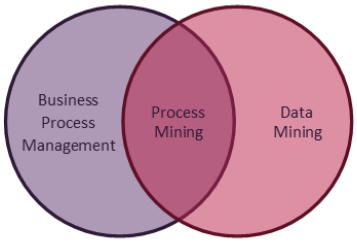


Fig. 2. Intersection of BPM and Data Mining

2. Business Process Management

Business Process Management (BPM) is an important aspect of organizations excellence and global competitiveness. The main indicator of an efficient BPM in place is the level of conformance of its implementation to its original process model.

Business Process Management (BPM) is the discipline that combines approaches for the design, execution, control, measurement and optimization of business processes. But Process mining can be best related to BPM by looking at the so-called BPM life-cycle at Fig.4:

The BPM life-cycle shown above is often used to describe how BPM iterates through multiple phases in designing, implementing, analyzing, and then re-designing the processes. Process mining clearly fits into the analysis/diagnosis phase. While traditional BPM

approaches start with modeling the process, process mining starts by understanding the processes that are already there by discovering the actual processes from data.

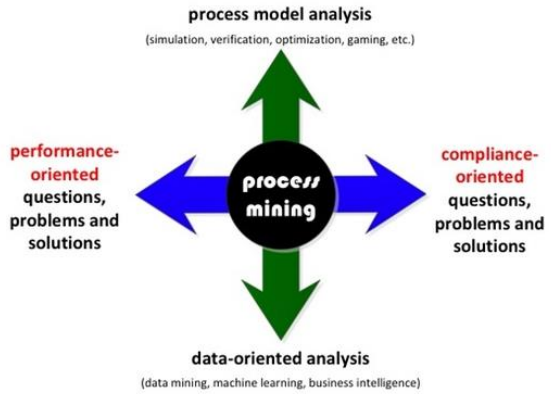


Fig. 3. Process Mining - Illustration by Wil van der Aalst [1]

Initially, the main focus of BPM was on process design and implementation. Process modeling plays a key role in the design/redesign phase and directly contributes to the implementation phase. There is now a clear trend in the BPM community to focus more on the monitoring, adjustment, and diagnosis phases. These phases are more data-driven and process mining techniques are frequently used in this part of the BPM life-cycle.

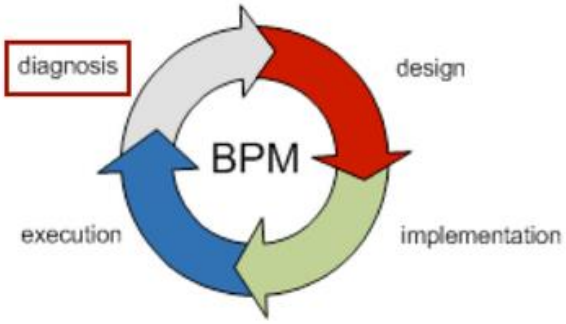


Fig. 4.BPM life-cycle

3. Process Mining definitions and types

3.1. Theoretical definitions

Theoretically, Process mining can be described in various ways. Weijters [10] describes business process mining as a technique that attempts to extract non trivial and useful process information from event logs. Moreover, van der Aalst [10] suggests that the aim of business process mining is to extract information in the form of process models from event logs. According to Lassen [11], the aim of business process mining is to extract information about processes from transaction logs, in order to discover the control flow of a process model. Furthermore, van der Aalst and Weijters [12] stated that business process mining aims to extract information from event logs to capture the business process as it is being executed.

In more practical way, the Process mining can be described as a relatively young research discipline that sits between computational intelligence and data mining on the one hand, and process modeling and analysis on the other hand. The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's information systems.

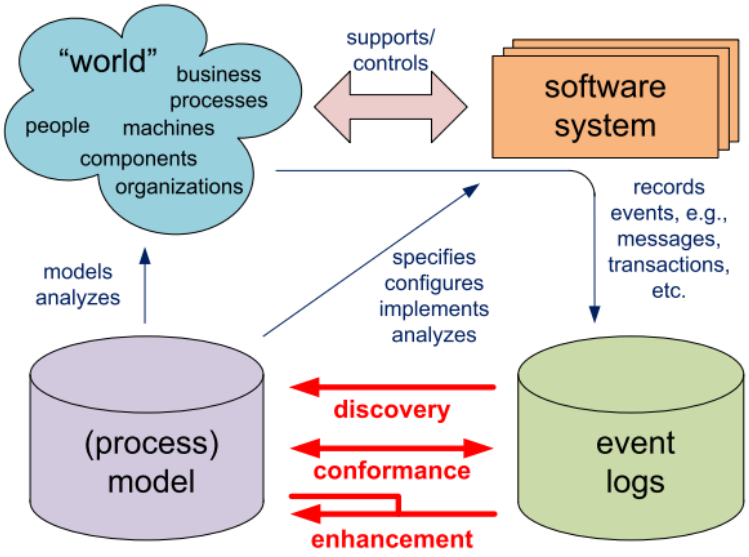


Fig. 5. Positioning of the three main types of process mining: discovery, conformance and enhancement [1]

Process mining includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/organizational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations.

Basically, we can say that Process Mining provides an important bridge between data mining and business process modeling and analysis. The figure above (Fig.5) shows that process mining establishes links between the actual processes and their data on the one hand and process models on the other hand. Today's information systems log enormous amounts of events. Classical workflow management systems, BPM systems, ERP systems and etc. provide detailed information about the activities that have been executed. Figure 5 refers to such data as event logs. Most information systems store such information in unstructured form. The event data is scattered over many tables or needs to be tapped off from subsystems exchanging messages. In such cases, event data exist but some efforts are needed to extract them. Data extraction is an integral part of any process mining effort.

3.2. Process Mining Types

Wil van der Aalst describe three basic types of business process mining:

- **Process Discovery**

No a-priori model exists. Mining is based on event logs without the use of a template of the complete process. Van der Aalst gives the example of their algorithm where a process model can be discovered from an event log without prior knowledge of the process itself.

- **Conformance checking**

The second type of process mining is conformance. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. For instance, there may be a process model indicating that purchase orders of more than one million Euro require two checks. Analysis of the event log will show whether this rule is followed or not. Conformance checking may be used to detect, locate and explain these deviations, and to measure the severity of the process disparity.

- **Enhancement/Extension**

The third type of process mining is enhancement. Here, the idea is to extend or improve an existing process model using information about the actual process recorded in some event log. Whereas conformance checking measures the alignment between model and reality, this type of process mining aims at changing or extending the a-priori model. One type of enhancement is repair, i.e., modifying the model to better reflect reality. For example, if two activities are modeled sequentially but in reality can happen in any order, then the model may be corrected to reflect this. Another type of enhancement is extension, i.e., adding a new perspective to the process model by cross-correlating it with the log. An example is the extension of a process model with performance data. For instance, by using timestamps in the event log of the "request for compensation" process to show bottlenecks, service levels, throughput times, and frequencies.

4. Process Mining Techniques

Process mining techniques give the opportunity to extract knowledge from event logs commonly available in today's information systems. These techniques provide new means to discover, monitor, and improve processes in a variety of application domains. There are two main drivers for the growing interest in process mining. On the one hand, more and more events are being recorded, thus, providing detailed information about the history of processes. On the other hand, there is a need to improve and support business processes in competitive and rapidly changing environments.

Currently, one of the most recent developments in the field of process mining is the rise of commercial tools capable of performing process mining. The Gartner Group [13] already has highlighted Business Process Management Suite vendors who now provide automated methods for process discovery. Their tools implement some of the following process mining techniques:

1. **Transition systems and regions**

This concept of van der Aalst utilizes transition systems and regions to discover process models from event logs. This technique is used to mine models that offer a balance between over fitting (only modelling process transitions observed in the event log) and under fitting (modelling process transitions that may potentially be inferred from the event log) the event log, so providing a mined model that does not over generalize the event log it is based on.

2. **Clustering techniques**

Many techniques highlight the possibility of combining a number of process mining approaches to mine more challenging event logs, such as those that contain noise (event log data that contains incorrect or incomplete information about a process). Most notably, clustering techniques have been used in this way. The work of Greco [14] provided such a technique utilizing the κ-means algorithm to provide hierarchical views of a mined process model. The technique of Greco can utilize the Heuristic approach of Weijters and van der Aalst

[15] to successfully mine noisy event logs. The usage of clustering for process mining has been further investigated in the work of Alves de Mederios [16], addressing the mining of complex 'spaghetti-like' process models. An interesting new direction for clustering is its use with the Fuzzy Process Mining approach of Gunther and van der Aalst [17]. Their work concentrates on the performance analysis of business processes and the identification of bottlenecks within processes.

3. Heuristic approach

One of the most long standing process mining techniques is the heuristic approach of Weijters and van der Aalst [15]. In the work of Weijters and van der Aalst a set of rules are used to determine the precedence between tasks and overall task sequences. Further research in the field of heuristic process mining has been made by Burattin and Sperduti [18] who, with their *Heuristics++* miner, adapt the approach of Weijters and van der Aalst to allow the duration of process activities to become part of the process mining parameter set.

4. Evolutionary techniques

The use of evolutionary techniques for process mining is another new research direction. The approach of Alves de Mederios [10] utilizes a Genetic Algorithm (GA) to mine a variety of event logs containing noise and duplicate tasks. This technique allows for process patterns to be represented as chromosome strings, and successive populations of potential process mining solutions to be generated. Each individual process mining solution in a population is a mined process which is given a score based on how accurately it describes the event log on which it is based. The use of a GA allows for a global search of the event log allowing for the possibility of mining process constructs that span the entire length of the process. The GA approach is suited to mining complex highly parallel processes.

5. Declarative mining approach

The declarative mining approach allows for the introduction of artificially generated negative events. These are events that show that a particular path through a process (or activity recorded in a process trace) is not possible. The process logs rarely contain such negative traces and their introduction can aid the task of process discovery.

6. Conclusion

Sooner or later all the gathered data must be put in use and that's what Process Mining is about. It builds a bridge between big data and business process analytics, resulting in fewer errors throughout the process. Indeed, Process Mining is a loud word in the future of BPM analysis, and we believe it will soon be an indispensable method of keeping the workflow under control.

The presented research was an overview of researcher's findings in the field of business process mining, definition of business process management lifecycle and also overview of process mining techniques and their place in the process model improvement. In a conclusion we could say that process mining techniques have an increasing influence on many BPMS systems. A lot of vendors are putting their efforts to include that kind of modules in their systems and in that way to support their customers throughout the continuous business process optimization journey.

References

1. Wil M. P. van der Aalst, Process Mining, ISBN 978-3-662-49851-4, 2016.
2. Wil M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, 2011.

3. Stefanov, G. Business Process Automation with BPMS, Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE), ISSN 2367-7635 (PRINT) ISSN 2367-7643 (ONLINE), 2014.
4. Business Process Management Workshops, Clermont-Ferrand, France, Revised Selected Papers, Part I, 2011.
5. Wil M. P. van der Aalst, Process Mining Manifesto, 2012.
6. Omar AlShathry, Process Mining as a Business Process Discovery Technique, 2016.
7. Andrea Burattin, Process Mining Techniques in Business Environments, Springer, ISBN 978-3-319-17482-2, 2015.
8. Turner CJ, Tiwari A, Business Process Mining: From Theory to Practice, Business Process Management Journal Vol.18, 2012.
9. Wil M.P. Van der Aalst W. M. P., Supporting the Full BPM Life-Cycle Using Process Mining and Intelligent Redesign, Eindhoven University of Technology, Department of Technology management, 2007.
10. Tiwari, A., Turner, C. J. and Majeed, B., A review of business process mining: State of the art and future trends, Business Process Management Journal, Vol. 14, 2008.
11. Alves de Medeiros, A. K., Guzzo, A., Greco, G., van der Aalst, W. M. P., Weijters, A. J. M. M., van Dongen, B. F. and Sacca, D., Process mining based on Clustering: A quest for precision, Business Process Management Workshops, 2007.
12. Lassen, K.B, van Dongen, B.F, and van der Aalst W.M.P., Translating Message Sequence Charts to other Process Languages Using Process Mining, Beta Working Paper WP207, Eindhoven University of Technology, 2007.
13. Alves de Medeiros, A. K., van Dongen, B. F., van der Aalst, W. M. P. and Weijters, A. J. M. M., Process mining: Extending the a-algorithm to mine short loops, BETA Working Paper Series, WP 113, Eindhoven University of Technology, 2004.
14. Gartner Research, https://www.gartner.com/en/documents/3899484, 2019
15. Greco, G, Guzzo, A, and Pontieri, L., Discovering Expressive Process Models by Clustering Log Traces, IEEE Transactions on Knowledge and Data Engineering, Vol. 18, 2006.
16. Weijters, A.J.M.M, and van der Aalst, W.M.P., Rediscovering workflow models from event-based data using little thumb, Integrated Computer-Aided Engineering, Vol. 10, 2003.
17. Alves de Medeiros, A. K., Guzzo, A., Greco, G., van der Aalst, W. M. P., Weijters, A.J. M. M., van Dongen, B. F. and Sacca, D., Process mining based on Clustering: A quest for precision, in: Business Process Management Workshops, 2007.
18. Christian W. Gunther and Wil M.P. van der Aalst, Fuzzy Mining – Adaptive Process Simplification Based on Multi-Perspective Metrics, Eindhoven University of Technology.
19. Burratin, A. and Sperduti, A., Heuristics miner for time intervals, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2010.

Application of Information Technologies for Analysis and Assessment of Return on Capital

Rositsa Ivanova

University of National and World Economy, Sofia, Bulgaria
rosi_ivanova@abv.bg

Abstract. The issue of efficiency of every enterprise's business is topical at all phases and stages of its development. Business efficiency may be analyzed and assessed in two interrelated aspects. The first of them measures achieved results on the basis of used resources, and the second one – input resources on the basis of results achieved from the enterprise's business. Enterprise's business efficiency may be analyzed and assessed by applying different approaches and models for analysis. These are the accounting approach based on the use of accounting information, the financial approached based on the value-based management, as well as the management approach based on the balanced system of indicators for analysis of the efficiency for achievement of the enterprise's mission, vision and strategy. Different approaches allow analyzing and assessing business efficiency from different perspectives by using the respective metrics and models for analysis. Each of them applies various types of information sources, and the information obtained as a result of the analysis meets the needs of various internal and external users. This publication highlights the return on total capital invested in all enterprise's activities, and its subject matter covers the methodology for analysis of return on capital through the application of the accounting approach. It describes opportunities for using information technologies to analyze the enterprise's return on capital. The aim is to reveal the need to use, the significance and usefulness of the accounting approach for analysis of return on capital.

Keywords. Return on capital, Accounting approach, Methodology, Analysis, Factors, Information technologies

1. Introduction

The return on total capital invested in all activities of the enterprise is a key business indicator. Analysis of return on total capital is made with the data from the enterprise's financial statements. The information contained in these financial statements should meet the needs of both the internal (the management, financial management, internal supervision bodies, etc.) and the external (creditors, investors, suppliers, clients, staff, the Government and its authorities, the public, auditors, etc.) users of understandable, reliable, comparable, neutral and authentic information.

2. Accounting approach for analysis and assessment of return on total capital

Return on total capital is calculated as a percentage ratio between profit (gross or net) and the amount of total capital invested in all activities of the enterprise.

Gross return on capital $\left(\mathrm{PR}^{\mathrm{br}}\right)$ is calculated with the following formula – **model 1**:

$$PR^{br} = \frac{P^{br}}{\bar{C}} \times 100, \quad (1)$$

and the net return (PR^n) – with the following formula – **model 2**:

$$PR^n = \frac{P^n}{\bar{C}} \times 100, \quad (2)$$

where:

P^{br} is the accounting (gross) profit;

P^n – the net (book) profit, and

$\bar{C}$ – the average amount of the total capital.

Two factors have influence on the deviation of the gross, respectively, the net return on total capital for the current in comparison to the reference period – changes of the amount of gross, respectively, the net profit, and the changes of the average amount of total capital of the enterprise. The amount of profit (gross and net) is formed during the reporting period, i.e. it is an interval indicator. For the purposes of achieving comparability among absolute indicators used in the formulas for calculation of the gross and net return on total capital, we should take into account the average amount of capital. It may be calculated as an arithmetic mean or chronologic mean.

Depending on the determined objectives and tasks, and the necessity of information for the needs of the management in order to make decisions for the enterprise's business management, the analysis of return on total capital may be deepened in different aspects.

The formula of the net return on total capital may be transformed as follows - **model 3**:

$$PR^n = \frac{P^n}{\bar{C}} \times 100 = \frac{\frac{P^n}{\bar{I}} \times 100}{\frac{\bar{C}}{\bar{I}}}, \quad (3)$$

where: I is the total amount of enterprise's income,

$\frac{P^n}{\bar{I}} \times 100$ – income-based net return, and

$\frac{\bar{C}}{\bar{I}}$ – capital consumption of income.

This formula may be further extended by differentiating the total invested capital as equity and borrowing. We will then have the following formula – **model 4**:

$$PR^n = \frac{P^n}{\bar{C}} \times 100 = \frac{\frac{P^n}{\bar{I}} \times 100}{\frac{\bar{E}}{\bar{I}} + \frac{\bar{L}}{\bar{I}}} = \frac{\frac{P^n}{\bar{I}} \times 100}{\frac{1}{\frac{\bar{I}}{\bar{E}}} + \frac{1}{\frac{\bar{I}}{\bar{L}}}} \quad (4)$$

Formula – model 4 shows that the following factors have affected the dynamics of the net return on total capital:

1. Changes of the income-based net return ($\frac{P^n}{\bar{I}} \times 100$).

2. Changes of capital consumption of income on the basis of equity ($\frac{\bar{E}}{\bar{I}}$) – changes of the amount of equity invested in the company's business for the purposes of gaining income of one lev. The reciprocal indicator – equity load ratio ($\frac{\bar{I}}{\bar{E}}$), describes the amount of income gained with equity of one lev.

3. Changes of capital consumption of income on the basis of borrowings ($\frac{\bar{L}}{\bar{I}}$) – changes of the amount of borrowings invested for generating income of one lev. The reciprocal indicator – borrowing load ratio ($\frac{\bar{I}}{\bar{L}}$), describes the amount of income gained with borrowings of one lev of the enterprise's capital.

Model 4 may be presented on Fig. 1.

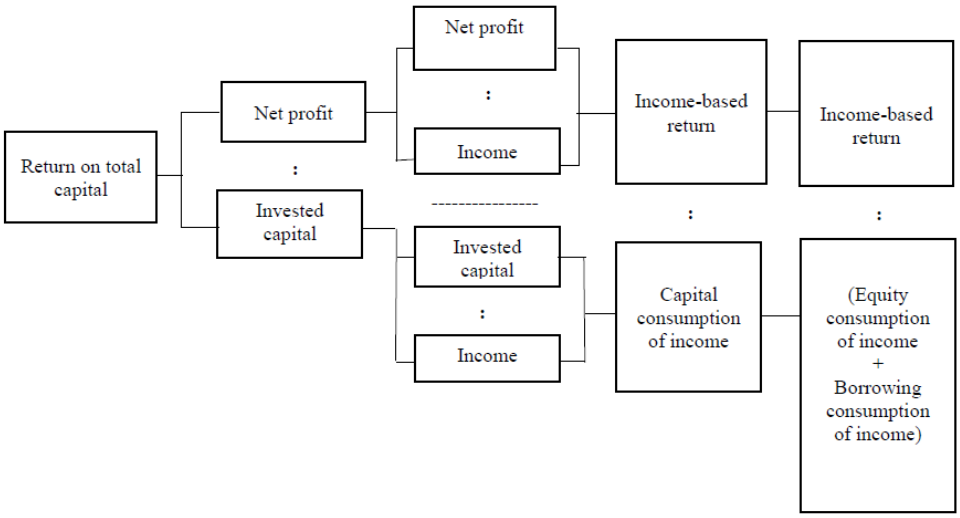


Fig. 1. Extended model for analysis of return on total capital

Depending on the objectives and tasks of the analysis, the numerator of formula – model 4 may be transformed, and for this purpose we can use the objectively existing dependence between the costs per income of BGN 100 (efficiency of income) and the income-based net return. In general, the difference between 100 and the indicator for costs per income of BGN 100 is the income-based net return.

The above statement allows to extend formula – model 4 in accordance with the following formula – **model 5**:

$$PR^n = \frac{100 - \frac{Cost}{I} \times 100}{\frac{\overline{E^{fc}}}{I} + \frac{\overline{E^{wc}}}{I} + \frac{\overline{L^{lt}}}{I} + \frac{\overline{L^{st}}}{I}} = \frac{100 - \frac{Cost}{I} \times 100}{\frac{1}{\frac{I}{\overline{E^{fc}}}} + \frac{1}{\frac{I}{\overline{E^{wc}}}} + \frac{1}{\frac{I}{\overline{L^{lt}}}} + \frac{1}{\frac{I}{\overline{L^{st}}}}} \quad (5)$$

where:

Cost is the total amount of costs;

I – total amount of income.

This formula shows that the following factors have influenced the changes occurred in the net return on total capital:

1. Changes of income-based net return $(100 - \frac{Cost}{I} \times 100)$ – income type efficiency.
2. Changes of absorption ratio $(\frac{\overline{E^{fc}}}{I})$, respectively, load ratio $(\frac{I}{\overline{E^{fc}}})$ of basic equity.
3. Changes of employment ratio $(\frac{\overline{E^{wc}}}{I})$, respectively, turnover ratio of working equity $(\frac{I}{\overline{E^{wc}}})$.
4. Changes of absorption ratio $(\frac{\overline{L^{lt}}}{I})$, respectively, load ratio of long-term borrowings $(\frac{I}{\overline{L^{lt}}})$.
5. Changes of employment ratio $(\frac{\overline{L^{st}}}{I})$, respectively, turnover ratio of short-term borrowings, i.e. working borrowings $(\frac{I}{\overline{L^{st}}})$.

It is possible to take the information from the enterprise's financial statements directly in electronic form and in real time by means of the respective application, and to arrange it in

appropriate tables for the purposes of the analysis. By means of information technologies, this information may be processed in operational manner thus enabling the users to make timely decisions with regard to the enterprise. On the basis of these data we can calculate the indicators for analysis and assessment of return on total capital and to arrange them in the same tables.

We will illustrate the methodology for analysis of return on total capital in accordance with the financial statement data of an industrial enterprise. The information necessary for the analysis is summarized in table 1. Based on the input information, key business indicators in the table are classified in the respective groups: indicators for analysis of the efficiency of income and cost; indicators for analysis of net income-based return; indicators for analysis of the capital turnover – in general and by types of capital; indicators for analysis of return on capital, calculated with the help of various models.

Table 1

Indicators	Previous year	Current year	Change	Dynamics percentage
I. Output data, thousand BGN:				
1. From the profit and loss account:				
A. Total revenue	59625	61215	1590	102,67
B. Total cost	45820	46470	650	101,42
C. Accounting (gross) profit	15339	16383	1044	106,81
D. Balance (net) profit	13805	14745	940	106,81
2. From the balance sheet:				
A. Capital invested, including:	159000	162900	3900	102,45
a. equity	89700	95700	6000	106,69
b. attracted capital (liabilities), incl.	69300	67200	-2100	96,97
b1. long - term liabilities	37200	35700	-1500	95,97
b2. short - term liabilities	32100	31500	-600	98,13
B. Assets, including:	159000	162900	3900	102,45
a. fixed assets	100125	102375	2250	102,25
b. short-term assets	58875	60525	1650	102,80
II. Additionally calculated indicators:				
A. Revenue and cost efficiency indicators,%				
3. Expenses per BGN 100 of revenue	76,8470	75,9128	-0,9342	98,7843
4. Revenues per BGN 100 expenses	130,1288	131,7301	1,6014	101,2306
5. Net profitability based on revenue,%				
a. model 1 (item D: item A) x 100	23,1530	24,0872	0,9342	104,0349
b. model 2 (100 - item 3)	23,1530	24,0872	0,9342	104,0349
B. Capital turnover indicators, BGN				
6. Capital intensity of revenues (item 2A: item 1A)	2,6667	2,6611	-0,0056	99,7917

7. Capital intensity of own funds income (item 2A.a.: item 1A)	1,5044	1,5633	0,0589	103,9178
8. Capital intensity of attracted capital (item.2A.b.: item.1A)	1,1623	1,0978	-0,0645	94,4510
9. Capital utilization ratio (item 1A: item 2A)	0,3750	0,3758	0,0008	100,2087
10. Equity burden ratio (item 1A: item 2A.a)	0,6647	0,6397	-0,0251	96,2299
11. Ratio of attracted capital (item.1A: item.2A.b)	0,8604	0,9109	0,0505	105,8750
C. Equity composition indicators, BGN thousand				
12. Equity capital (item 2B.a - item 2A.b1)	62925	66675	3750	105,9595
13. Working capital (item 2A.a. - item 12)	26775	29025	2250	108,4034
14. Net working capital (item 2B.b - pok 2A.b2)	26775	29025	2250	108,4034
D. Equity turnover indicators, BGN				
15. Equity absorption ratio (item 12: item 1A)	1,0553	1,0892	0,0338	103,2073
16. Equity burden ratio (item 1A: item 12)	0,9476	0,9181	-0,0294	96,8924
17. Equity working capital ratio (item 13: item 1A)	0,4491	0,4741	0,0251	105,5877
18. Turnover ratio of working capital (item 1A: item 13)	2,2269	2,1090	-0,1178	94,7080
E. Indicators for attracted capital turnover, BGN				
19. Absorption ratio of attracted capital (item 2A.b: item 1A)	1,1623	1,0978	-0,0645	94,4510
20. Absorption ratio of long-term attracted capital (item 2.BA.b1: item 1A)	0,6239	0,5832	-0,0407	93,4751
21. Employment ratio of attracted working capital (item 2.BA.b2: item 1A)	0,5384	0,5146	-0,0238	95,5820
22. Capital utilization ratio (item 1A: item 2A.b)	0,8604	0,9109	0,0505	105,8750
23. Long-term attracted capital ratio (item 1A: item 2.A.b1)	1,6028	1,7147	0,1119	106,9804
24. Ratio of attracted working capital (item 1A: item 2A.b2)	1,8575	1,9433	0,0859	104,6222
F. Capital return indicators,%				
25. Gross Profitability - Model 1 (Item 1B: Item 2A) x 100	9,6472	10,0571	0,4099	104,2491
24. Net profitability:				
a. Model 2 (Item 1G: Item 2A) x 100	8,6824	9,0516	0,3692	104,2520
b. Model 3 (Item 5A : Item 6)	8,6824	9,0516	0,3692	104,2520
c. Model 4 [Item 5A : (Item 7 + Item8)]	8,6824	9,0516	0,3692	104,2520
d. Model 4 [item 5A: (1: item 10 + 1: item 11)]	8,6824	9,0516	0,3692	104,2520

Obtained results confirm the authenticity of the calculations we made for the enterprise's net return on total capital – determined in accordance with the presented models for analysis.

3. Conclusion

The proposed methodology for analysis gives the opportunity to identify the power and direction of influence of the direct factors (income efficiency, income-based return, capital consumption of income on the basis of equity and borrowings, turnover of basic equity and working equity, etc.) on the dynamics of return on total capital. This methodology allows to make the analysis of return on total capital deeper in other aspects. For example, if all long-term payables have been used as a source of formation of fixed assets (golden rule in business funding), in its nature the net working capital expresses the own working funds (working equity) of the enterprise, [1] i.e. the working equity is equal to its net working capital. The calculations so made show that the enterprise under review has implemented the golden rule for business funding.

The availability and sufficiency of net working capital in the enterprise are one of the criteria for financial stability. On one hand, the management is interested in the analysis and assessment of return on total capital with regard to the efficiency of use of the net working capital. On the other hand, the net working capital is an absolute indicator for liquidity. At the same time, liquidity "is a prerequisite for ensuring solvency. Solvency analysis is related to the analysis and control of financial position..." [2] These objective connections require theoretically justified binding of methodologies for analysis of return on capital, capital turnover, liquidity, solvency, financial stability and financial position of the enterprise. This is especially useful with view of the achievement and maintenance of the enterprise's financial equilibrium.

Of course, depending on the set objectives and tasks, and depending on the necessity of information both for the needs of the management and of external users, analysis of return on total capital may be also deepened in other directions, for example, with regard to the efficiency of use of fixed capital.

References

1. Chukov, K., Financial and Business Analysis. Sofia, University Publishing House Stopanstvo, 2011
2. Nachkova, M., Financial and Accounting Management. Sofia, UNWE Publishing House, 2018
3. Danailov, D., Corporate Financial Management. Sofia, Lauren, 1994
4. Chukov, K., R. Ivanova, Financial and Business Analysis. Sofia, UNWE Publishing House, 2014
5. Chukov, K., R. Ivanova, Financial and Business Analysis. Sofia, UNWE Publishing House, 2017
6. Accounting Act, promulgated in SG no. 95/8.12.2015, as amended no. 74/20.09.2016, as amended and supplemented no. 95/29.11.2016, no. 97/6.12.2016, as amended no. 85/24.10.2017, no. 92/17.11.2017, as amended and supplemented no. 97/5.12.2017, as amended no. 15/16.02.2018, as supplemented no. 22/13.03.2018, as amended and supplemented no. 98/27.11.2018, no. 13/12.02.2019, no. 37/7.05.2019.
7. Ordinance on the determining of limits of return, efficiency and financial autonomy ratios, the requirements to submitted evidence, special cases, methods and ways for determining the ratios and the net cash flow, Council of Ministers' Decree № 85/17.04.2006.
8. National Accounting Standards, promulgated in SG no. 30/2005, as amended and supplemented no. 86/2007, no. 3/12.01.2016, no. 15/19.02.2019.

Approaches to Research the Costs and Benefits of Implementation of Information Technologies

Rosen Kirilov

University of National and World Economy, Sofia, Bulgaria
rkirilov@unwe.bg

Abstract. The present study examines the issues of investment in information and communication technologies. As far as several studies pay predominant attention to hardware investments, the focus here is on the use of software and network connectivity. Based on data from official statistics, summaries are made, and finally, conclusions are drawn.

Keywords. Internet, Information technologies, Software

1. Business access to the Internet

The present study draws attention to trends in the use of the Internet and network connectivity by enterprises. After Bulgaria's membership in the European Union, the Bulgarian national statistics to a large extent synchronizes the conducted surveys and the collected data with the European statistics. On this basis, data from the National Statistical Institute [3] can be used very successfully to analyze persistent trends in the information society and its elements.

During the period 2014 - 2019, the state of Internet use by enterprises is improving. If in 2014 an average of 91.2% of enterprises had access to the Internet, then in 2019 this percentage was increased to 93.7% (Fig. 1).

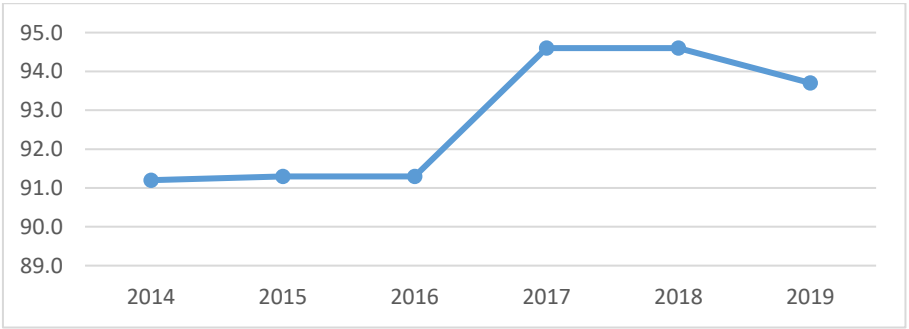


Fig. 1. Enterprises with Internet access (%) [3]

The increase is mainly due to improved internet access in small and medium enterprises.

Providing access to the Internet is not the only condition for exploring investment in information technology. It is also important to perform a constant analysis of the use of the Internet and network connectivity. The analyzed data show that in terms of the type of connectivity in 2019, on average about 80.5% of enterprises have and use broadband. In 2014, this percentage was only 70.2% (Fig. 2).

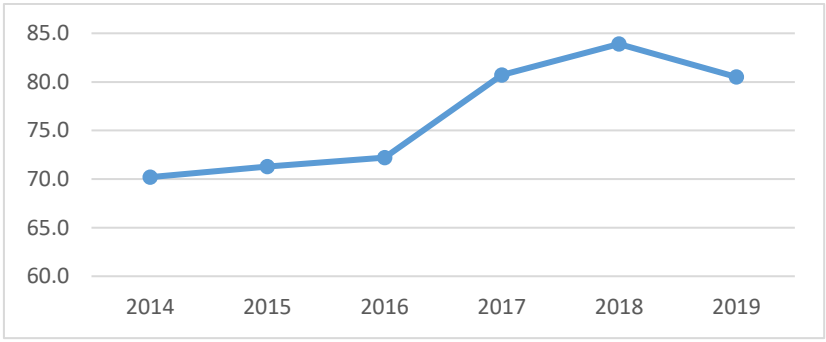


Fig. 2 [3] Businesses with broadband internet access (%) [3]

As can be seen from the presented data, about 20% of enterprises still do not have a broadband connection, and this percentage is lower for large enterprises and higher for small and medium enterprises.

2. Use of Internet technologies by enterprises

Regardless of the access provided, it is important to study the main forms of Internet use in enterprises. This can be a starting point for looking for opportunities to develop new web-based services. In this regard, one of the latest studies is by Kirilova [1]. The author examines the relationship between business process reengineering and digital service transformation. The last topic is directly related to the use of cloud technologies and big data. Some authors explore the possibilities of applying additional architectural levels in the form of a service, which can also be cloud [2]. In this regard, according to official statistics as of 2018, only 8.3% of enterprises use cloud services. This percentage has increased, as in 2014 it was only 7.7%. The distribution between the three types of micro and small enterprises, medium and large is different. In 2018, only 6.6% of small enterprises use cloud services, while for large enterprises this percentage is 29.7%. The graphical distribution of the data is given in Fig. 3.

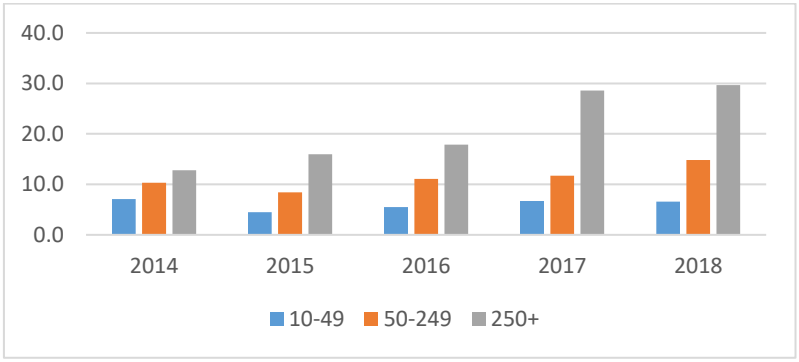


Fig. 3 Use of cloud technologies by enterprises (%) [3]

Similar results are observed with regard to the use of large data by enterprises. In 2018, the average percentage of enterprises using big data is 6.7% [3]. The percentage is higher for large enterprises (24.8%), while for micro and small enterprises the percentage is only 5.4%. The graphical distribution of the results is presented in Fig. 4.

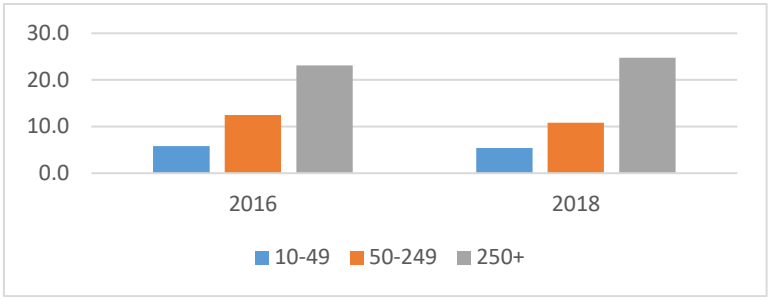


Fig. 4 Use of big data by enterprises (%) [3]

3. Costs and benefits of implementing information technology

This section examines the main costs incurred by businesses in terms of providing jobs with internet and network security. The costs are in the following areas:

- Broadband internet access costs;
- Mobile internet costs;
- Hardware costs for desktop workstations. This mainly includes the purchase of desktop computers;
- Hardware costs for mobile workstations. This mainly includes the purchase of laptops, boards and other smart equipment;
- Staff training costs, etc.

The main benefits of the implementation of information technology in these areas can be:

- Achieving a very high network connectivity and a high level of Internet use;
- Opportunities to use cloud services;
- Possibilities for application of approaches for work with big data;
- Opportunities for creating new web services, etc.

The cost-benefit comparison can provide reliable information about the level of effectiveness of decisions made.

4. Conclusion

Because of the research we can draw the following conclusions:

- There is a need to invest and increase the level of broadband internet access, especially for small and medium-sized enterprises;
- All enterprises need to expand the use of cloud technologies and services, as well as their work with big data, as an opportunity to increase overall efficiency;
- Each enterprise should regularly analyze the costs and benefits of implementing information technology, especially in the field of network connectivity and the use of web services.

References

1. Kirilova, K., Possibilities for Improvement of Administrative Processes in Local Administration, Economic and Social Alternatives, Issue 1, 2020, ISSN 1314–6556.
2. Milev, P., Integration of software solutions via an intermediary web service, Trakia Journal of Sciences, 2019, Vol. 17, Suppl. 1, pp. 181 - 185, ISSN 1312-1723.
3. https://www.nsi.bg

Current Trends in the Career Development of Students

Rosen Kirilov

University of National and World Economy, Sofia, Bulgaria
rkirilov@unwe.bg

Abstract. The present study examines the problems of career development of students and conducting practical training in a real work environment. The main goal is to present the possibilities for creating auxiliary software tools to support the main processes. An approach for creating software tools that is useful for the work of the teams from the higher schools is described. A technological solution for building a record-keeping software system of the project is offered. Finally, conclusions are formulated.

Keywords. Software platform, Information technologies, Career development

1. Conducting practical training in a real work environment

The present research is dedicated to the current trends in the career development of students. One of these trends is the conduct of many practical trainings in a real work environment during the training of students. This helps significantly for an easy transition to the labor market and allows to increase the realization of students. Project BG05M2OP001-2.013-0001 "Student practices – Phase 2" funded by the Operational Program "Science and Education for Smart Growth" is aimed at conducting practical training by students in a real work environment. The main objectives of the project are [3]:

- Supporting the improvement of the quality of higher education by providing opportunities for gaining practical experience;
- Facilitating the transition from educational institutions to the workplace and increasing the successful realization of young people in the labor market;
- Supporting the building of stable partnerships between educational institutions and business;
- Increasing the incentives of students to participate in additional practical training in a real work environment;
- Stimulating the growth of the number of students who find a job immediately after graduation;
- Providing prerequisites for updating both the curricula as a whole and the taught disciplines, courses and topics according to the needs of the labor market;
- Supporting the creation of sustainable mechanisms and opportunities for employers to select students.

Fig. 1 shows the start page of the web-based platform of the project. It publishes announcements for internships, applications for them, as well as concluding and administering all contracts and processes.

2. Possibilities for creating additional software applications

Work on such projects requires consideration of developing additional software applications outside the main project platform [3]. This requires a number of preliminary

studies. Some of them are related to analysis and modeling of less covered business processes. In the literature there is a number of studies on business processes, their role in the creation of public services and the design of information systems. Some of the latest studies in this direction are the developed productions by Kirilova [1]. It proposes to apply a broader view on the analysis of the components of the processes and their connection with the developed services. At the same time, the trends in the development of increasingly large web-based information systems, which have different architectures and allow searching for web content [2], should be taken into account.

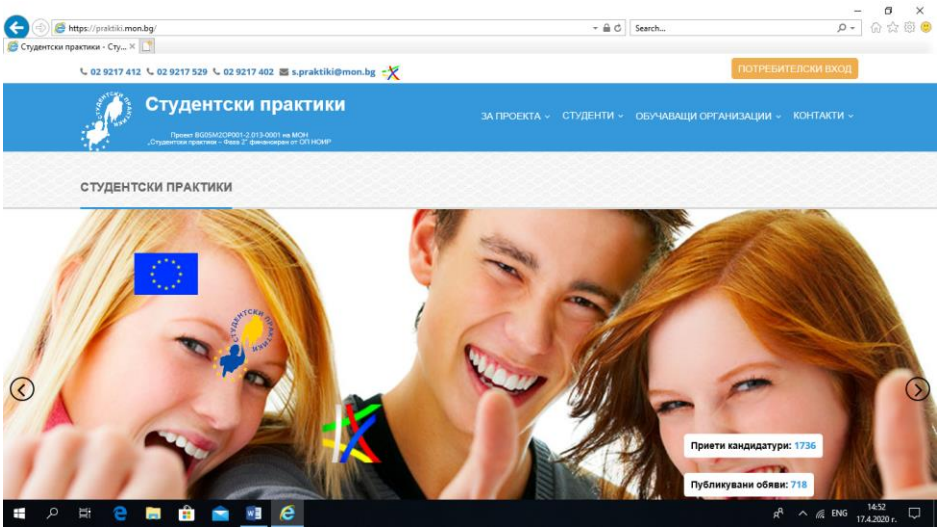


Fig. 1. Information system of the project

Based on what is stated in the present study, it is proposed that the development of additional software applications go through the following several stages (Fig. 2):

- Analysis and modeling of basic business processes;
- Software application design;
- Designing the connection and integration with other applications;
- Program implementation;
- Testing of the created software;
- Actual use of the created system.

The six main stages defined are exemplary. Depending on the individual specifics and features of the existing systems, they can be changed.

3. Technological realization of the project register system

The technological implementation of the project register system is carried out in compliance with the sequence of stages described above. Its main purpose is to allow electronic record keeping, as well as the generation of various reports on accepted, signed contracts and other documents (Fig. 3).

From an architectural point of view, the system includes the following components: a database for storing data on contracts and other documents; business logic and calculation procedures; desktop interface for data entry and reference. Fig. 4 presents a fragment of the program code of business logic.

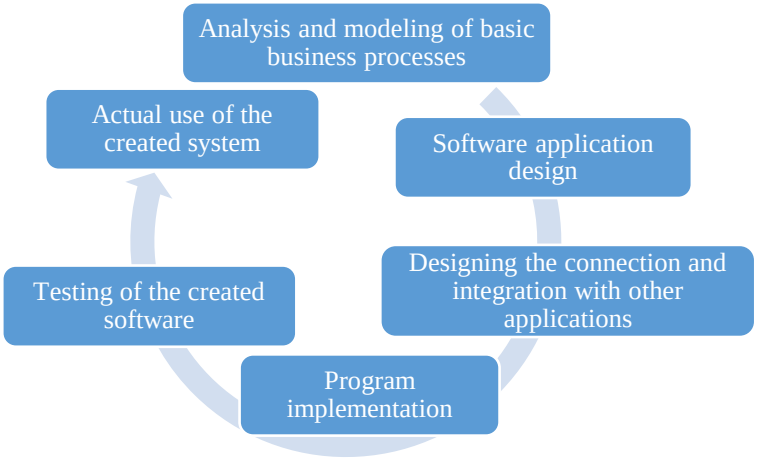


Fig. 2 Main stages of creating additional software applications

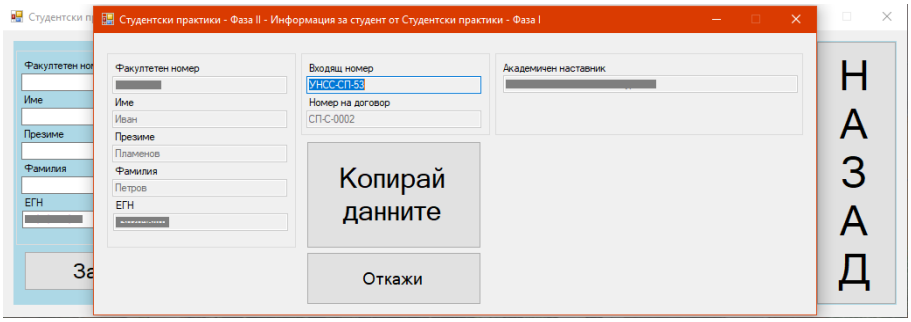


Fig. 3 Screen from the register system

```
LiteDatabase db = new LiteDatabase(LITEDBF1);
      using (db)
       {
        LiteCollection<Student> students = db.GetCollection<Student>(STUDENT);
        Student s = students.FindOne(Query.EQ("IdNumber", this.textBoxEGN.Text));
          if (s != null)
          {
             copyData = false;
             FormStudentF1 fs = new FormStudentF1();
             fs.rNumber = s.ReferenceNumber;
             fs.fMain = this;
             fs.ShowDialog();
             if (copyData)
             {
                this.textBoxEGN.Text = s.IdNumber;
                this.textBoxName1.Text = s.FirstName;
                this.textBoxName2.Text = s.MiddleName;
                this.textBoxName3.Text = s.LastName;
             }
          }
            }
```

Fig. 4 A fragment of the program code of business logic

4. Conclusion

Because of the research we can draw the following conclusions:

- Conducting practical training in a real work environment should be perceived as an element of students' career development;
- The introduction of modern information technologies and systems creates preconditions for improving the management of the processes of career development of students;
- The introduction of a record keeping system in which to register all project contracts and other documents significantly improves the project work and its efficiency.

References

1. Kirilova, K., Methodological Aspects of Administrative Service in the Public Sector, Economic and Social Alternatives, Issue 2, 2018, ISSN 1314–6556.
2. Milev, P., Conceptual approach for development of web scraping application for tracking information, Economic Alternatives, Issue 3, 2017, ISSN 1312-7462.
3. https://praktiki.mon.bg

Conceptual Approach for Application of Data Science in Web Analytics

Plamen Milev

University of National and World Economy, Sofia, Bulgaria
pmilev@unwe.bg

Abstract. The paper presents research in the context of the application of data analysis in web applications. Some traditional approaches to data analysis are presented. Publications from several public internet media are used for this purpose. The first part of the paper presents the results of these traditional approaches for data analysis. In the second part of the paper the possibilities for application of data science analysis are researched. Some results with the means of data science are presented. The conclusion outlines the need for the development of this approach in the context of modern technological trends.

Keywords. Data science, Data analysis, Web application, Dashboard, Web analytics

1. Introduction

Current trends outline the need for presentation of more data analysis of web-based information systems. The concept of web analytics is not limited to measuring web traffic. This concept is increasingly associated with the ability to track the presence of Internet media. There are different ways of presenting web data regarding the content of the respective publications. Given the large volume of such publications in various Internet media, their actual reading is practically almost impossible. As a result, various means are used to present data from web publications in the form of analysis of this data. In this regard, there are well-established traditional approaches to presenting data analysis in web applications. In the present study we will pay attention to them. We will also pay attention to the possibilities for presenting data analysis with the application of data science.

The aim of the research is to highlight some specifics in the presentation of data science analysis in web applications. The possibilities for presenting data analysis in the client part (web browser) of web-based information systems are related to the performance of the web application. Modern ways of using technologies to change performance systems have been investigated by some authors in their researches [1]. Other authors focus on the performance of database management systems in handling unstructured (web) data [5, 6].

Data integration in information systems is one of the main challenges of modern software solutions, because of which this topic is represented in many researches [7]. Information systems with data analysis can be referred to a specific kind of web-based information systems. In this sense, such software solutions may also have some additions to their architecture.

Many of the web applications of the public sector are also relevant to data analysis capabilities. Issues related to web-based information systems for public services have been investigated by different authors [3, 4]. We are about to research the traditional analysis in the context of web analytics.

2. Traditional data analysis in web applications

Traditionally, there are many software tools that provide data analysis capabilities. Many of these tools allow this analysis to be presented in web applications using appropriate technical means. In the present study we will consider some of the possible analysis, namely:

- Quantitative analysis, which shows the number of publications in which a keyword occurs, as well as the number of occurrences of that keyword for a selected period of time;
- Quantitative analysis, which shows the number of publications and occurrences of a keyword in them, but only in publications where there is another word from a list of words that we will call tags;
- Source analysis, which shows in which sources a keyword is most common;
- Content analysis that shows how often a word from a list of words (tags) occurs along with a keyword.

For the purposes of the study, we use data from publications from the list of web-based information systems presented in Table 1.

Table 1. List of used in the research web sources of publications

http://abc.net.au
http://economist.com
http://euractiv.com
http://express.co.uk
http://money.cnn.com
http://reuters.com
http://theguardian.com
http://uk.reuters.com
http://wsj.com/europe

The first analysis that we will consider is a quantitative analysis that shows the number of publications in which the keyword “data” occurs, as well as the number of occurrences of this word for the period September – November 2019. This analysis is presented in Fig. 1.

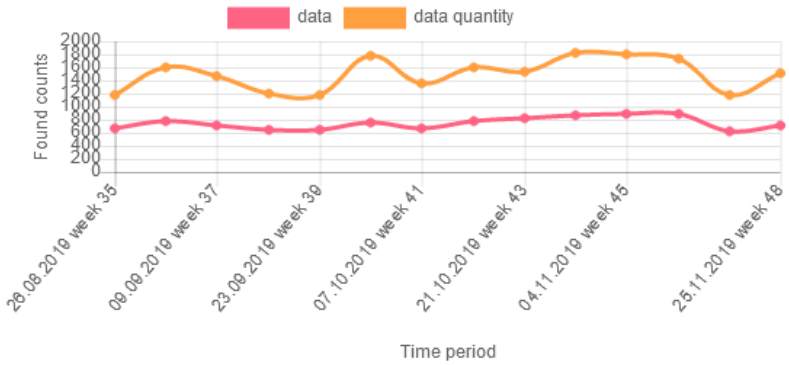


Fig. 1. Quantitative data analysis

The data in the presented analysis are grouped by weeks. We can see that the red graph of keyword occurrences is almost linear. This analysis is widely used because it gives a visual

perspective of the popularity of a keyword.

The second analysis we will look at is a quantitative analysis that shows the number of publications and occurrences of the keyword "data" for the same period, but only in publications where another word from a list of words occurs. In the context of data analysis, these words are called tags. The list of tags used for the purposes of the study is presented in Table 2.

Table 2 List of used in the research tags

software
hardware
database
website
project
concept
architecture

The results of the quantitative analysis with the attached tags are presented in Fig. 2.

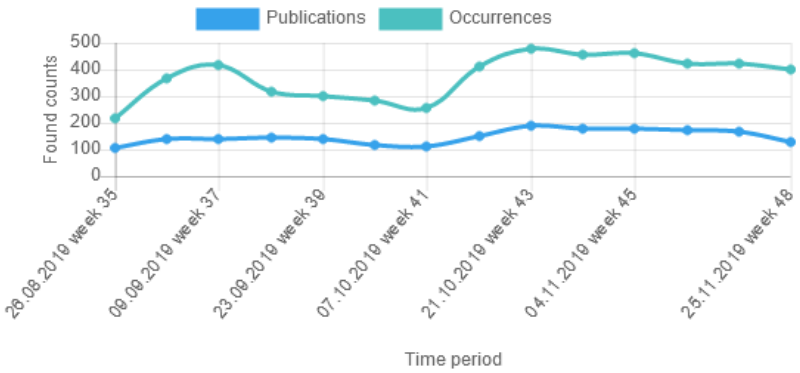


Fig. 2 Quantitative data analysis with application of tags

The data in the presented analysis are again grouped by weeks. The blue graph of the keyword occurrences along with some of the tags is almost linear. The results are less because of the application of the tags. Only those publications in which the word "data" occurs with any of the selected tags participate in the new results. This analysis is widely used because it gives a visual perspective of the popularity of a keyword in the context of other words (tags). These tags are selected according to the purpose of the study for which the relevant analysis is intended.

The third analysis that we will consider is an analysis by source, which shows in which sources the keyword "data" for the period September – November 2019 occurs most often. This analysis is presented in Fig. 3. The results of this analysis show that most publications with the keyword "data" are found on the website reuters.com. This is because for the purposes of the study, the largest number of publications for analysis belong to this web source. This analysis is useful because it shows in which internet media a keyword is most common.

The fourth analysis we will look at is content analysis, which shows how often each tag occurs along with the keyword "data". The illustration of this analysis is known as word cloud. It is shown in Fig. 4.

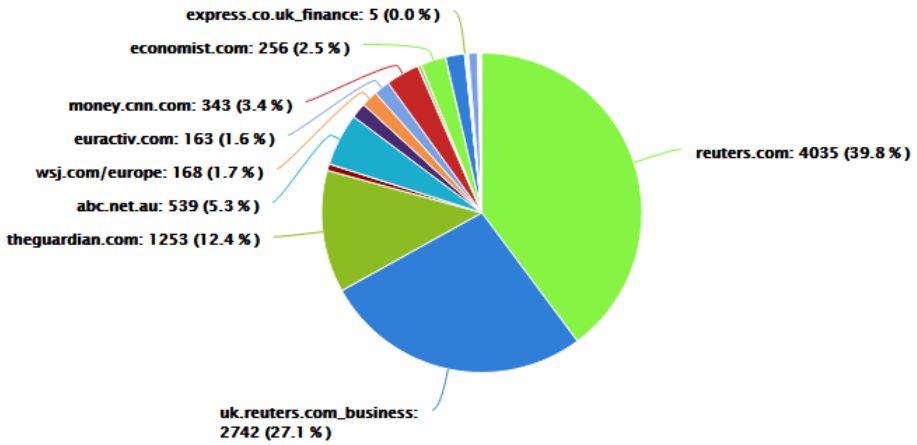


Fig. 3. Data analysis by source

Fig. 4 Traditional word cloud

The data from the last analysis illustrate that the most common tags with the word "data" are the tags project, software, and website. They are followed by the tags database and hardware. The rarest tags are the concept and architecture. This analysis is increasingly used because it gives a very good visual perspective of the presence of tags in the relevant sample of publications from Internet media.

The considered analyses give a good visual perspective of the presence of a respective keyword in publications of selected Internet media. Also, tracking certain tags can be important for users and organizations. If these analyses are rather traditional, we should explore the possibilities of processing analyses with the use of data science.

3. Data science analysis

The application of data science in the analysis of Internet media publications can be crucial. This approach would have a key advantage in some situations. Many researchers consider data science as part of the concept of artificial intelligence. Artificial intelligence includes many areas that are the focus of a large amount of research. In all of them it is said that this concept requires a significant computational resource [2]. For the purposes of the present study, we will consider data science analysis as the next level of data analysis at which it is possible to identify dependencies that are not known in advance. Earlier in the study, we looked at traditional word cloud analysis. The result was that we understood the importance

of a list of tags in the context of a keyword. However, this list was known in advance. We can repeat this analysis, but this time without a preliminary list. The purpose of the analysis will be to find out which are the most common words in combination with a keyword. In the previous analysis period was 3 months. However, this period can be very long for data processing when it comes to applying a data science approach. For this reason, we will examine the presence of the keyword "data" in the selected sources for the first week of September 2019. If we go directly to this analysis without prior preparation, the result will be like these shown in Fig. 5.

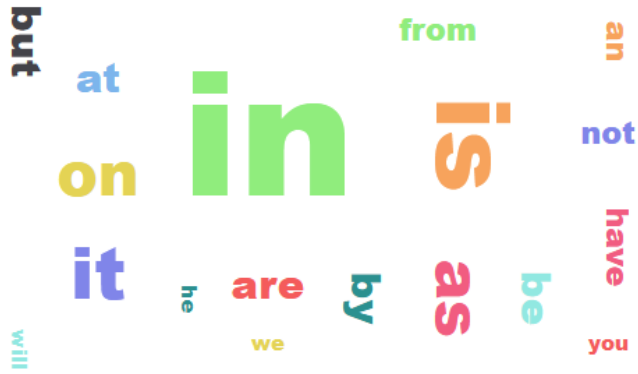


Fig. 5. Word cloud without excluding stop words

The data show that the most common words along with the word "data" are the most common words in any web publication in general. Such an analysis would be of no practical use. To get a better result, we will repeat the analysis, but this time we will exclude from the result tags the words that are not of interest for the study. Such words in the context of data science analysis are often called stop words. These words are different and should be carefully selected. Their choice depends largely on the context of the study. For the purposes of this study, we will use a list of commonly used stop words in many software solutions with capabilities for data analysis. This list includes prepositions, pronouns, adverbs, participles, and other parts of the speech that should be ignored when creating a word cloud. In this sense, it should be noted that stop words are different for each language. The new result is shown in Fig. 6.

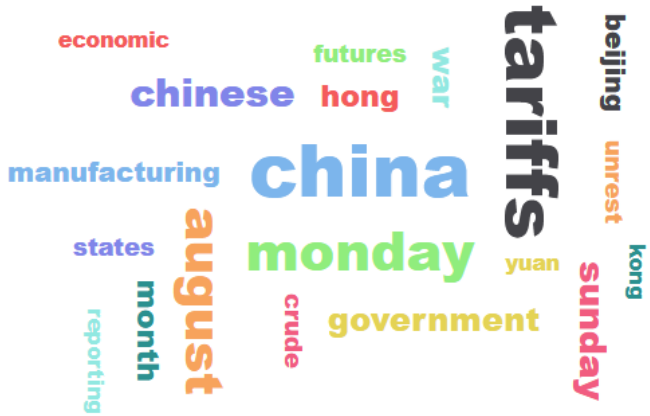


Fig. 6. Word cloud for "data"

The data shows the most common words along with the word "data". This analysis already looks in a way that can be useful for the relevant study. Such analyses are especially preferred for establishing the mood behind various organizations and celebrities. For the present study we will repeat this type of analysis, but this time with the word "software". The results are shown in Fig. 7.

Fig. 7. Word cloud for "software"

The data show that the most common words together with the word "software" differ significantly from the most common words together with the word "data". In the context of data science, there are many other types of analysis that can bring significant usefulness. The advantages of using data science analysis find serious application in systems with business intelligence, in terms of extracting knowledge from data.

4. Conclusion

In conclusion, it should be noted that the analysis using data science is very specific. It is more complex to implement, requires much more computational resources than traditional data analysis. Nowadays, data science analysis is extremely preferred and useful in the context of the development of the concepts of big data and artificial intelligence.

References

1. Belev, I., Software business process management approaches for digital transformation, Yearbook of UNWE, 2018, ISSN 1312-5486.
2. Boyanov, L., Opportunities and Threats from Implementing Internet of Things, Economic and Social Alternatives, Issue 2, 2018, ISSN 1314-6556.
3. Kirilova, K., Administrative Aspects of Public Electronic Services, Research Papers of UNWE, Issue 2, 2019, ISSN 0861-9344.
4. Kirilov, R., Technological Opportunities for the Digitization of the Career Development Processes, Economic Alternatives, Issue 1, 2020, ISSN 1312-7462.
5. Marzovanova, M., Advantages of using a system for intelligent tagging and search in unstructured data, International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2013), UNWE - Sofia, 2014,
6. Mihova, V., Dashboard design for dynamic management of database performance, International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2014), UNWE - Sofia, 2015, ISSN 2367-7635.
7. Stefanov, G., Marzovanova, M., Information system for detecting heterogeneities during multi-dimensional data mart integration, International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2015), UNWE - Sofia, 2016.

Opportunities for Presentation of Tag Cloud in Public Information Systems

Plamen Milev

University of National and World Economy, Sofia, Bulgaria
pmilev@unwe.bg

Abstract. The paper presents a study of the possibilities for applying a specific type of data analysis in a web-based environment. The specifics of tag cloud for the purposes of data analysis are considered. A sequence model for the purposes of tag cloud processing is presented. This type of analysis is useful for many organizations, including public sector organizations. The paper presents a concept for processing word cloud analysis in two steps. In conclusion, the advantages of this approach are highlighted.

Keywords. Tag cloud, Word cloud, Web application, Data analysis, Information system

1. Introduction

The possibilities for presenting different types of analysis within information systems is an actual issue. Nowadays, the Internet media create many publications every day that are of interest to various organizations from both the private and public sectors. Given the huge flow of publications in the Internet media, it is very useful to be able to review only the highlights (accents) of these publications. The presentation of these accents is appropriate to be realized by means of the so-called tag cloud. Tag cloud has many varieties. In the present study, we will consider the tag cloud as a list of tags (words) that are relevant in the context of a specific analysis. Such approaches are increasingly part of the systems of many business organizations that use different types of data analysis daily in their work. There are many studies by authors working in this field [1]. A large part of the public sector is also interested in or is already using data analysis applications. Issues related to data analysis systems for public services have been investigated by different authors [2, 3]. We are about to consider the possibilities for using tag cloud in the field of data analysis.

2. Use of tag cloud in field of data analysis

Traditionally, tag cloud is used for the purpose of presenting various accents in the respective subject area. In the context of the present study, it is appropriate to use the tag cloud to visualize the highlights of publications in Internet media. These accents are important for all organizations, but they are especially useful for public institutions, which monitor the leading events in the media daily. Many authors draw attention to the leading role of cloud resources in the preparation of various analysis [4]. Other authors explore the possibilities of indexing unstructured (web) data [5]. In this study we will look at the technological features that must be considered to represent the tag cloud within the architecture of information systems. Nowadays, due to the many available information, applications for data analysis often use a data warehouse. A data warehouse typically presents integrated data in multi-dimensional form [6]. Information systems with data analysis occupy a specific niche in the system of software solutions. In this sense, the architecture of these applications may differ in

comparison to traditional solutions. Data analysis are often implemented with the use of business intelligence technologies [7]. Fig. 1 presents a traditional view of the tag cloud. It consists of words that are set by default within the software solution for designing information systems Microsoft Visio [8].

Fig. 1. Traditional tag cloud

Usually within the tag cloud the more important tags in the context of the respective analysis are presented more expressively. The idea is to focus on those tags that are presented in a larger size. The present study proposes a classical approach to tag cloud processing and then an approach to word cloud processing in two steps. Fig. 2 presents a process of sequence diagram of tag cloud processing.

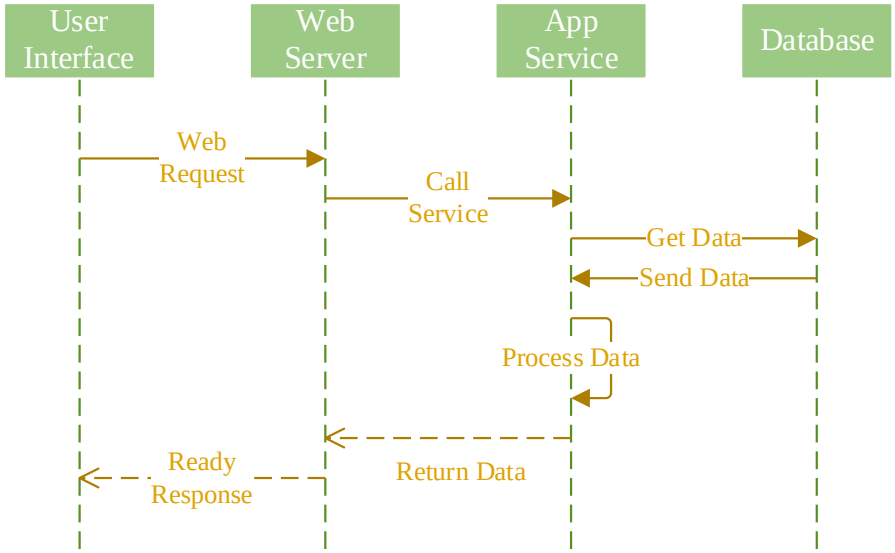


Fig. 2 Sequence diagram of tag cloud processing

The presented process consists of the following main elements:

- User interface;
- Web server;
- Application service;
- Database.

The presented process flows according to the following scenario:

1. A request for data analysis is created. This request includes the input data needed to create the corresponding tag cloud. The request is made from the user interface of the system;
2. The web server receives the request and accesses the application service, which has the business logic to perform the relevant analysis;
3. The application service turns to the database to obtain the data necessary for the preparation of the relevant analysis;
4. The database sends the data to the application service. Given the fact that this service is not web-based, there is no risk of any timeout occurring while the database is processing the relevant requests;
5. The application server performs the relevant analysis based on the set requirements;
6. The web server receives a signal that the analysis is ready to be provided to the user interface;
7. The user interface visualizes the corresponding result.

The presented sequence of actions satisfies the processing of tag cloud analysis. In many cases, however, resource difficulties would arise. This is the true for cases where there is too much data to be included in the result to be visualized. For this reason, the present study considers the possibility of applying a tag cloud presentation approach in two separate steps. In particular, the study considers the possibilities for presenting one of the varieties of this type of analysis, namely a word cloud.

3. Concept for processing of word cloud in two steps

Word cloud is a variant of tag cloud. For the purposes of this study, we will consider the word cloud as a list of words that are most common in publications in Internet media. This type of analysis would be very useful for the purpose of establishing highlights from selected publications in the Internet media for a certain period. Given the possibility that the data from these publications are in a very large volume, the study presents an approach for processing word cloud in two steps, namely:

- Step 1, which takes care of preparing the chart part of the analysis;
- Step 2, which takes care of preparing the data part of the analysis for a specific word selected from the already presented chart.

The first step of processing of word cloud includes the following actions (Fig. 3):

- The client (web browser) makes a request to the web server with a description of the searched keyword and period;
- The web server transmits this request to the application service, which has the business logic for preparing the analysis;
- The application service prepares a chart part of the analysis with information about the available words and their importance (number of publications in which they occur or number of occurrences in total in all publications).

After completing the first step, the client has the chart part of the relevant analysis within a web-based user interface. In cases where this is necessary, the client can make a second request for access to the data. This data usually includes information about a specific word from the word cloud chart. In the present study, these data are links to publications in the Internet media in which the relevant words are found.

The second step of processing of word cloud includes the following actions (Fig. 4):

- The client (web browser) makes a request to the web server with a description of the word for which the relevant data should be obtained;
- The web server transmits this request to the application service, which has the business logic for preparing the analysis;
- The application service prepares the data part of the analysis with information about the links to the publications from the internet media in which the respective word occurs.

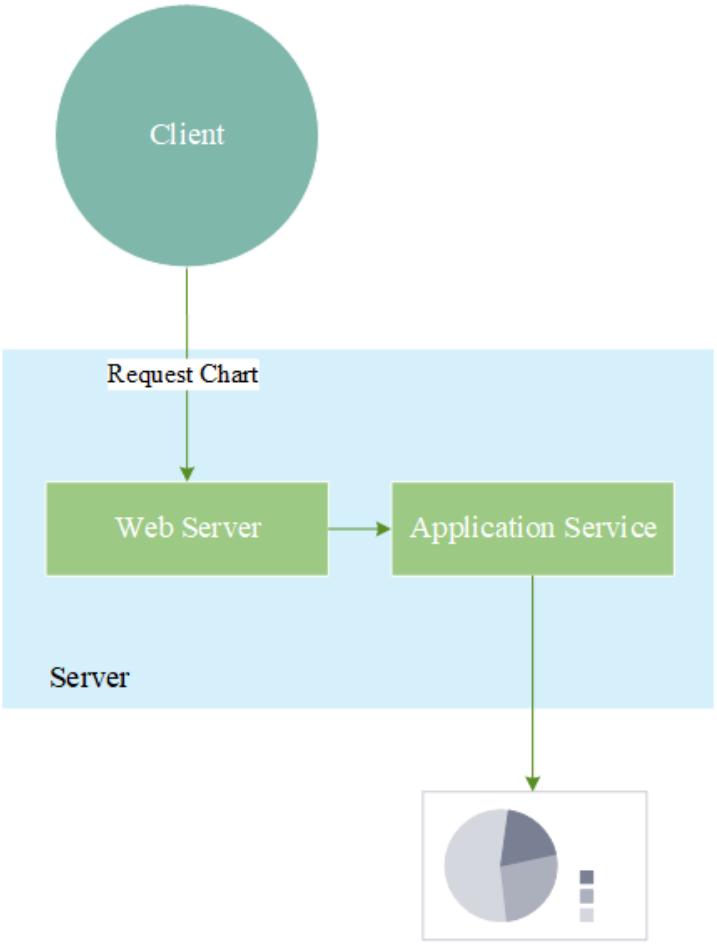


Fig. 3 Step 1 of the concept (chart processing)

After completing the second step, the client has the chart part of the relevant analysis within a web-based user interface, as well as the data part of the respective selected word from the chart. The separation of the analysis into two separate parts has some key advantages, namely:

- Faster initial preparation of the chart part of the word cloud analysis, without loading all the data for which it is not known whether it will need to be presented;
- Fewer queries to the database, because data is extracted for the data part of the tag cloud analysis only for those words that are of interest in the context of the analysis;

- More optimized complete operation of the web-based information system, which is due to the availability of an intermediate application service, which processes separately the two steps of the process of preparing the word cloud analysis.

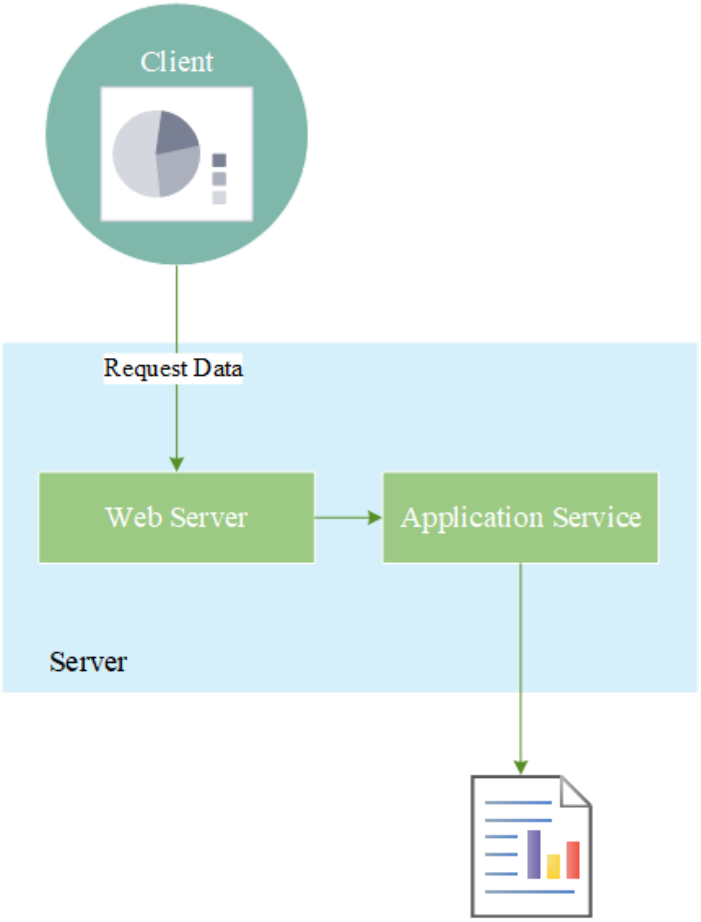


Fig. 4. Step 2 of the concept (data processing)

After completing the second step, the client has the chart part of the relevant analysis within a web-based user interface, as well as the data part of the respective selected word from the chart. The separation of the analysis into two separate parts has some key advantages, namely:

- Faster initial preparation of the chart part of the word cloud analysis, without loading all the data for which it is not known whether it will need to be presented;
- Fewer queries to the database, because data is extracted for the data part of the tag cloud analysis only for those words that are of interest in the context of the analysis;
- More optimized complete operation of the web-based information system, which is due to the availability of an intermediate application service, which processes separately the two steps of the process of preparing the word cloud analysis.

4. Conclusion

In conclusion, it should be noted that tag cloud analysis and its variants are becoming increasingly popular. These analyses are very useful because they visually illustrate the meaning of the respective tags (words). This type of analysis is often used in both the private and public sectors. The present study presented an approach of dividing word cloud design into two separate steps. In this way the chart part and the data part are processed relatively independently. This can lead to significant advantages in some situations.

References

1. Belev, I., Software for business process management and the potential for implementation in Bulgarian universities, Economic and Social Alternatives, Issue 3, 2017, ISSN 1314-6556.
2. Kirilova, K., Methodological aspects of administrative service in the public sector, Economic and Social Alternatives, Issue 2, 2018, ISSN 1314–6556.
3. Kirilov, R., Features of application of software solutions in public projects risk management, Economic and Social Alternatives, Issue 2, 2017, Sofia, ISSN 1314-6556.
4. Lazarova, V., The Expansion of the E-business Information Systems in the Cloud – Problems and Solutions, Economic and Social Alternatives, Issue 4, 2015, ISSN 1314-6556.
5. Marzovanova, M., Creating annotations and indexes in system for intelligent tagging and search in unstructured data, Proceedings of International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2013), UNWE - Sofia, 2014, ISBN 978-954-644-586-5.
6. Mihova, V., Methods of using business intelligence technologies for dynamic database performance administration, Economic Alternatives, Issue 3, 2015, ISSN 1312-7462.
7. Stefanov, G., Methods for heterogeneity detection during multi-dimensional data mart integration, Economic Alternatives, Issue 1, 2015, ISSN 1312-7462.
8. Visio, 2019, 01.12.2019, https://www.microsoft.com/bg-bg/microsoft-365/visio/flowchart-software.

Study of the Status of the Administrative Service Processes in the Municipalities

Katia Kirilova

University of National and World Economy, Sofia, Bulgaria
katia.kirilova@unwe.bg

Abstract. In the present study the issues of the management of the administrative processes in the municipalities are considered. Their analysis and modeling can be considered as a basis for their reengineering. This is also a starting point for the development, offering and introduction of new digital administrative services. The performed study and researches give grounds for defining some recommendations and conclusions.

Keywords. Public administration, Administrative service processes, Municipalities

1. Providing local digital administrative services

The basis for studying the status and trends in the provision of digital local administrative services is the Unified Portal for Access to Electronic Administrative Services, maintained by the State Agency for Electronic Government [3] (Fig. 1).

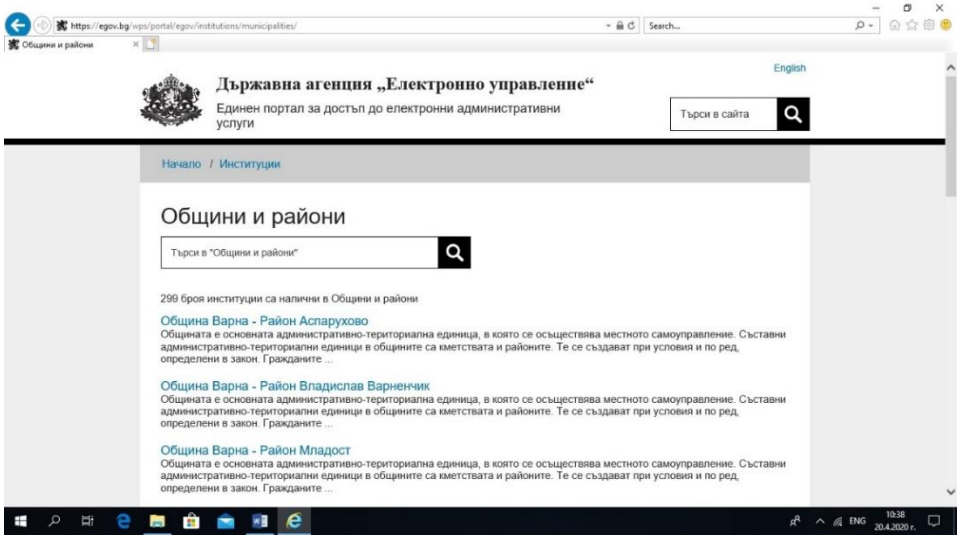


Fig. 1. Unified Portal for Access to Electronic Administrative Services [3]

The services provided in the register are categorized in the following areas:

- Business and entrepreneurs;
- Civil rights;
- Civil status;

- Taxes and fees for citizens;
- Health and health insurance;
- Property, housing and utilities;
- Education and qualification;
- Environment;
- Agriculture and forests;
- Social activities;
- Transport and vehicles;
- Spatial planning, construction and cadastre.

It is important to point out that the municipal administrations may include the administrative services provided by them under each of the above. If we analyze the services in the section "Name, address registration, identity documents" it is impressive that the registered services in this mass direction are only 28, provided by a limited number of administrations (Fig. 2).

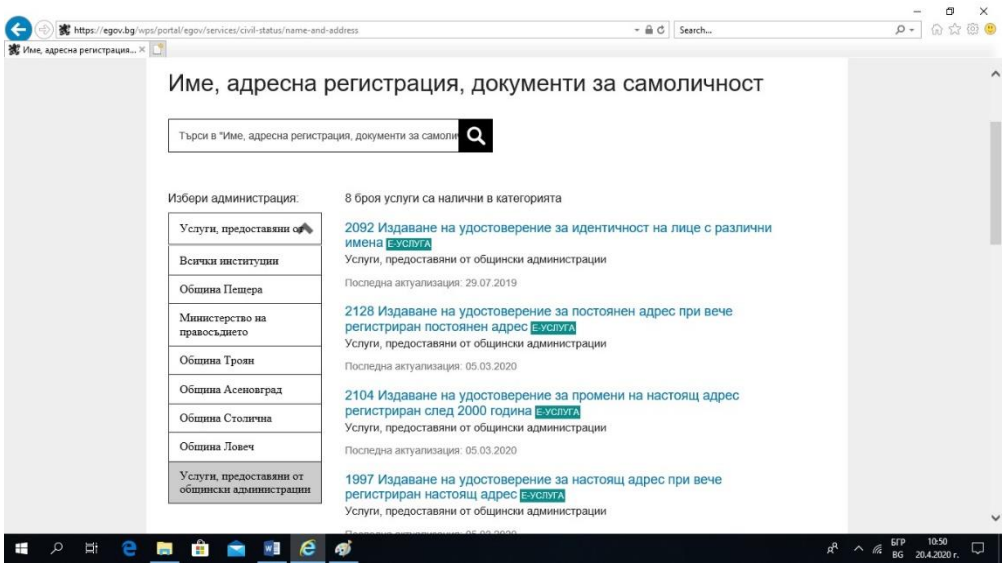


Fig. 2. Services provided in a specific section [3]

The situation is similar in several other sections and categories, with a significant number of them in which there is no digital administrative service. These facts are quite worrying and show that now the processes of digitalization of local administrative services have been slowed down for various reasons. The main ones can be:

- Lack of funds;
- Lack of sufficiently qualified specialists;
- Limited use of services;
- Low percentage of persons who would perform digital administrative services, etc.

The disclosure of these and other reasons can be done with a high degree of accuracy only when a survey is conducted among a representative sample of municipalities. Such a

study can also provide answers to the life cycle questions of the developed digital administrative services.

2. Life cycle of digitalization of administrative services

The following more important characteristics from the stages of service development can be set as an element of the study (Fig. 3):

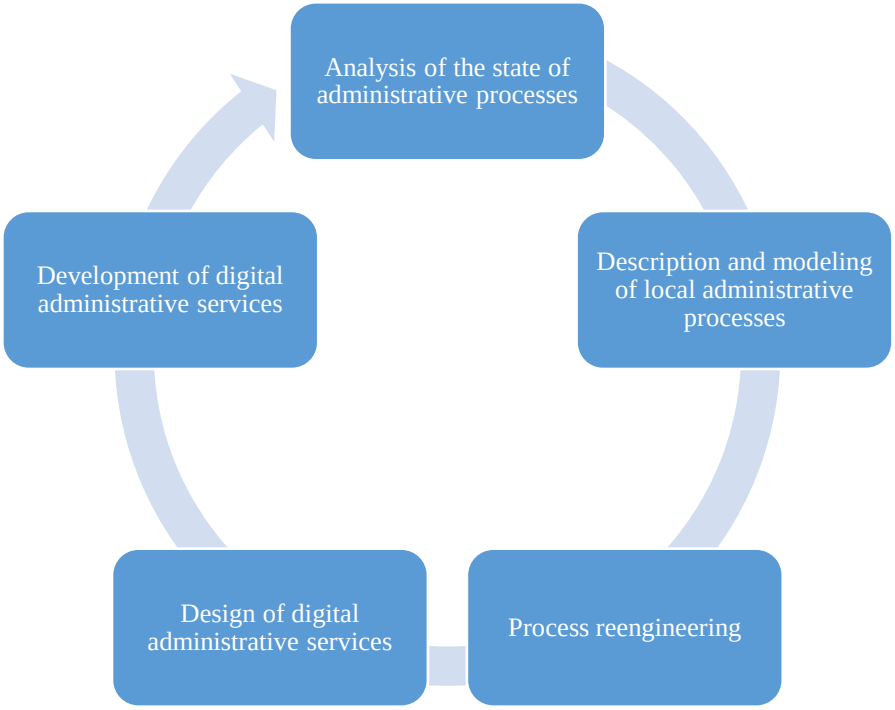


Fig. 3 Life cycle of digitalization of administrative services

- Analysis of the state of administrative processes;
- Description and modeling of local administrative processes;
- Process reengineering;
- Design of digital administrative services;
- Development of digital administrative services.

The initial phase is related to the analysis of the current state of the administrative processes. This phase can also be defined as "identification". Its purpose is to establish the administrative processes that lead to the creation of local administrative services. It is important to point out that at this stage it is possible to identify several processes that do not lead to the creation of services but are of an ancillary nature.

After the identification phase, it is necessary to proceed to the analysis and modeling of the administrative processes, and for this purpose appropriate software tools should be used. In the literature there are several studies on the subject, one of the last in this direction is by Kirilov [1]. It addresses the issues of digitalization of processes using specific information technologies. Other authors study the different trends for modeling from an architectural point

of view [2]. The main result of the modeling of administrative processes is the creation of process maps with their complete design specification.

The next phase is related to the reengineering of the processes that are modeled. This is a difficult process and in most cases the results are optimized processes. They can be successfully used in the next stage to design new digital administrative services in the local administration. It is necessary to carefully specify all the characteristics of the designed services in advance. The last phase is software implementation of digital administrative services and monitoring of their effectiveness in the real work process.

It is important to note that these five steps are indicative and may be modified or adapted at any time according to the specific conditions of application.

3. Conclusion

Because of the research we can draw the following conclusions:

- Local administrations in Bulgaria face the challenge of developing more and more digital administrative services;
- Serious analytical tools should be developed and applied for the analysis of the considered topic. It should allow for empirical research among municipal administrations;
- Each municipal administration can create its own methodology for analysis and modeling of administrative processes. This is an important prerequisite for the creation of new digital administrative services.

References

1. Kirilov, R., Technological Opportunities for the Digitization of the Career Development Processes, Economic Alternatives, 2020, Issue 1, pp. 184-195, ISSN (print): 1312-7462, ISSN (online): 2367-9409, DOI: https://doi.org/10.37075/EA.2020.1.10.
2. Milev, P., Trends in the development of architecture solutions for web publishing systems, Business Management, 2019, Issue 2, pp. 17 - 33, ISSN 0861-6604.
3. https://egov.bg/wps/portal/egov/home

Opportunities for Development of Soft Skills of the Administration Employees

Katia Kirilova

University of National and World Economy, Sofia, Bulgaria
katia.kirilova@unwe.bg

Abstract. This study analyzes data on the skills of administration staff. Emphasis is placed on skills for working in a digital environment and the search for opportunities to develop soft skills. An approach for a sequence of steps to develop soft skills in an administrative environment is proposed. Based on data from official statistics, several summaries are made, concluding with outlines.

Keywords. Public administration, Administration employees, Soft skills

1. Citizens' digital skills

In the last few years, business and the public sector are in a state of constant transformation of processes. Some of these trends are provoked by the digitalization of services and the increasing use of Internet technologies in business, education, administration, and social life. To establish these trends and dependencies, the National Statistical Institute conducts annual surveys on the level of e-skills of citizens. They are in several areas and are mainly related to the development of digital competencies for working with modern information technologies and solutions. The summarized data for 2019 has been be given in Table 1 [3].

Table 1. E-skills of individuals according to NSI data for 2019 [3]

Types of e-skills (%)	2019
Copy or move a file or folder	44,1
Using word processing software	27,3
Using spreadsheet software	15,9
Use software to edit photos, videos, or audio files	10,6
Creating a computer program	1,1
Transfer files between a computer and another device	43,2
Change or check the configuration parameters of software applications	7,4
Creating electronic presentations using the appropriate software with images, sound, video or graphics included	14,5
Installing software or applications	21,5

The comparison of these data with the data from the previous similar researches shows that the e-skills of the persons on the separate components increase. On the other hand, the National Statistical Institute also maintains long-term data on the interaction of citizens with public institutions. In 2019, such an interaction was performed by 25.4% of the respondents, and in 2014 this percentage was only 21%. The formed tendency is upward, as the graphic representation is given in Fig. 1.

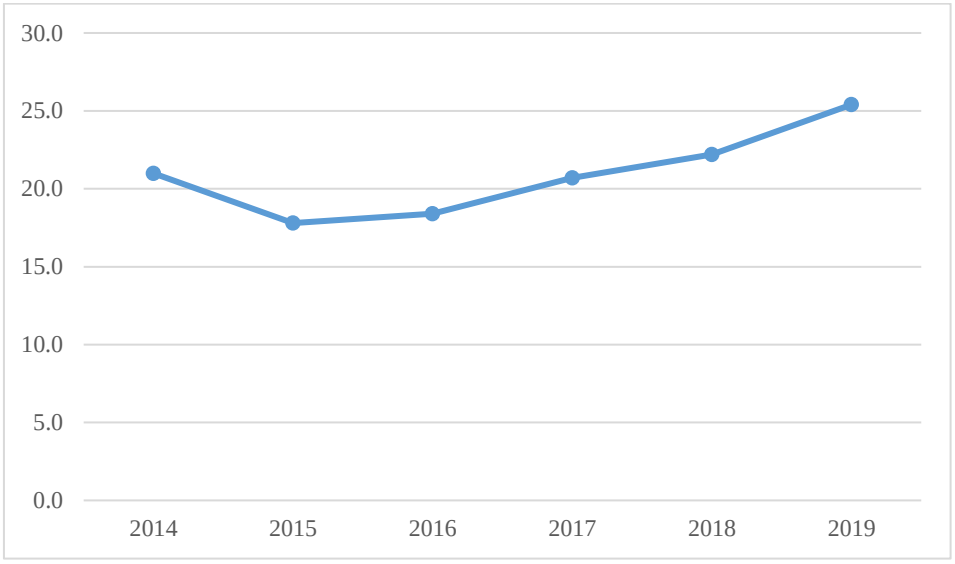


Fig. 1 Interaction of citizens with public institutions (%) [3]

The shown tendency imposes and requires the creation of a comprehensive policy to increase the e-skills, as well as the soft skills of the employees of the administration in response to the changes in the environment.

2. Basic digital and soft skills for administration staff

Improving the digital and soft skills of administration staff is a complex process. It requires the development of special training programs, as well as for the acquisition and improvement of both digital skills and soft skills to deal with new realities. One of the interesting studies in the field of soft skills development is by Alison Doyle [4]. It develops a detailed description of different categories of soft skills, and the most important for the work of employees in the administration can be summarized in the following three groups:

- Communication (Listening, Negotiation, Nonverbal communication, Presentation, Public speaking, Verbal communication, Visual communication and Writing skills);
- Leadership (Conflict management, Conflict resolution, Decision making, Motivating, Project management);
- Work Ethic (Business ethics, Competitiveness, Independence, Multitasking, Organization, Planning).

The development of the indicated soft skills in the three groups will lead to the improvement of the administrative service and an opportunity for participation of the employees of the administration in various projects and programs for administrative service. Along with this, the question of the search for opportunities for digitalization of basic processes, such as those described by Kirilov [1], is also important. In this regard, it is necessary to make efforts to expand digital competencies in the following directions:

- Working with integrated environments and platforms;
- Digitization of basic processes;

- Working with digital signatures and electronic identity;
- Participation in the development of digital administrative services;
- Working with electronic control systems, Internet monitoring [2], etc.

The development and upgrading of these digital competencies requires the implementation of specific policies and programs to achieve the required level of effectiveness.

3. Opportunities to increase the digital and soft skills of the employees of the administration

Increasing the digital and soft skills for the employees of the administration will allow a significant improvement in the provided administrative services. On the other hand, it is a starting point for the creation of new digital administrative services at central and local level. The opportunities to increase these skills should be done through a certain sequence of actions and stages, such as the one shown in Fig. 2.

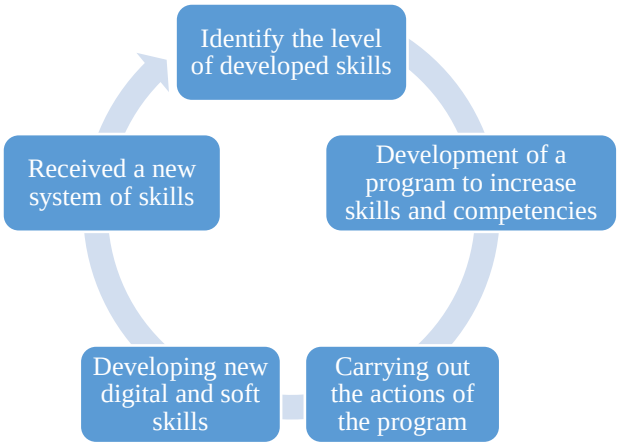


Fig. 2 Main stages for developing digital and soft skills

The proposed system includes five main stages for the development of digital and soft skills for the employees of the administration. They are:

- Identifying the level of developed skills;
- Development of a program for increasing skills and competencies;
- Carrying out the actions of the program;
- Development of new digital and soft skills;
- Received a new system of skills.

The initial stage is related to identifying the level of developed skills for employees. This can be done through the use of questionnaires or interviews. Identification is important in order to be able to suggest the characteristics of the developed program for increasing skills and competencies. It aims to provide opportunities for specific actions and trainings. Each of them is carried out according to a specially developed plan or program and helps to create or develop a specific digital or soft skill. Once the program has been developed, the specific

actions of the program should be carried out for each employee or group of employees. This is a lengthy process and under normal conditions can take a long time - from a few weeks to several months. The actions at this stage should also be perceived as a form of investment in the human capital of the specific administration. As a result of the conducted trainings the development of new digital and soft skills in the different groups is obtained, which can be summarized in a specific system.

The proposed system of actions should be taken as an example, and in each case this sequence of actions may be modified and changed.

4. Conclusion

Because of the research we can draw the following conclusions:

- The identification of the available digital and soft skills of the employees of the administration should be perceived as a starting point for proposals for their improvement;
- The development of digital skills by employees can be an important prerequisite for increasing administrative capacity, because it is based on the application of modern technological environments;
- The development of new skills and competencies, as well as the expansion of existing ones is a form of investment in human capital.

References

1. Kirilov, R., Technological Opportunities for the Digitization of the Career Development Processes, Economic Alternatives, 2020, Issue 1, pp. 184-195, ISSN (print): 1312-7462, ISSN (online): 2367-9409, DOI: https://doi.org/10.37075/EA.2020.1.10.
2. Milev, P., Opportunities for implementing Internet monitoring in public organizations, Economic and Social Alternatives, Issue 2, 2018, ISSN 1314-6556.
3. https://www.nsi.bg
4. https://www.thebalancecareers.com/list-of-soft-skills-2063770

Adopting e-Learning Tools into Marketing Courses to Encourage Students Experience Their Own Decision-Making Attitudes

Vanya Slantcheva-Baneva

MTM College Sofia, Bulgaria
v.slantcheva@mtmcollege.bg

Abstract. The paper resumes experiential adoptions of selected e-learning tools used into the educating process of undergraduate and postgraduate classes in Marketing for a decade. Couples of specialized marketing tools that support academics into both the teacher-student interrelations offline, and the peer-to-peer interactions online are outlined. The paper review includes: a Marketing Simulation Tool, a Conjoint Analysis Module, an Operations Management Simulation, a Positioning Game. A discussion of student experiences is generated. Crucial learning outcomes are represented by means of theory acknowledgement and practice embodiment while students test their own decision-making.

Keywords. Higher education, Marketing simulations, Experiential learning, Conjoint analysis, Positioning game.

1. Introduction

Adopting simulation tools in marketing courses as a ground for experiential learning exercise is becoming a more than necessary task. It is turning into a demanding *modus operandi* when it comes to the issue of when students are supposed to understand decision-making causalities. Marketing simulations are developed to encourage students to experience their own decision-making attitudes. They put them into a hypothetical market context where, as learners, they are to generate situations and unfold model decision loops. The *modus* is built on setting clear marketing goals and leaning towards an "analysis-decision-analysis" iteration. Such an interactive procedure could include the following steps: a) conducting a situation analysis, b) setting goals and choosing a strategy for further policy execution, c) coming to a decision-making based on a wisely spread budget, d) submitting decisions, and e) reflecting on feedback.

The purpose of the report is to outline the academic experience of using marketing simulation tools addressed to facilitate Marketing courses in undergraduate and postgraduate classes. The ultimate goal is to perform and track (to announce) the symbiosis of mediative teaching process and self-debriefing learning activity when the decision-making issue is tackled. As a ground of the experiential learning paradigm, it has its favourable influence on students' disposition to make their decisions in Marketing cases.

2. Experiential learning for decision-making in academic classroom

Using specialized simulation tools in the academic space, a teacher provides a supportive learning environment for students as a ground for an active experiential learning

called “situational learning” [5]. As an incorporation of active, participatory opportunities in a course, experiential learning occurs whenever the student is aroused from the role of a passive listener to that of an active respondent. Learning creates opportunities for "data learning" when students are intended to participate in determination of the learning process itself.

Considered as role-playing instruments, simulation tools, on one hand, engage students with a proactive behaviour helping them to express opinions and use inductive reasoning [10]. And, on the other hand, they provide opportunities for the teacher to improve his/her teaching productivity. Moreover, they support students to experiment with their ideas, advancing and employing concepts into real situations. Learning by situation simulating considerably enhances the students’ engagement with the problematics studied and makes them participate in what Michaly Csikszentmihályi called “flow experience” [6].

Academics who use simulations employ characteristics of authentic pedagogy when designing their teaching and assessment activities. Most use simulations to provide a real world experience that allows students to develop the kinds of skills and competencies they will need in their working lives. Academics provide a full range of tasks to capture a breadth of student learning preferences and to allow for knowledge and skill development. They see reflective activities as a means for students to really interrogate their own learning and to make connections between theory and practice. Although possible group problems that could emerge (generally because one or more members does not contribute) [12], all see the benefits of teamwork and collaboration in helping students to learn and fill the gaps in their learning.

Diversity of learner groups is a characteristic of higher education today. Students within programmes are therefore likely to come with a range of different learning characteristics and backgrounds. These characteristics include differences in prior knowledge, abilities, personality and household background, language competence, motivation, expectations towards achievement, preferred ways of learning, willingness to engage in collaboration, and current levels of understanding. Students also approach learning differently.

They could take a surface or a deep approach to learning. In the context of a simulation, students might have different levels of knowledge and understanding of the concepts that underlie the game and international students may struggle with language and terminology. Other students may have less prior training or skill in the financial and numerical concepts that could help them interpret results. Nevertheless, marketing simulations allow students to engage in sufficient and varied activities to ensure that they cover all the associated learning outcomes. [12] Making regular decisions, whether over a short intensive period or over 6-10 weeks, and then having the opportunity to interpret results, are examples of the regular activity provided by simulation's themselves.

Marketing simulation tools used in higher education become a powerful drive that contributes to better academic education of principles and practical aspects in marketing rationale. Playing simulations, students experiment differentiated implementations of marketing strategy. They make their tactical marketing decisions and receive a prompt feedback from the simulation, which measures their individual performances by means of selected marketing KPIs. Role-playing helps students to overcome the abstractness of the terminology, studying the terms in context and interrelations. Prior to the experiential learning process, the teacher delivers critical instructions to the participants, paying attention by sole discretion to selected particulars that are to be interactively experienced and cognitively loaded.

3. Selection of Marketing Simulation Tools Carried Out in Educational Process

Marketing Simulation: Managing Segments and Customers

The simulation is an online-based interactive tool [1, 2]. In using it, students make decisions regarding a company's marketing strategy. They make important decisions – regarding product design, pricing, discount structure, marketing expenditures, sales-force size – that should collectively support an overall marketing strategy designed to achieve a combination of sustainable revenues and profits over 12 fiscal quarters. Performance is measured using both qualitative and quantitative indicators, including profitability, revenues, unit sales, market share, and customer satisfaction (Fig. 1).

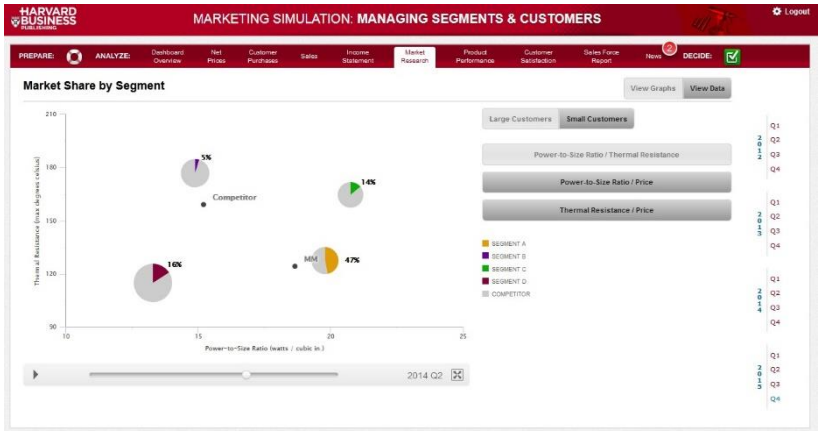


Fig. 1. Marketing Simulation: Managing Segments and Customers, V2. Market Research

The main learning objective is for students to understand the link between marketing strategy formulation and its effective implementation and execution. They need to gain an understanding of segmentation, targeting, and positioning issues. Moreover, students learn to use segment/customer needs analysis to make product design decisions. They manage customer acquisition and customer retention strategies.

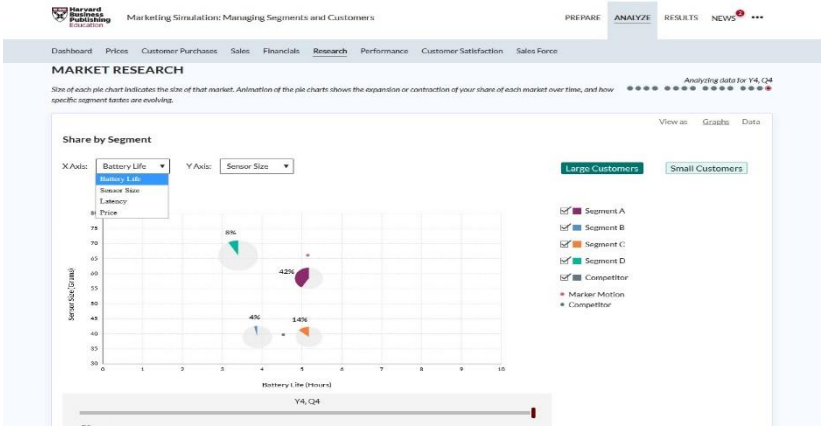


Fig. 2. Marketing Simulation: Managing Segments and Customers, V3. Market Research Tab.

Along the presented decade, two versions of the tool were announced within the Harvard Business Publishing of Educators and adopted in considered Marketing classes. Moreover, a third one has already been out since November 2019 (Fig. 2). The main update is the new Market Motion Ltd. case that makes students take into amount three product features, instead of two, as they were in previous versions of the Minnesota Micromotors Ltd. Case (Fig. 1). Hence, a new learning objective should pursue the issue of interdependency between combinations of variables (product feature and product price) [2].

Marketing Simulation: Using Conjoint Analysis for Business Decisions

This simulation is part of a comprehensive Conjoint Analysis Toolkit, which includes two modules: The Conjoint Analysis Online Tutorial and the Conjoint Analysis Do it Yourself Guide [8, 9]. The simulation is designed to reinforce student understanding and use the conjoint analysis as one of the most popular market research methods in academia and practice (Fig. 3).

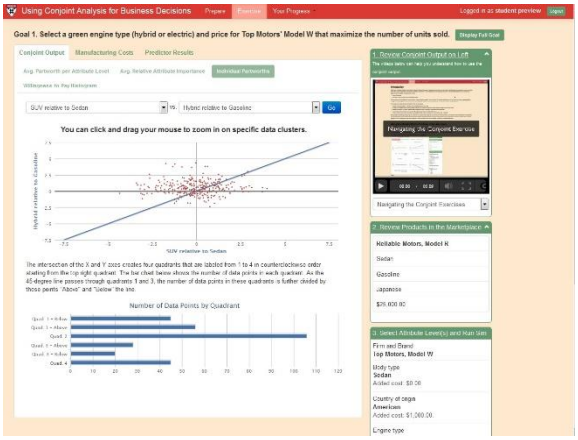


Fig. 3. Interface of the Marketing Simulation: Conjoint Analysis for Business Decisions.

Its goal is to provide students with an appreciation of how conjoint analysis output can be relevant in practice, and to provide an interactive experience of using conjoint analysis for managerial decision-making. Delivered online, the simulation gives students exposure to key business concepts such as demand curves, segmentation, profit functions, competitive responses, vertical and horizontal differentiation, optimal pricing, niche vs. mass market strategies, product portfolio management, and brand equity.

Operation Management Simulation: Benihana

That simulation tool helps 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 (Fig. 4).

The teaching plan is to help students systematically unearth the elements of service 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 how terms such as throughout capacity, demand variability, capacity utilization and service time apply to a service production process. The most important message the simulation conveys to the learners is that the profitability could be the KPI for designing and evaluating scenario-driven servuction strategies. Students learn to analyse capacity, demand rates, cycle time, and throughput in a service operation. Running the simulation, they optimize the capacity in an attempt to optimize multiple variables in an

operation and to ensure the consistency of the overall strategy [11].

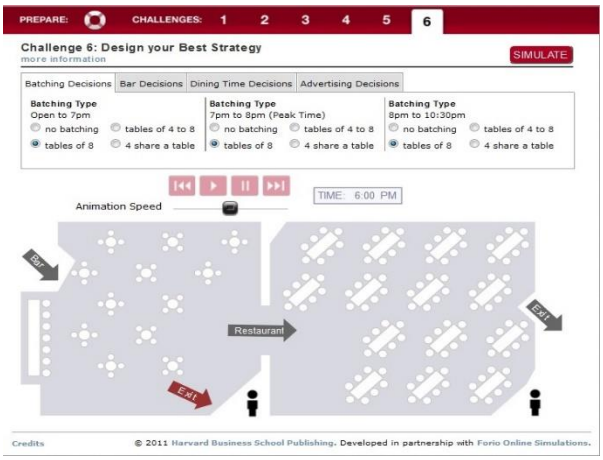


Fig. 4. Simulation Interface of Challenge 6: Design your Best Strategy.

Marketing Simulation: The Positioning Game

Using perceptual maps, students make decisions about launching new products in the context of impending market competition. It is a multi-player tool where students compete in the launch of a new or enhanced product by positioning their brand at an ideal place in the market. [3] Decisions are made in real time, and costs are incurred with positioning changes. Launching new products into new or established markets provides the greatest opportunity for a company to think strategically and to reap the benefits of these decisions [Fig 5.].

Fig. 5. Positioning Game Interface: Results. Darden Business Publishing.

A core learning objective of the simulation is to introduce the critical role of market structure and product positioning in marketing strategy. Students need to analyse markets and recognize consumer segments, underserved segments, and ideal points for targeting opportunities. They learn to comment on brand perceptions, similarities, and differences, and recognize perception problems.

4. Discussion

Table 1 illustrates the total number of students enrolled into the four Marketing simulations during the decade 2010-2019. Approximately 300 students (73 undergraduates and 230 postgraduates) participated in the experiential learning experiment. Undergraduate students were those who enrolled on specialized Marketing programme. A common feature of the postgraduate students was that they attended a Marketing course or programme, but they had a non-Marketing or IT-educational background.

Table 1. Coursepacks' launches of marketing simulations. Students enrollments (2011 – 2019).

MARKETING SIMULATIONS:	Students enrolled (numbers)		Simulation coursepacks launched (2010-2019)	
	Undergraduate	Postgraduate	Per academic year	Total years
Managing Segments and Customers	35	230	1.7	9
Using Conjoint Analysis for Business Decisions	30		1	2
Operation Management Simulation: Benihana	1	0	1	1
The Positioning Game	7	0	1	2
	73	**230**		

The tool primarily carried out in Postgraduates' classes was *Marketing Simulation: Managing Segments and Customers*. It had got 15 launches or appropriate 1,7 times each year between 2010 and 2019, except the academic 2012/2013. Until then the simulation exercise was proposed as an optional one. Students were supposed to select whether or not to play and report on it as an optional assignment for final exam. Since the academic 2013/2014 year the instructor set the simulation as a formal assignment of the course and students participated in it as part of the course requirements. They had to play the game successfully, to write a report and to present their best score completion. Having the Marketing Simulation as an optional one, let some students neglect learning benefits of the tool and make inconsistent decisions – to enrol and purchase but not to play the simulation. They refused participation, sometimes arguing that the Marketing course is a marginal one for them. On the other pole, there were IT-postgraduates who perceived the simulation as an experiential learning opportunity to learn about the Marketing paradigm. Most of them performed a progressive engagement with the platform. Their completions resulted in an active experimentation with the model of decision causality. *The Marketing Simulation: Using Conjoint Analysis for Business Decisions* was purposefully introduces in a Marketing Analyses course in a specialized Bachelor programme in Marketing. The experiential learning started with an initial adaptation of the comprehensive Conjoint Analysis Toolkit. The instructor predominantly integrated it into the course syllabus during the academic 2014/2015 and 2015/2016 years. But the interaction with the simulation tool itself remained as an elective assignment for final exam. The Conjoint Analysis theme within the course syllabus appeared to be highly assessed as relevant, reasonable, and practice-oriented one. Another simulation exercise launched in 2013/2014 and 2014/2015 for the purposes of Marketing Analyses course, was the *Marketing Simulation: The Positioning Game*. Only 4 students took place in the spring term of 2013/14, and on the next academic year – 3. The Positioning Game was an optional exam assignment and the students were separately instructed how to play it and what theory frame it is grounded on. It was a multi-player exercise and students in 2013/14 class ensured extra-curricular coordination. Facilitated by the instructor, a Skype connection was pre-established and the main points of the interaction

were clarified and considered. The interface was intuitively perceived with drag-and-drop pointing gestures. Students began playing with fun driven by receiving prompt performance feedback of the submitted decisions. Amazingly, the fun turned into a contest mood boosted by the applied positioning tactics.

Marketing Simulation: Operation Management Simulation: Benihana was carried out in the academic year of 2013/2014. It was an optional assignment for the final exam with in Services Marketing course. The student who chose to learn through that simulation approached it ambitiously. Rehearsing a couple of scenarios, she experienced 70 runs and accepted a best-practice service process with superior profit.

5. Conclusion

The package of the four Marketing Simulations considered above are single-player exercises, except the Positioning Game as a multi-player one, although the instructor can have the students play them in teams. All four simulations provide students with an array of tactical decisions to make and focus on marketing practice and theory. The structure of the simulations is interactive and enables multiple opportunities for self-debriefing and learning. They are flexible enough to allow students to use a variety of strategies successfully. There is no single correct solution, but the students look for the better one.

The retrospective view of the experimented experiential learning methodology in Undergraduate and Postgraduate classes studying Marketing reveals some important issues. Students need to be committed to the experiential learning paradigm. Otherwise, they would proceed pragmatically within the learning process. They have to be encouraged to demand simulation archetypes to self-debriefing. In that respect, Higher Educational Institutions should build up an adaptive infrastructure to facilitate and support both academics and students integrate software games into reciprocal academic interrelations.

References

1. Das Narayandas. *Marketing Simulation: Managing Segments and Customers V2*. Online Simulation, HB Publishing, 2010.
2. Das Narayandas. *Marketing Simulation: Managing Segments and Customers V3.*, Product No. 8880-HTM-ENG, HB Publishing 2019.
3. Farris, Paul W. *Marketing Simulation: The Positioning Game*. HB Publishing, Product No. UV6715-HTM-ENG, Darden School of Business, 2014.
4. Fink, L. Dee *Integrated Course Design.* The IDEA Center: Kansas, adapted with permission from John Wiley & Sons 2003, 2005.
5. Hawtrey, Kim. "Using experiential learning techniques", *The Journal of Economic Education*. 38(2), pp. 143-152, 2007.
6. Kolb, A. Y. and D. A. Kolb "Learning styles and learning spaces: Enhancing experiential learning in higher education", *Academy of Management Learning & Education* 4(2), pp. 193-212, 2005.
7. Ofek, E., O. Toubia, *Conjoint Analysis: A Do-It-Yourself Guide.* Product No. 515024-PDF-ENG, Harvard Business Publishing, 2014.
8. Ofek, E., O. Toubia, *Conjoint Analysis: Online Tutorial.* Product No. 514712-HTM-ENG, Harvard Business Publishing, 2014.
9. Ofek, E., O. Toubia, *Marketing Simulation: Using Conjoint Analysis for Business Decisions.* Product No. 515713-HTM-ENG, Web Based HTML, Harvard Business Publishing, June 2015.
10. Reynolds, M., and R. Vince. *The Handbook of Experiential Learning and Management Education*. Oxford: Oxford University Press, 2007.
11. Sasser Jr., W. E., R. Ernst, *Operations Management Simulation: Benihana V2*, Product No. 7003-HTM-ENG, HB Publishing, 2011.
12. Voss, L., Simulation games in business and marketing education: How educators access student learning from simulations", *Int. Journal of Management Education*, Issue 13, pp. 57-74, 2015.

Adopting Techniques to Engage Students when Teaching with Cases Online

Vanya Slantcheva-Baneva

MTM College Sofia, Bulgaria
v.slantcheva@mtmcollege.bg

Abstract. The paper summarizes the contemporary issue of moving case teaching online that accounts of designing student-teacher interaction by means of keeping students engaged with. The consistency of the case method conduction both in-person and online is explored. In that respect, the design of the traditional framework is addressed by focusing on the first two stages of case teaching – the foundation and the flow. The issue of moving the case teaching online boosts the research perspective to the contemporary discussions about creating transformational learning experiences and student engagement.

Keywords. Case teaching, In-person, Online teaching, Student engagement, Transformational learning experiences.

1. Introduction

The purpose of the report is to examine the consistency of the case teaching paradigm comparing the traditional and the transitional environments of conduction – in-person and online. Referring to the traditional composition of case teaching is necessary to delineate those steps that are to become convertible from in-person to online context. The focus of the paper is primarily the design of the preparation (foundation) and the discussion (flow) stages. There the corresponding sub-stages within are to be unzipped and fragmented as case paradigm episodes to be translated in terms of the online teaching layout. Particularly, case method components such as "establishing classroom infrastructure" should be remodeled in an intangible and inseparable mode as to be provided online, and should directly concern the nature of the student engagement issue.

2. Consistency of the Case Teaching In-Person and Online

Bringing the case method online encounters the challenge of scaling under pressure to maintain the highest level of participant-centered learning possible [3]. Examining the traditional case teaching journey is necessary here to identify those components that are to be transformed for the purposes of online teaching. Moreover, blueprinting the case teaching process make it possible to compare the traditional and the synchronous journeys from their practical guidance perspective in terms of understand the way we teach and learn by the case method (Table 1).

Foundations

The Foundation stage in case teaching is a grounding one. Comparing the two modes in teaching – in-person and online – it appears to be consistent, prepared formally in advance. It becomes convertible when the point comes to establish online participation requirements and online "infrastructure" establishment, which is the actual online preparation.

Taking Ownership is a valuable feature of the foundation stage in case teaching. It means the case teacher to make the content and the message of the course clear to him- or herself and others. He has to ensure that he has the authority and self-confidence to take control of the student experience, understanding where his support and opposition will be. Moreover, taking ownership regards setting the students' expectations about the content and conduct of the course, and their own role in it.

Table 1. Comparison of online and in-person case teaching design, based on the traditional practical guidance and the contemporary online teaching experience [1]

ELEMENT	CASE TEACHING JOURNEY	
	TEACHING IN-PERSON	TEACHING ONLINE
FOUNDATION	• Taking ownership (self-confidence) • Making a contract • Developing course content • Motivating students • Helping the students prepare for case discussion • Establishing classroom participation requirements • Establishing infrastructure • Setting length and frequency of class sessions • Wrapping up	• Setting learning objectives • Sets clear tasks • Cold calling to reticent students
FLOW DURING THE CLASSROOM	Preparing yourself • Crafting a teaching plan • Demeanor in class Starting to teach • The first class • Opening classes • Managing discussion • Beyond discussion Concluding discussion	• Facilitate discussion • "Sitting in" as a manager, auditor, client, or consultant. • Interacts: warm calling; questioning, creating polls, videos, presentations, breakout grouping. • Suggests processes to make students productive.
FEEDBACK	Designing Evaluation • Designing Assignments – oral and written; individual and group; before, during and after case teaching process. • Designing Exams – selecting cases for exam and writing exam questions; • Designing Term Papers Grading • Scaling • Structuring feedback	Scopes the deliverables to prove meaningfulness and usefulness of the case.

Making a contract acknowledges being explicit about expectations about the course, while describing it in the formal syllabi. The case teacher as a faculty member sends a letter to the students before the course starts explaining the contract (syllabi). It requires that the student devote time in the first class to explain the course structure and how the case discussions are to be conducted. Further, the faculty member *develops the course content*. He lines up the issues to be covered by dividing the course into parts, and structures the syllabus. The teacher selects and assigns the cases, as well as the readings and textbooks.

Having the contract and the course content in store, the case teacher needs to market the course in order to attract students, believing that it will succeed on its own merits. —by word of mouth and course description only. *Motivating* students is important because they should invest scores of hours in a course, and why it is worthwhile. To convince students, the case teacher as a faculty member should *demonstrate enthusiasm* for the material for further *student engagement* with the case method. He or she should focus on the case method relevance, as well. The case teacher is committed to the value of learning for its own sake, but emphasizes that the real-world relevance of case studies and discussions will teach students about the practical implications of what they are learning, is crucial. For those students motivated by the trade aspects of what they are taught, the teacher emphasizes how grounded case discussions are in the reality of practice. Moreover, case discussions help students to build valuable skills presenting students with the opportunity to improve their ability to speak publicly, think on their feet, and improve their problem-solving and pattern-matching skills. Teacher also focuses on keeping teaching environment safe where students can overcome their typical anxiety in case discussions.

Foundations in case teaching are set on *helping the students prepare for case discussion.* The teacher becomes an instructor and reminds students to dedicate enough time, read the case several times, pay attention to exhibits with numbers, answer any preparation questions, ask what the main challenges are for the case protagonist(s) and what he would do about them, formulate an outline for a strategy, and be ready to talk about it in class. Moreover, he states the *classroom participation requirements* as a part of the grade. He makes it clear that he reserves the right to call on any student at any time on any question or course material for which the student has requested preparation. In-person, classroom participation does not just mean that the student is physically in the classroom. The student has to say something to show preparation, and he has to contribute to the discussion by building on other comments from classmates or the teacher, and to work towards understanding problems and generating alternative solutions. While excluding the physical attendance, the same remains valid for the *virtual classroom participation requirement:* good virtual participation means moving the discussion forward; successful virtual participation is achieved when everyone, including the teacher, learns something from the discussion. The dialog with peers to advocate for case solutions should occur regardless of whether it is in-person, or online. The classroom becomes a simulation of the reality, and students should be ready to discuss, submit decisions and analyze them.

Discussion-based teaching needs a supporting infrastructure. Schools dedicated to case teaching have special classrooms designed for discussion and the administrative support necessary for handling the many little details involved in a successful case course. *Establishing virtual infrastructure* then becomes crucial to adapt the teaching process and the virtual classroom to each other. The question *is* about establishing "seating arrangements" in which the students face each other, but the pattern in the virtual classroom responds to the functionality and reliability of the teaching platform, that should "simulate" a horseshoe or a circular pattern, e.g. breakout rooms. Setting length and frequency of class sessions remains equal to the both in class and online case teaching.

The preparation phase is not just about the structure, content, and context of the course, but also about what comes after it. *Wrapping up* foundations, flow and feedback stages is important to be clarified before the actual process of case teaching in-person or online begins.

Flow

Preparing the course is a scene-setting exercise, expressed in the syllabus. The teacher prepares for each case he is going to teach: to understand the content of the case and how it relates to other topics and cases in the course; to decide on a teaching plan, including a board

plan; and to plan how he/she is going to run the discussion. Preparing as an instructor should result in a *crafted teaching plan* for the case. The plan is more like a framework giving some milestones the instructor wants to reach rather than a detailed script.

Usually the plan contains: 1) the main points the students have to take away from the class; 2) links to other cases and topics in the course. 2) the topics to be discussed, in proper order and in terms timeframe; 3) the *opening question*, who should probably answer it; 4) a rough board plan of the structure to be written on the board(s) during the classes or sessions. 5) frameworks, models, theories, presentations, videos, and other supporting material included in the course and in case teaching. Preparing for the class, the case teacher understands that his role as an instructor is based on his own style and *demeanor* in class. During the *first class* students learn the mechanics of discussions. Including rules or guidelines in the written syllabus, the case teacher plans a separate session with the students, in class or in online session, before the course starts. A *case discussion* should be cumulative: each comment should build on the preceding one. In order to do that, students listen to each comment by putting their hands up when they want to say something. It serves to underscore the mechanics of the discussion and the role the students have in giving and receiving, learning to and from each other [5].

Further, the instructor spends some time and asks the participants what their *concerns and expectations* are. An effective technique is to go around the room (in teaching live) and in both environments – to write down their answers on the whiteboard, and then revisit the list to check that the student addresses at least most of the expectations. This is to make it clear that the learning outcome of a seminar or a session is a joint responsibility of the audience and the discussion leader.

The classic way to start a case is to have a student *open the case discussion*. The instructor has three main ways of doing this: a "cold call," a "warm call," and an "open call." A *cold call* occurs when an instructor starts by asking a student directly. A *warm call* is when the case teacher starts by naming the student who will open the discussion, but give the person some time to collect his or her thoughts. He can also notify students when and how he plans to call on them further in advance, by email [5]. An *open call* is when the teacher simply asks for a volunteer to start the discussion. A great way to start the class is to *have a vote* on some question or *poll* the students on some numerical point, finding the thinking behind the analysis.

The core of case teaching lies in *managing discussion*. The instructor decides who gets to speak, when, and for how long; uses boards to keep track of the discussion; asks questions that drive the discussion forward; and is conscious of his own movement and position in the classroom or of his presence during the online sessions. The most important content of case teaching are the student contributions—their suggestions, comments, and questions.

Using the board is a privilege of case teaching. The board is its vital tool; as the students come up with various points about the case, the teacher writes them down (using key words, drawings, tables, or whatever methods he or she can think of to do it fast). Taking notes in public serves a number of purposes. First, it provides a shared overview of what students have been discussing. A second purpose is to guide the discussion. The instructor can also use the board to introduce frameworks and models into the discussion. The purpose of using boards and developing board plans on it is to drive the discussion.

In classroom using the *body language, silence, and movement* is much of the art of discussion management. The instructor can direct a discussion well by using body language, not being very explicit. A powerful tool not to be underestimated is silence in both in classroom, and in online sessions. If the presenters pose a question to the audience and when nobody answers, it is preferable not to start rephrasing the question or answering it himself. Instead, he or she needs to wait and keep quiet, until somebody blurts out something.

In most schools, classrooms are designed to be faculty-centric and support lecture pedagogy. During the case discussion the case teacher tries to open up the space and physically signal that he will be *sharing the space* with the group. In online case discussions "sharing the space" amounts to managing cold and warm callings, or *asking questions*. Choosing questions for a case discussion is similar to designing an interview guide for a field research project. Questions serve many purposes. The instructor uses them to start discussions, to guide the discussions deeper or higher, to follow up on student contributions, and to make transitions between students and topics.

Case teaching is about much more than the core method of a single teacher guiding individuals in a discussion of a case in a plenary setting. It is about *using theory* and other non-case material, but not reciting theories and frameworks. It is about the ability to examine a real-world situation (represented by a case), apply a theoretical perspective to it, and draw conclusions from this about how to proceed. Despite using theories in case discussion, a technique that creates an immersive environment, bringing the case and related analysis to life and fostering empathy for the players involved in the case, is *role-playing*. It can build insight and increase commitment to the case. Moreover, the case teacher can break off the plenary discussion and have the students discuss the topic between themselves or in breakout rooms online [5, 6]. He may also want to generate detailed ad hoc group discussion among the students on some smaller points.

The flow of case discussion is to be concluded at determined point. The goal of the conclusion of a case class is not only to summarize and focus on the key learnings, but also to finish as to have the students continue the discussion among themselves after class and to provide a transition to the next class.

3. Designing Case Teaching Online by means of Transformational Learning Experience

Creating a Transformational Learning Experience

One of the main goals when teaching online is making classes as interactive as possible "to keep students from tuning out". Finding pedagogically meaningful ways to engage them is essential to online learning. This means to make a synchronous class, which is more discussion-based avoiding *just-speaking* mode and using the available technology. In the flow stage of managing the discussion, case teaching online adopts particular techniques of student-teacher interaction by means of the transformational learning experience (Fig. 1).

Basic interaction with students could be provided through *audio* and *video content* shared on the screen. *Screen sharing* becomes a must so that the whole class can view presentation materials. *Hand writing* or drawing on the screen could be practiced by Whiteboard tools or substituted by a tablet. *Chat windows* are to keep track of, but they are a powerful tool to allow students who aren't speaking to bring up ideas or follow up on previous points. The best way for the teacher to keep the students engaged online is to look at the camera and *to sell* the course for the students to become convinced that it is valuable. The teacher needs to present involvement by conveying the sense of the commitment within the course. Using *warm calls* by letting students know ahead of time that the teacher has planned to call on them; or *cold-calling* by emailing students ahead of time about the topic to be discussed. The teacher has either to interject discussions to reengage students, or to make the discussion a central part of the lecture especially when he applies the case method or the peer-to-peer learning. Creating *polls* ahead of time to use during class or writing them on the spot if something interesting comes up during the discussion is also an engaging technique. Moreover, when online students' attention fades, teachers could *summarize* what have been discussed.

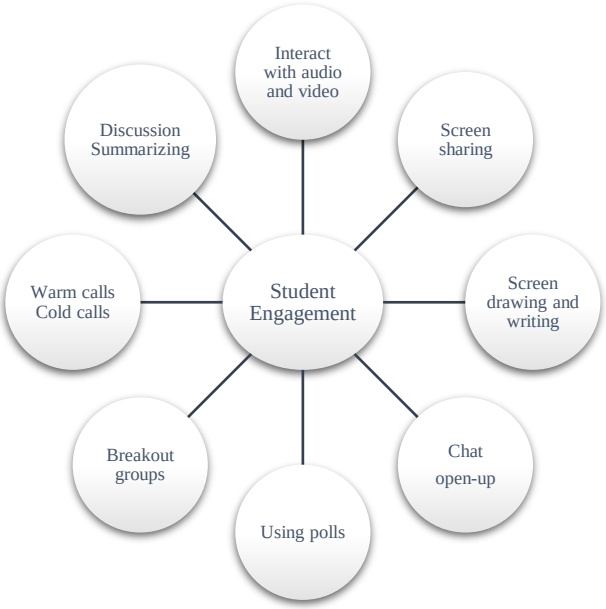


Fig. 1. Student engagement techniques supporting teaching with cases online

Most online teaching platforms integrate *breakout rooms* to get students into virtual groups. An effective tactic to get students more actively involved in case discussions is to use a breakout session, where students can talk to one another in a group of 4 to 7 students. The case teacher pops in on each group to check on progress and to ask if they have questions, "sitting in" as group manager, auditor, client, or consultant. He sets a clear task; suggests processes that might make students more productive; scopes the deliverable so that it would be meaningful and useful [6].

In an interactive lecture the teacher sets a clear learning goal for the day, shares expectations around how that goal will be achieved, and provides questions for groups to answer. Students then work to discuss an idea, reflect, and try to draw closure. The teacher requests that the students learn or students engage the teacher with what they have already learned. Learning interactions are rarely concentrated between the students directly. Rather, informal cooperative moments are generated in classrooms. Hence, in online teaching educators carefully dedicate a larger proportion of course time to formalize those "informal moments" into cooperative learning structures. Those structures would keep students "united toward a common goal" [4]. To ensure a more cooperative learning experience educators employ practices, which promote coordination, interdependence, and accountability. These practices can work well in synchronous online teaching settings, in-person, or in a blended format (Table 2).

With any given version of in-person teaching educators may be tempted to rely on naturally occurring group dynamics and generate enough informal cooperation to hold more instructor-focused courses again, as the case teaching is. But teaching "as usual", even in a hybrid world that involves small groups with physically distanced seating, is no longer enough. Reciprocally, students will have undoubtedly higher demands for the quality of the face-to-face time. [4]

Table 2. Formal practices to ensure cooperative learning experience [4]

8 Formal Cooperative Practices in Teaching Online	Poll-group-repoll
	Value lane
	Students Teams-Achievement Devisions
	Learning Roles
	Role Play
	Expertise-Specialization Project
	Peer Grading
	Group Prcessing

Designing a student-teacher interaction in online case-teaching journey

Moving the case teaching online is about maintaining the benefits from teaching with cases in the traditional classroom. Usually, academic case studies have three or four decision points that students have to discuss, by using the knowledge they have, and then exercising judgment. Students coming to class prepared to having read the case debate, discuss. Moreover, they listen to alternative points of views, shape and deepen their thinking based on the arguments made. The case teacher plays the role of trying to make sure that the learning is deepened by questions that students are asking. He intervenes to drive the case to a rich discussion of the issues that helps peers hone their judgment and they're learning skills. Case teaching online is being designed as an interactive medium and a participant centered *modus operandi* (Table 3). Transforming the learning experience of the case method online, more or less, it is about retaining all of the in-person experiences rather than change the case method approach [2, 3].

Table 3. Designing a student-teacher interaction in online case-teaching journey

ELEMENT	TASK		PRINCIPLES
	STUDENT'S ROLE	FACULTY MEMBER'S ROLE	
PREPARATION	• Need to become aware of the text of the case	• Setting learning objectives • Cold calling to reticent students • Sets clear tasks	Consider a business context or a situation
DISCUSSION	• Exercise decision points of the case. • Use the knowledge gained. • Exercise judgment.	• Facilitate discussion • "Sitting in" as a manager, auditor, client, or consultant. • Interacts: warm calling; questioning, creating polls, videos, presentations, breakout grouping. • Suggests processes to make students productive.	*Students* debate decision points. *Faculty member* ensures the learning is deepened by students' questions *Students* listen to alternative points of views *Faculty member* makes interventions to drive the case to a rich discussion of the issues that helps students hone their judgments.
JUDGEMENT	Write down an argumentation on a decision made upon the situation	Scopes the deliverables to prove meaningfulness and usefulness of the case.	*Students* shape and deepen thinking based on the arguments made *Faculty member* perform a mentorship

There are guiding principles that help the faculty member support variable courses. In the beginning, the educator considers that he/she is going to cover 75% of that material, as a minimum, because of the technology gaps that are to come. Also, the teacher has to get as much data about what students think they are coming into class with, based on their pre-class preparation, doing polls inside or outside the e-platform. In this manner, the educator already knows what students think and consider, so that he can draw out much more quickly than what he might ordinarily need to do in the physical classroom.

Using *breakout room* for the case discussions help teacher to establish "circular seating arrangements" for student engagement. It is a feature of video and audio conferencing platforms as Zoom, open source web conferencing as BigBlueButtonBN within Moodle, or Skype, used in online teaching, or distance education. The case teacher breaks the class in small groups of 4-6 students online either randomly or in a preselected fashion. He announces before the class the issues that are to be discussed or the technique that is to be studied and approaches the generated virtual teams.

After completing the task, the class is called back, the small groups are reassembled and students share than what they did in smaller groups. In breakout room the case teacher cannot see everyone at the same time, but he pops in on each group to check the progress and to enrich the discussion that leads the group to the demanded judgements. The key is to constantly keep the students engaged to cooperate around a common goal.

4. Conclusion

Despite the differences between in-person and online case discussions, there are important similarities between them. These are still case discussions [5] that are to be prepared, managed and closed. Pattern recognition and development of solutions in cooperation with others remains continuous in both. Students keep learning it by "distilling key problems from fuzzy data, formulating solutions, and advocating for them in dialogue with peers" [1]. Most cases focus on decisions, and it's possible in an online environment to put even more of the work on students to decide what to do. If students feel like they are accountable, they are likely to be more invested. In keeping students engaged, the case teacher uses the need for interactivity and format changes [7] to his/her advantage in keeping them engaged. The virtual classroom then becomes a simulation space where the case teaching expounds on the experiential learning paradigm.

References

1. Andersen, E. and B. Schiano, *Teaching with Cases: A Practical Guide*. Harvard Business School Publishing 2014.
2. Busuttil, L. G. and S. J Van Weelden, *Student Guide to the Case Method*. Ivey School of business Foundation, 2018.
3. Datar, S. and B. Kenny, "Bringing the Case Method Online", Podcast, *Harvard Business Publishing Education*, April 8, 2020.
4. Peshkam, A. "What Your Pre-COVID Course Design Was Missing: 8 Cooperative Learning Practices to Enrich Your Online or Hybrid Classroom", *Harvard Business Publishing Education*, June, 19 2020.
5. Schiano, Bill "Mastering the Case Teaching in Online Classes: Don't Sweat the Small Stuff and Let Your Students Do More of the Work", *Harvard Business Publishing: Education*, 6 p., April 30, 2020
6. Schiano, Bill "You can still engage students when teaching online – here's how", *Harvard Business Publishing: Education*, 6 p., March 13, 2020.
7. Yael Grushka-Cockayne, "Designing a Curriculum for Online and Hybrid Classes" Webinar, *Harvard Business Publishing: Education*, July 9, 2020.

Business Processes Digitalization in Accounting

Vanya Lazarova

University of National and World Economy, Sofia, Bulgaria
vlazarova@unwe.bg

Abstract. The report discusses the main stages of digitalization of business processes in accounting. The factors that determine the underlying processes to be digitized are identified and the life cycle of the processes that have to be digitized are traced.

Keywords. Business processes digitalization, Digitalization in accounting.

1. Introduction

To define business processes in accounting that have to be digitalized is not an easy task. Accounting is limited by many requirements, standards, laws, regulations, and any change requires many factors to be taken into consideration.

The purpose of this report is to discuss the main stages of digitalization of business processes in accounting. The factors that determine the underlying processes to be digitized are identified and the life cycle of the processes that have to be digitized are traced.

2. Determining the corporate functions and processes ahead for the initial or advanced digitalisation

Determining the corporate functions and processes ahead of initial or advanced digitalisation depends on the following several factors:

1. Degree of completion of the processes that precede digitalisation - electronicization and automation in the company.
2. Motivation and need of digitalization according to the company's management team.
3. Availability of free funds for digitalization.

The processes prior to digitalization must have been completed. In accounting, for example, there must be invoicing systems [7], automatisatoin for recurring transactions, automatic balance etc.

When examining the organization's digitalization capabilities, including in the context of choosing accounting software solutions, the purpose of the software, its core functionality, its architecture, its use, its training and maintenance should be taken in mind [1].

In the process of digitalization and digital transformation processes of a company, the following questions must be answered:

- Are information systems available and for which activities - production management, accounting and finance, human resources, customer and supplier [6]. Are the business processes digitized? Is there a connection between the different systems? Is there an integrated company information management system (ERP)?
- What are the connections with external systems?
- Are business intelligence and Big Data analytics systems used?
- Are cloud-based systems in use?

The methodology that can be used to do this research and find the answers to these questions is company research, interviews with management and department heads. The results have to be presented to the management of the company using new technologies such as dashboards [6], augmented reality and others.

After finding the answers of the questions above, company should take decision about the degree of required digitalization.

Determining the degree of digitalization required answering the following questions:

1. Which existing business processes will evolve further to successful digitalization?

2. Which business processes will be transformed so they can be successfully digitized?

3. Which business processes need to be created?

Insufficiently optimized business processes represent the biggest obstacle of digitalization in accounting, so in order to answer questions, the management team of the company must determine whether existing business processes are sufficiently optimized to meet the requirements for further management.

When the company's management team is thinking about digitalization the team also should think about integrating with existing systems, because the company will not digitize a single process, but interconnected processes [2].

Accounting handles paper and non-paper documents, since everything involving money should be well documented. Without a system in place, it is difficult to capture and manage all documents. In addition, the business logic throughout accounting should be traced and all business processes - new and changed – should be covered.

The system that supports those activities is the Business Process Management System (BPM).

3. Business Process Lifecycle Tracking Using BPM System

The consistent implementation of the steps and the use of the BPM system should lead to:

- Achieving control over chaotic and difficult processes.
- Creating, mapping, analyzing and improving business processes.
- Perform day-to-day operations more efficiently.
- Realization of larger organizational goals.
- Improve and optimize intricate operations.
- Carefully track individual items as they move through the business process.

The lifecycle of business processes should be traced using BPM system by going through the following steps:

Step 1: Design

In order to design the business process, the process should be broken into tasks. A questionnaire can be used to collect data and process it. At this point, a unique questionnaire (form) for the organization must be created. Should be identified who will be responsible for each task in the business process.

Creating new and modifying existing data and processes in digitalization in accounting

In order to create new and modify existing data or processes in order to better prepare for digitalization, it is necessary to use the following features that each business process management tool has:

- Creating diagrams for visualizing processes
- Form Designer
- Role-control access
- Mobile support

- Administrator functions
- Single Registration
- Integration with existing software systems
- Reports and Analytics
- Opportunities to work with big data
- KPI for calculating efficiency of the business process.

Once the data and business processes that need to be modified are identified, a model is created.

Step 2: Model

A business process model is its description, presented graphically, verbally or in a mixed form. In the model some details such as terms and conditions are fixed, to give a clear idea of the sequence of events and the flow of data used in the business process.

Step 3: Implementation

Implementation of the business process. First, the process is tested with a small group of people and then opened to all users. Access to confidential information should remain restricted.

Step 4: Monitoring

Real-time monitoring of the whole process. Using KPI to identify progress, measure performance, and find problems. At this step the business process should be analyzed.

Step 5: Optimizing

As a result of the previous steps, conclusions are drawn about any changes that need to be made to the business process in order to become more efficient. Steps to improve business processes are being considered.

4. Conclusion

The review clearly shows the possibilities to map fields to search for specific quantitative models of sustainability. These models have to be based on in-depth qualitative analysis of the concepts that make up the modern understanding of corporate sustainability. The discussed above is a model of this kind.

References

1. Milev, P., Opportunities for Implementing Internet Monitoring in Public Organizations, Economic and Social Alternatives, Issue 2, pp. 58 – 69, 2018, ISSN 1314-6556.
2. Kouzmanov S., M. Tsaneva. Impact of Business Problem Characteristics on the Architecture and Specification of Integration. Computer Science and Information Technology, vol. 4, no. 1, pp. 9-14, 2016, ISSN 2331-6063
3. The New IT Operating Model for Digital, CEB analysis, https://www.tribalmind.co/S3/tribalmind-prod/Discoveries/C6AcBPeRhk-Q9b1C-RuWZQ.pdf
4. Pratt. Digitalisation in Accounting. Munich: KPMG, 2017
5. Turchi P. The Digital Transformation Pyramid: A Business-driven Approach for Corporate Initiatives. 2018. https://www.thedigitaltransformationpeople.com/channels/the-case-for-digital-transformation/digital-transformation-pyramid-business-driven-approach-corporate-initiatives
6. Mihova, V., G. Stefanov, M. Marzovanova. Cognos Mobile-dashboards design and implementation technology, 6th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE 2016) UNWE, Sofia, Bulgaria, 2-3 December 2016, pp. 514-519. http://icaictsee.unwe.bg/pastconferences/ ICAICTSEE-2016.pdf.
7. Tsaneva M., A. Murdzheva. Tekhnologiya za sazdavane na fakturi vaz osnova na dinamichni danni za balansirane na uslugite na klientite. Balgarski schetovoditel, pp. 21-23, 2008, ISSN-1310-7186.

Approach for Automatic Identification of Unstructured Data

Yavor Tabov

University of National and World Economy, Sofia, Bulgaria
jtabov@unwe.bg

Abstract. The paper explores issues related to the possibilities for structuring data in web-based information systems. An approach for automatic identification of unstructured data is presented. This approach is practically applied with the help of an experimental prototype. The paper presents the results of this study. In conclusion, the more important role of data structuring approaches in web-based information systems is emphasized.

Keywords. Internet, Unstructured data, html

1. Introduction

Much of the useful data today is in an unstructured form. This leads to the need to develop different approaches to processing this data. This processing aims to present the data in a more structured way. There are many software solutions that aim to structure data in web-based information systems. These software solutions have a specific architecture. This architecture usually encompasses specific data structuring components. These architects are also applicable in the public sector [1, 2]. The software architecture consists mainly of software components, interfaces provided by these components and relations between them in the context of structuring the data in web-based information systems.

In this paper we will look at the result of the application of such a software architecture for web-based data retrieval.

A similar approach is presented by Milev in his research with focus on the issues of development of web scrapping applications [3]. We perceive all web-based sources which do not belong to an organization as external. That is why we can generally accept web-based sources as external sources. Since the external data for business intelligent systems is particularly valuable for analyzing at senior management level [4], we believe the problem of structuring web-based data is of crucial importance for this type of systems. In this context, we will consider an approach for automatic identification of unstructured data with the help of an experimental prototype developed for the purposes of the study.

2. Application of approach of automatic identification of unstructured data

To perform the experiment, it is necessary to be able to functionally adjust the provided data sources from web-based information systems (with key characteristics, such as source name, web address, home page, etc.). View of showing parameters of web-based source is presented in Fig. 1.

To implement this approach, the web pages of selected web sources were studied. Relevant data components are identified in the selected sources. A list of sources with the relevant components is presented in Table 1.

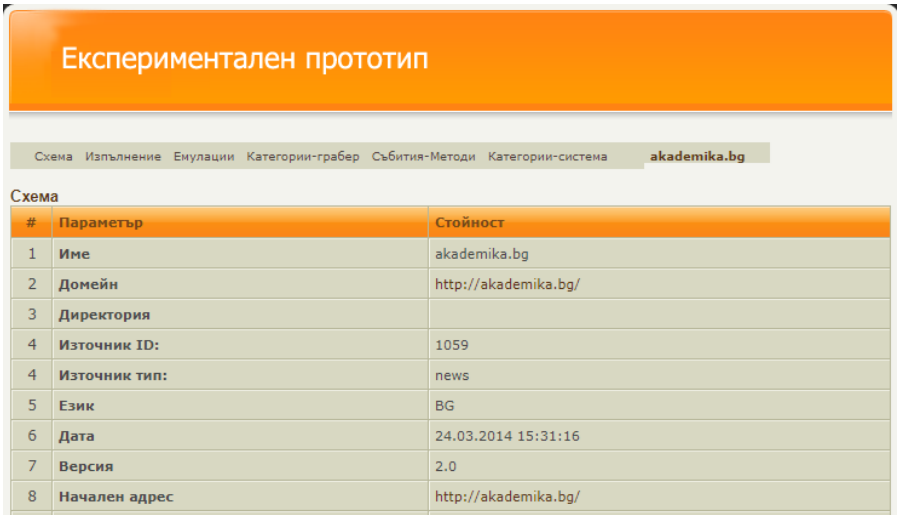


Fig. 1. View of the settings of a web source

Table 1. List of web sources and their identified components

Name	Element	Type
agronovinite.com	div	DataField
alarmanews.com	div	DataField
antenneair.eu	div	DataField
antenneair.eu	li	Target
autoclub.bg	h1	DataField
az-jenata.bg	h2	DataField
az-jenata.bg	li	Target
beer2beer.org	div	Target
bglov.com	div	Record
bgvolleyball.com	tr	Target
big5.bg	div	DataField
blagoevgrad.utre.bg	a	Link
blagoevgrad.utre.bg	div	DataField
blagoevgrad.utre.bg	a	Title
blog.100beers.bg	article	Target
bnr.bg/plovdiv	div	Container
bnr.bg/varna	ul	Container
bnr.bg/varna	div	Target
caralyze.bg	ul	Container
clubz.bg	div	DataField

defakto.bg	ul	Container
defakto.bg	div	DataField
deistvie.bg	div	Target
dnevnik.bg	div	Container
dnevnik.bg	div	DataField
electronics-bulgaria.com	div	Container
electronics-bulgaria.com	div	Target
engineering-review.bg	ul	Container
eratv.bg	div	Container
eratv.bg	div	Target
financebg.com	div	Container
glamour.bg	ul	Container
glasove.com	ul	Container
gradat.bg	ul	Container
it.dir.bg	div	DataField
izvesten.com	table	Container
kliuki.bg	div	Target
mignews.info	div	DataField
mignews.info	img	DataField
money.bg	ul	Container
monitor.bg	a	Category
monitor.bg	a	Link
monitor.bg	article	Record
monitor.bg	h4	DataField
monitor.bg	a	Target
monitor.bg	a	Title
plovdivskinovini.com	div	Record
plovdivskinovini.com	h2	DataField
plovdivskinovini.com	span	DataField
potv.eu	table	Record
presstv.bg	div	Container
presstv.bg	div	Target
radiok2.bg	div	Target
radiomixx.net	div	Target
rusofili.bg	div	Record
sense4style.com	div	Container

skandal.bg	div	Container
skandal.bg	div	Target
sofialive.bg	div	Container
stroiteli.bg	ul	Container
stroiteli.bg	article	Record
stroiteli.bg	img	DataField
stroiteli.bg	time	DataField
stroiteli.bg	h1	DataField
stroiteli.bg	div	DataField
stroiteli.bg	div	Target
stznews.bg	div	Container
tractor.bg	div	Container
tribali.info	div	Container
tribali.info	ul	Container
vevesti.bg	div	Container
zapernik.com	article	Record
zelenite.bg	div	Container

To study the approach for automatic identification of unstructured data, a source is selected, and settings of some basic parameters are set, as shown in Fig. 2.

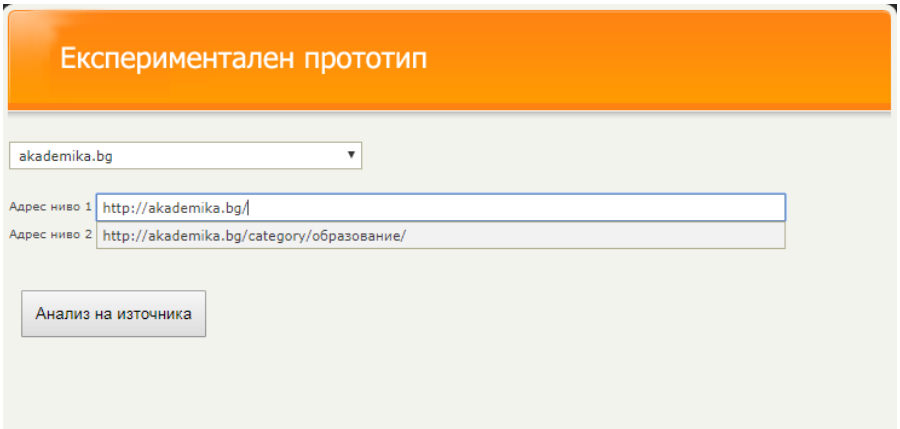


Fig. 2 View of the analyze part of the prototype

Once the source for analysis is selected, random web addresses are set from its pages. Then, a process is initiated in which the prototype, based on the embedded data for the building components described in Table 1, tries to propose at least one template to be recognized in the html code structure of the respective page. The result is presented in Fig. 3.

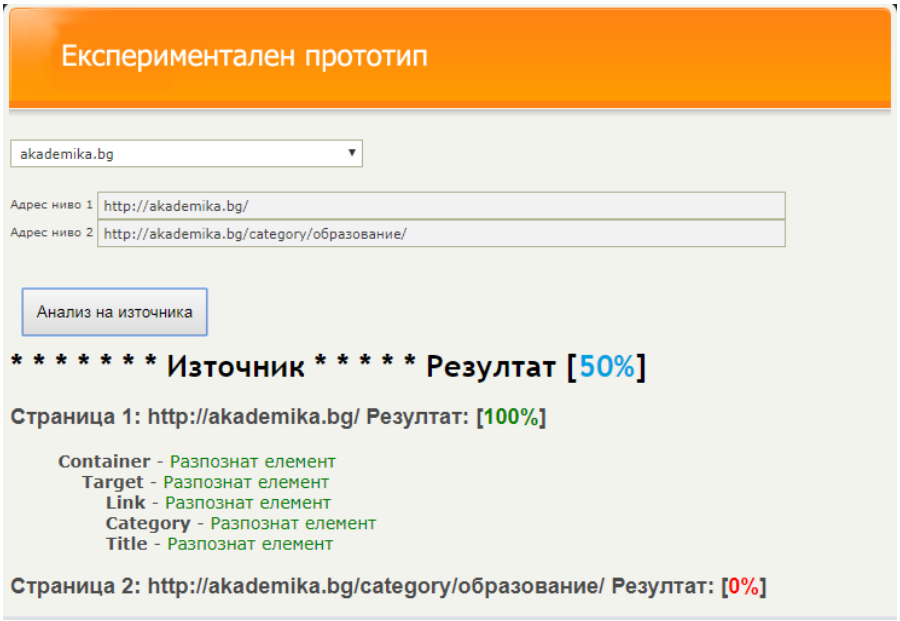


Fig. 3 View of the results of the approach

The result is considered successful if a template is created through which the required sets of certain data can be identified and distinguished.

3. Conclusion

The purpose of this approach is to provide the ability to automatically identify data within a web page. In this context, the presented approach for automatic identification of unstructured data aims to create an algorithm for recognizing fragments of structures within a webpage that makes sense to be extracted. The result of the implementation of the experimental prototype is a plurality of fragments of structures, which includes a description of the extracted elements and their contents in the form of hypertext.

References

1. Kirilova, K., Methodological aspects of administrative service in the public sector, Economic and Social Alternatives, Issue 2, 2018, ISSN 1314–6556.
2. Kirilov, R., Software solutions for managing projects co-financed under the European union's operational programs, Business Management, Issue 3, 2016, ISSN 0861-6604.
3. Milev, P., Conceptual approach for development of web scraping application for tracking information, Economic Alternatives, Issue 3, 2017, ISSN 1312-7462.
4. Stefanova, K., Kabakchieva, D., Business intelligent systems, Publishing complex UNWE, Sofia, 2015.

Application of Concept for Structuring of Web Data

Yavor Tabov

University of National and World Economy, Sofia, Bulgaria
jtabov@unwe.bg

Abstract. The paper explores the possibilities for structuring data in web-based information systems. This topic is relevant because much of the data today is in an unstructured form within web publications on the Internet. The paper presents some concept approaches for structuring this type of data. The study presents the results of research on this topic. In conclusion, some important aspects of the need for structured web data are highlighted.

Keywords. Web, Unstructured data, html

1. Introduction

The aim of the research is to present the results of practical application of the possibilities for structuring data in web-based information systems. For this purpose, a prototype is prepared for structuring data from web-based information systems, which are news websites. Structuring methods should be applied to their content in order to separate the main fragments that make up the data of their published articles, namely web addresses and section names (for example categories "Sports", "Politics", "World", etc.), web addresses from lists of publications to detailed pages (when a list of titles and hyperlinks is presented in the source category), characteristics of news from a detailed page (title, date and text). The structured data is typically managed with a database, but the web data is usually managed by a search engine due to the lack of structures. The search engine allows the user to find useful information from the data collection by using a keyword query. Improving the efficiency and effectiveness of the search engine is a central research topic in the field of information retrieval [1, 3], as many search-related issues such as text clustering, text categorization, summarization, and recommender systems are also a subject of research [2, 7]. Many authors note in their researches that public organizations also use different types of data for the implementation of their services [4, 5]. The two most common techniques for analyzing and retrieving information from the web are text mining and natural language processing (NLP). Some authors also research the opportunities for assessment of information systems in this area [6]. In this regard, we will consider some approaches to structuring data in web-based information systems.

2. Approaches for structuring of web data

For the purposes of the research, it is necessary to define a way to describe the required fragments. An example of such fragments is shown in Fig. 1 – hyperlinks to certain sections of the source (categories). By searching for fragments of a web page we mean searching for certain fragments of the html tree of the page. Thus, in this example, the source categories in question represent html elements of a tree structure and usually, no matter how many, have common characteristics (identical element names, identical attribute names, identical or similar attribute values). For the experimental prototype, it is necessary to describe these

structures in some way. For this purpose, xml-based templates are defined that describe search fragments. The following is a description of these templates, which can be:

- Containers;
- References;
- Container with publication details;
- Sub-elements for description of characteristics.

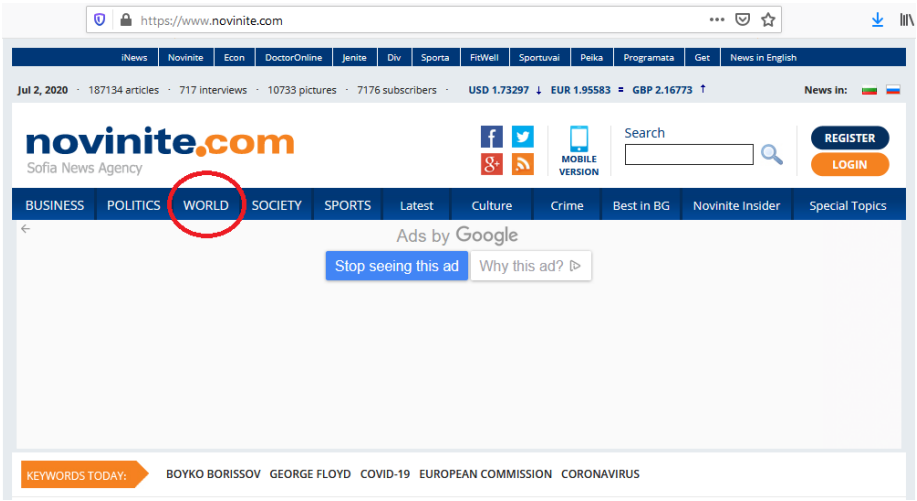


Fig. 1. Example of a fragment of web data

Containers are areas of the page that have some distinctive feature and store the data you are looking for. Fig. 1 shows the area of such a container with an ellipse enclosing a category fragment described above. An exemplary description of this type of elements is presented in Fig. 2.

```
<container objectid="2" pathid="l1-body-0/l2-nav-0" normalsearchmethod="NONE"
advancedsearchmethod="SEW">
</container>
```

Fig. 2. Description of a container type element

This type of element is denoted by the keyword container. References are areas that are usually more than one in number, have similar or identical characteristics, and are located at the same level in the html tree structure. They can be hyperlinks or other html elements that store hyperlinks. They represent sets of web addresses that link to categories of lists or to detailed pages of publications. A description of this type of elements is presented in Fig. 3.

```
<target designation="category" pathid="l1-body-0/l2-nav/l3-a-0" pageid="page1"
modifier="getfirstchild" cycled="True" iter="0" advancedsearchmethod="SBF">
        <link pathid="l1-body-0/l2-nav/l3-a-0" pageid="page1"
advancedsearchmethod="SBF" />
        <title pathid="l1-body-0/l2-nav/l3-a-0" advancedsearchmethod="SBF" />
        <category pathid="l1-body-0/l2-nav/l3-a-0" pageid="page1"
advancedsearchmethod="SBF" />
</target>
```

Fig. 3 Description of reference type elements

This type of element is denoted by the keyword target. It includes sub-elements named link, title, category, which describe the location of key features – hyperlink, title, category name. A publication details container is an area that stores data intended for structuring. Unlike a regular container, the description of this type also includes the description of the sub-elements that are part of it and contain textual data such as title, date, and description. An example of a container with publication details is shown in Fig. 4.

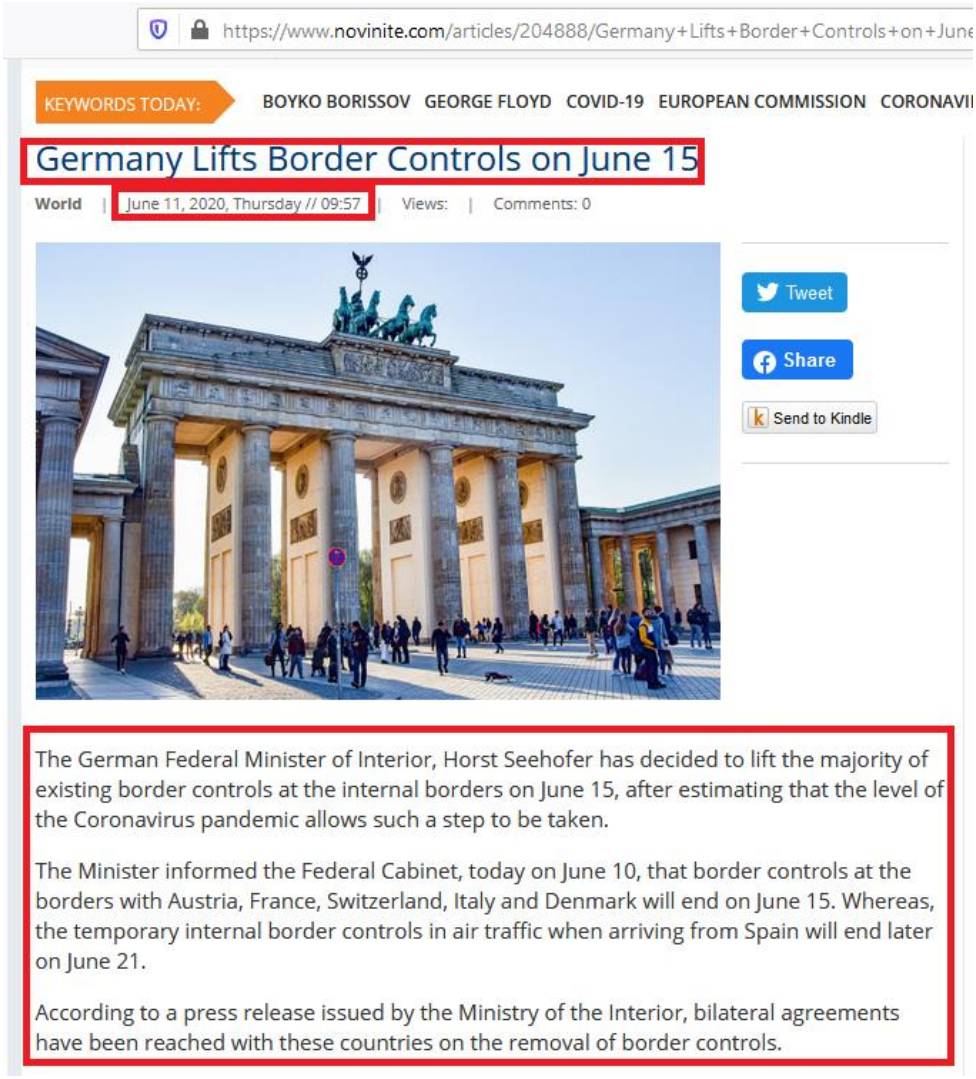


Fig. 4 Container with publication details

A description of this type of elements is presented in Fig. 5.

This type of element is denoted by the keyword record. It includes sub-elements named data field, which use the type attribute to determine what type of data fragments they belong to: title, date, text, hyperlink to an image. Sub-elements for the description of characteristics are denoted by the keyword checker and represent elements that in the tree-like xml structure describe characteristics for their elements - parents (elements from the previous level in the

hierarchy). The one shown in Fig. 6 code describes a container type element that has a class attribute with the text value "news-item".

```
<record objectid="5" pathid="l1-body-0/l2-article-0" normalsearchmethod="NONE"
advancedsearchmethod="SEW" >
        <checker function="AND" type="containattribute class" value="article" />
        <datafield pathid="l1-body-0/l2-article-0/l3-img-0" type="image_url"
reportoptions="noerror" advancedsearchmethod="SEW">
            <checker function="AND" type="containattribute class"
value="article__gallery-image" />
        </datafield>
        <datafield pathid="l1-body-0/l2-article-0/l3-h1-0" type="title"
advancedsearchmethod="SEW">
            <checker function="AND" type="containattribute class" value="page-
header__title" />
        </datafield>
        <datafield pathid="l1-body-0/l2-article-0/l3-div-0/l4-div-1/l5-h4-0" type="subtitle"
advancedsearchmethod="NONE" reportoptions="noerror">
        </datafield>
      <datafield pathid="l1-body-0/l2-article-0/l3-time-0" type="date_published"
advancedsearchmethod="SEW">
        <checker function="AND" type="containattribute class" value="info-
block__count%" />
      </datafield>
      <datafield pathid="l1-body-0/l2-article-0/l3-div-0" type="text"
advancedsearchmethod="SEW">
        <checker function="AND" type="containattribute class" value="article__text" />
      </datafield>
</record>
```

Fig. 5 Description of a container with publication details

```
<container pathid="l1-body-0/l2-nav-0" normalsearchmethod="NONE"
advancedsearchmethod="SEW">
<checker function="AND" type="containattribute class" value="news-item" />
</container>
```

Fig. 6 Sub-element for description of characteristics

3. Conclusion

In conclusion, it should be noted that for the actual implementation of the experiment it is necessary for the experimental prototype to implement the following basic functionalities:

- Ability to access web pages on the web;
- Ability to search for components with data within a web page using the defined templates for searching for components;
- Ability to store found data;
- Possibility for settings and references through user interfaces.

References

1. Baeza-Yates, R., Ribeiro-Neto, B., Modern Information Retrieval - the concepts and technology behind search, Second edition, Pearson Education Ltd., 2011, ISBN: 978-0321416919
2. Grossman A., Frieder, O., Information Retrieval: Algorithms and Heuristics, Springer, 2004, ISBN: 978-8181289179
3. Jones, K., Willett, P., Readings in information retrieval, Morgan Kaufmann Publishers Inc, 1997, ISBN:1-55860-454-5
4. Kirilova, K., Methodological issues in development of public electronic services, Economic and Social Alternatives, Issue 2, 2016, ISSN 1314–6556.
5. Kirilov, R., Features of application of software solutions in public projects risk management, Economic and Social Alternatives, Issue 2, 2017, Sofia, ISSN 1314-6556.
6. Milev, P., An assessment of information systems for the indexation and analysis of online publications, Business Management, Issue 1, 2017, ISSN 2534-8396.
7. Stefanova, K. and Yordanova, S., Knowledge Discovery from Unstructured Data using Sentiment Analysis, Economic and Social Alternatives, Issue 1, 2017, ISSN 1314–6556.

One Opportunity for Forecasting the Total Industrial Production Index (IPI) based on Business Indicators in Industry

Sonia Chipeva[1], Valentin Chavdarov[2]

[1] University of National and World Economy, Sofia, Bulgaria
[2] NSI, Sofia, Bulgaria
soniachip@unwe.bg, VChavdarov@NSI.bg

Abstract. Knowing the economic situation and anticipating the direction of its development in the short term is a key for good managing and planning business. In this context, the information on the changes in the state of the economy and its major sectors, provided by the NSI's short-term business indexes, provides a reliable economic analysis within the relevant frameworks. The short-term forecast of these indices is an essential tool of the current analysis of the economic situation on the one hand and the possibility of publishing early-warning estimates on the other. A range of econometric models have been developed for modeling the changes of the Total Industrial Production Index (IPI) based on key Business Indicators (BI) in Industry - Present business situation, Present production activity, Level of orders, Level of export orders, Level of stocks of finished goods, Expected business situation over the next 6 months, Expected production activity over the next 3 months, Business Situation in Industry. The models provide an opportunity for producing early estimates of the Total IPI comparing with produced values by the NSI within the current technological process. Their role will be rather directed towards the tendency to change the volume of industrial production in order for a timely and adequate response of the business.

Keywords: Total Industrial Production Index (IPI), Business Indicators in Industry, Econometric models, Short-term Forecasting.

1. Introduction

Knowing the economic situation and anticipating the direction of its development in the short term is a key for good managing and planning business. In this context, the information about changes in economy state and its major sectors on short-term, providing by the NSI's short-term business indexes, is an important base for reliable economic analysis within the relevant frameworks. Short-term business indexes express the short-term effects due mainly to seasonal factors as well as long-run tendency in phenomena development. Forecast of these indexes is a useful tool of the economic analysis and enrich the information base for right decision taking.

NSI produces monthly estimates of the short-term business indexes based on representative sampling of business units in the country. Data gathered in the survey comes from firms accounting books and present their real activity. Major problem related to the estimates is the delay of their producing due to the specific of the technological process of the data processing. According the NSI's calendar the estimates of the short-term business indexes are published some time after the end of the relevant period to which they relate. Delay is usually about a month or more. This is a serious obstacle for the current economic analysis, as well as for the short-term forecasting of the phenomena and processes in the economy.

One of the most important short-term business indexes is the Total Industrial Production Index (IPI). On-time estimation of this index provides essential information about current economic situation as well as it might serve for presenting the short-term tendency of the economic development in the country. Early estimate of the index will provide to interested users preliminary information about tendency of the economic situation which in turn will be an additional lever in decision taking process.

Along with the survey for short-term indexes NSI conducts monthly survey among entrepreneurs and managers on the business situation in the country. Data gathered in the survey is summarized in a form of special balance estimates of the respondent opinion and summary indicators for the current and expected business situation. Key feature of information from the survey is that it expresses the judgment of the people that organize and manage business activities but not the results of the activities themselves. Advantage of the indicators for business situation comparing to short-term business indexes is the faster and more simplified process of gathering and processing the collected data. This allows the business indicators to be produced and published earlier of time by 1 month or more, i.e. within the current month they concerned for.

Purpose of this paper is to build econometric model of the Total Industrial Production Index (IPI) based on a range of key Business Indicators (BI) for business situation in the country. The model will provide opportunity for producing estimates of the IPI ahead of time comparing to the estimates produced by NSI within the current technological process. Of course, it should be kept in mind that these are just preliminary estimates of the index preceding the actual ones produced according to currently set methodology. The main role of the early estimates will be rather indicative in respect of the industry production trend in order to timely and adequate response of business.

2. Methodology

Data used in the study spans period of January 2000 to June 2018. Time series of Total Industrial Production Index (IP) and the next Business Indicators are analyzed:

- Current business status in industry (b02)
- Current production activity in industry (b2)
- Order level in industry (b31)
- Level of orders from abroad (b33)
- Expected business situation in industry in the next 6 months (b7)
- Expected production activity in industry in the next 3 months (b8)
- Business climate in industry (bk)

The Total Industrial Production Index is presented by the seasonally adjusted time series published on the NSI site. Time series of Business indicators consist of summary characteristics calculated from the "balances" of respondent opinions. The answers of the most questions are presented on a simplified 3-stage nominal scale usually with categories "*increase*", "*decrease*" or "*without change*" or "*more than sufficient*", "*sufficient*" or "*insufficient*". "Balances" are calculated as differences between proportions of extreme response options. Some of indicators are quantitative estimates as "provision of production with orders (in months)", "load capacity (%)" and etc.

Developing of the econometric model of the Total Industrial Production Index in dependence on a set of key Business Indicators had been preceded by application of range of statistical analysis. First, all the time series were explore for seasonality using Ljung-Box Q-test and Friedman test. Both tests check the null hypothesis for lack of seasonality in the

respective time series. The true null hypothesis has been rejected when significance of the test statistics is smaller than the significance level $\alpha = 0{,}05$. Based on the results, the time series with proven seasonality have been seasonally adjusted using regARIMA models. The regARIMA models consist of 2 components – first one for trend modeling and second one for seasonality modeling. Application of such models enables the seasonal component to be modeled and thus the dynamic of the seasonal effect to be expressed and handled. Specification of the models presents the structure of both components – first part for the trend and second one for the seasonal factor. The general model specification is as follows:

regARIMA (p d q)(P D Q),

where specification in the first brackets concerns the *trend*:

p – number of AR elements in respect of the trend;

d – order of integration (order of differences) in respect of the trend;

q – number of MA elements in respect of the trend;

and specification in the second brackets – *seasonal factor*:

P – number of AR elements in respect of the seasonal factor;

D – order of integration (order of differences) in respect of the seasonal factor;

Q – number of MA elements in respect of the seasonal factor.

Both parts of the regARIMA models are joined multiplicatively. Along with accounting for and modeling the effect of seasonal factors, trading day and holiday effects are taking into account in the models. Models are applied using TRAMO/SEAT methodology recommended by Eurostat.

Augmented Dickey-Fuller (ADF) test has been applied to explore the time series for unit roots presence. The strength of correlation between IPI and business indicators included in the analysis has been explored with Pearson correlation coefficient. Correlation matrix of business indicators has been calculated for detecting collinearity between nominated factors for the econometric models. A set of multiple econometric models of the Total Industrial Production Index (IP) depending on selected factors has been developed. All the estimated models have been analyzed for goodness-of-fit using Fisher's *F*-test and for statistical significance of their parameters applying Student's *t*-test. Durbin-Watson test has been applied for detecting serial correlation of the model residuals. Value of Durbin-Watson statistics around 2 (within interval 1.5 to 2.5) has been used as an empirical proof for lack of serial correlation. Homoscedasticity in the models has been analyzed using ARCH-test, where confirmation of the null hypothesis provides statistical proof for lack of heteroscedasticity. Finally, the best econometric mode of the Total Industrial Production Index (IP) has been defined based on both the largest value of R Square and the smallest value of Akaike Information Criterion (AIC).

Statistical analysis is realized using statistical software JDemetra+ and Eviews.

3. Results and discussion

Results of conducted seasonality tests in the time series are presented in table 1. Seasonal effect is proven for 3 of analyzed time series - *Total Industrial Production Index* (IP), *Current production activity in industry* (b2) and *Expected business situation in industry in the next 6 months* (b7). Seasonality tests are not definitive enough for the *Current business status in industry* but it might be assumed seasonal effect in it. An evidence for this assumption is the presence of seasonal ARIMA component in the specified regARIMA model for the time series (Table1).

Table 1. Results of seasonality tests in the analyzed time series and the specified reg ARIMA models for them

Indicator		Presence of seasonality	regARIMA model specification
Current business status in industry	b02	Not definitive	(011)(011)
Current production activity in industry	b2	Yes	(311)(011)
Order level in industry	b31	No	(110)(000)
Level of orders from abroad	b33	No	(011)(000)
Expected business situation in industry in the next 6 months	b7	Yes	(111)(011)
	b8	No	(100)(000)
Business climate in industry	bk	No	(011)(000)
Total Industrial Production Index	IP	Yes	(011)(011)

Results of the conducted ADF tests show that only 3 of the tested time series are not integrated at 5% significance level - *Current production activity in industry* (b2), *Level of orders from abroad* (b33) and *Expected production activity in industry in the next 3 months* (b8). Time series of *Total Industrial Production Index* is integrated of order 2 and all the rest time series are integrated of order 1.

Correlation coefficients between the *Total Industrial Production Index* and the business indicators nominated as factors for the econometric model are presented in Table 2.

Table 2. Correlation coefficients between the *Total Industrial Production Index* and the business indicators

Indicator		Correlation coefficient - R	p-value of t-test for stat.sig. of R
Current business status in industry	b02	0.7027	0.000
Current production activity in industry	b2	0.3551	0.000
Order level in industry	b31	0.7580	0.000
Level of orders from abroad	b33	0.7226	0.000
Expected business situation in industry in the next 6 months	b7	0.1037	0.123
Expected production activity in industry in the next 3 months	b8	0.0344	0.609
Business climate in industry	bk	0.5366	0.000

In the Table 2 it can be viewed that there is strong correlation (R > 0.7) of the *Total Industrial Production Index* with *Current production activity in industry*, *Order level in industry* and *Level of orders from abroad* and significant correlation (R > 0.5) of the *IP* with *Business climate in industry*. All the correlation coefficients are statistically significant at 5% significance level (see the last col. of Table 2). Factors *Current production activity in industry*, *Expected business situation in industry in the next 6 months* and *Expected production activity in industry in the next 3 months* show very weak, practically lack of correlation, with the *Total Industrial Production Index*. Moreover, both last correlation coefficients are statistically insignificant. One of the reasons for this result might be due to presence of delayed effect of these factors on the *Total Industrial Production Index*. This assumption has been analyzed below by including *lag* variables for both factors in the model.

Analysis of the correlation matrix of the factors nominated for the econometric model shows that there is strong correlation between indicators for the current business situation in the industry (Table 3). Values in the brackets are p-values of the t-tests for statistical significance of the respective correlation coefficients. On the other hand correlation between them and the indicators concern expectations of entrepreneurs is very weak.

Table 3. Correlation matrix of the factors nominated for the econometric model

R / **(p-value)**	**b02**	**b2**	**b31**	**b33**	**b7**	**b8**	**bK**
b02	1.0000						
	-----						
b2	0.7221	1.0000					
	(0.000)	-----					
b31	0.8241	0.6552	1.0000				
	(0.000)	(0.000)	-----				
b33	0.7004	0.5161	0.8979	1.0000			
	(0.000)	(0.000)	(0.000)	-----			
b7	0.5559	0.6398	0.3058	0.1896	1.0000		
	(0.000)	(0.000)	(0.000)	(0.004)	-----		
b8	0.4766	0.6556	0.2715	0.1450	0.7952	1.0000	
	(0.000)	(0.000)	(0.000)	(0.030)	(0.000)	-----	
bk	0.9407	0.7680	0.7104	0.5767	0.7937	0.6698	1.0000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	-----

A set of multiple linear models of the *Total Industrial Production Index* are developed. Estimated models along with their main characteristics are presented in Table 4. Values in the middle brackets below the components of the models are the standard errors of the respective model parameters and the values in the small brackets are p-values of the *t*-tests applied for testing statistical significance of the respective model parameters.

Table 4. Econometric models of the *Total Industrial Production Index* with their main characteristics

№	Model	F–test Prob of F	R Square (%)	AIC
1	IP = *109,2 + 1,22*b02 + *0,26*b31+ *0,26*b33- *1,22*bk [4,4] [0,15] [0,15] [0,12] [0,17] (0,000) (0,000) (0,080) (0,030) (0,000)	115,83 (0,000)	68,10%	7,124
2	IP=*113,5+0,15*b02 *+0,18*b31+*0,22*b33- *0,39*bk+*0,68*ε_{t-1} *+0,21*ε_{t-2} [7,7] [0,15] [0,17] [0,08] [0,19] [0,07] [0,07] (0,000) (0,315) (0,203) (0,007) (0,044) (0,000) (0,001)	187,90 (0,000)	84,11%	6,418
3	IP=*105,4 + 0,23*b33- *0,25*bk+*0,23*b8$_{t-1}$+*0,69*ε_{t-1} *+0,23*ε_{t-2} [7,5] [0,08] [0,13] [0,08] [0,07] [0,07] (0,000) (0,003) (0,062) (0,004) (0,000) (0,000)	228,78 (0,000)	84,30%	6,389
4	IP=*104,2 + 0,26*b31- *0,27*bk+*0,23*b8$_{t-1}$+*0,69*ε_{t-1} *+0,23*ε_{t-2} [8,1] [0,13] [0,14] [0,08] [0,07] [0,07] (0,000) (0,048) (0,059) (0,004) (0,000) (0,000)	222,72 (0,000)	83,94%	6,412
5	IP=*110,1 +0,15*b31+*0,20*b33- *0,30*bk+*0,21*b8$_{t-1}$+*0,68*ε_{t-1} *+0,24*ε_{t-2} [8,3] [0,13] [0,08] [0,14] [0,08] [0,07] [0,07] (0,000) (0,249) (0,013) (0,034) (0,007) (0,000) (0,000)	191,12 (0,000)	84,39%	6,392

Test-statistics values of both tests for diagnostic analysis of the developed models are presented in tab.5. Values in the brackets in the last column of the table are p-values of the respective ARCH tests.

Table 5. Results of diagnostic of the developed models of the *Total Industrial Production Index*

№	Model	DW statistics for serial correlation	ARCH test for homoscedasticity
1	IP = *109,2 + 1,22*b02 + *0,26*b31+ *0,26*b33- *1,22*bk	0,735 Positive s.c.	115,830 (0,000) Heteroscedasticity proven
2	IP=*113,5+0,15*b02 *+0,18*b31+*0,22*b33- *0,39*bk+*0,68*ε_{t-1}+*0,21*ε_{t-2}	2,190 Lack of s.c.	0,627 (0,429) Homoscedasticity proven
3	IP=*105,4 + 0,23*b33- *0,25*bk+*0,23*b8$_{t-1}$+*0,69*ε_{t-1} *+0,23*ε_{t-2}	2,187 Lack of s.c.	0,004 (0,950) Homoscedasticity proven
4	IP=*104,2 + 0,26*b31- *0,27*bk+*0,23*b8$_{t-1}$+*0,69*ε_{t-1} *+0,23*ε_{t-2}	2,189 Lack of s.c.	0,068 (0,794) Homoscedasticity proven
5	IP=*110,1 +0,15*b31+*0,20*b33- *0,30*bk+*0,21*b8$_{t-1}$+*0,68*ε_{t-1}+*0,24*ε_{t-2}	2,197 Lack of s.c.	0,038 (0,844) Homoscedasticity proven

First developed model includes 4 selected business indicators that have shown strongest correlation with the *Total Industrial Production Index*, namely *Current production activity in industry*, *Order level in industry*, *Level of orders from abroad* and *Business climate in industry*. All the model parameters except of *Order level in industry* are proven statistically significant

at 5% significance level. The model fits well empirical data and the included 4 business indicators in it explain 68,10% of the variation of the *Total Industrial Production Index* (see Table 4). Results of the model diagnostic show presence of positive serial correlation and heteroscedasticity in the model residuals. To overcome serial correlation in the model, two auto regression components of the residuals are included in it – of 1st and 2nd order. Durbin-Watson test on the modified model (the second one) shows that there is no more serial correlation. Inclusion of both auto regression components in the model get also the heteroscedasticity of the residuals off – the null hypothesis of ARCH test cannot be rejected, thus it can be conclude that the residuals variation remains constant (see tab.5). Parameters related to the factors *Current production activity in industry* and *Order level in industry* in the new model are statistically insignificant but the explanatory ability of the model increase compare to this of the initial model. In the next developed model (the third one) both factors with insignificant effect on the dependent variable are excluded from the model and the factor *Expected production activity in industry in the next 3 months* is included instead. The new factor is included by a one-lag variable ($b8_{t-1}$). Reason is the assumption for a lag-effect of this factor on the current change of the production activity and respectively on the total production in industry. Restructuring of the model increase, although a bit, its explanatory ability (R^2 = 84,30%) and also there is statistical evidences for lack of serial correlation and heteroscedasticity in it. As a result of the change in model specification, the parameter related to the *Business climate in industry* becomes statistically insignificant at 5% significance level but it is statistically significant at significance level higher that 6,2%.

One more model of the *Total Industrial Production Index* is developed in which the factor *Level of orders from abroad* is exchanged with factor *Order level in industry*. The change is made due to that both factors are strongly correlated with the *Total Industrial Production Index* and even the correlation with the *Order level in industry* is a bit stronger. At the same time both factors are strongly correlated in between too. That is why one of them was included in the model first and then it was replaced with another one. Model estimation results do not differ significantly from these ones of the previous model. Statistical significance of the parameter related to the *Business climate in industry* in this model is proven at error risk of more than 5,9%. The rest parameters in the model are statistically significant at 5% significance level and goodness-of-fit of the model is proven too. The explanatory ability of both models is almost the same with light predominance of the model that includes *Level of orders from abroad*.

As it was mentioned above, both factors related to the orders in industry are strongly correlated with the *Total Industrial Production Index*. Based on this, although the strong correlation in between both factors, they are included together in the last developed model. Along with them the factor *Business climate in industry* and the lag-variable of factor *Expected production activity in industry in the next 3 months* are included in the last model. This model shows the highest explanatory ability compare to the rest developed models. All the parameters in the model except this one related to *Order level in industry* are statistically significant. The model fits well empirical data and diagnostic analysis shows that there is no serial correlation and heteroscedasticity of the residuals.

4. Conclusion

Developed econometric models of the *Total Industrial Production Index* might explain significant part of index variation based on a selected set of business indicators. Thus, extrapolation of the models will provide reliable forecast of the *Total Industrial Production Index* through the relevant business indicators. Important advantage of the forecast based on these models compared to the trend models or ARIMA is that it takes into consideration effects

of the current economic situation through the business indicators included in the model while in the forecast based on trend models or ARIMA only the effects of the economic conjunction in the past are taken into consideration.

Results of the econometric analysis of the developed models show that the third and the last model are the best with the last one showing highest explanatory ability (84,39%). The last model differs from the third one only by one more factor included in it, namely *Order level in industry*. Based on the results it can be concluded that the model of the Total Industrial Production Index:

IP=110,1 0,15b31 0,20b33- 0,30bk 0,21b8t-1 0,68εt-1 0,24εt-2

is the best one and can provide reliable forecast for the Index that is expected to be relevant the current business situation.

References

1. Gujarati D., Basic Econometrics, 4th ed., McGraw-Hill, 2004
2. Chipeva, S., 2001. Opportunity for using data from studies of business situation for purposes of the economic analysis. International conference "Management and Marketing Aspects in Balkan Economic Development", 19-22.IX.2001, Kranevo, Proceeding, 371-388.
3. Chipeva, S., 2001. Statistical Analysis of the business indicators susceptibility to the industry Firms Difficulties in Bulgaria. Intern. Conference "Management and Marketing Aspects in Balkan Economical Development" in Economic Themes, year XXXIX, v.3.,2001, Nish.
4. Chipeva,S. 2003. Potentiality of Using the Business Indicators in the Econometric Analysis for Business Development Studies. *Intern.Conference "Management and Marketing Aspects in Balkan Economic Development",* Ravda, 11-14.IX.2002, *in Proceeding*, 2003, Univer.Press "Stopanstvo", Sofia, 158-166.
5. https://www.nsi.bg/sites/default/files/files/metadata/businesssurveys-metho.pdf
6. https://infostat.nsi.bg/infostat/pages/module.jsf?x_2=33

A Private Case of Conjecture2

Todor Stoyanov Stoyanov

Department of Mathematics, University of Economics, Varna, Bulgaria
todstoyanov@yahoo.com

Abstract. In this paper we prove the Conjecture2 when n=1and n=2. Conjecture1 and Conjecture2 are defined in Introduction. Besides Conjecture1 is already proved.

Keywords. Derivative, Integral, Complex polynomial, Zeros.

1. Introduction

In papers [4-5], we consider the Conjecture1:

If $a_k \geq 0,\ a_k \in \mathbb{R}$. Then we assert

$$\left| \int_0^{e^{i\varphi}} \prod_{k=1}^{n} (x + a_k) dx \right| \geq \frac{1}{n+1},$$

for arbitrary natural n, $\varphi \in [0, \pi/2]$.

Here we present some new results about Conjecture2:

If $\phi_k \in [\pi/2, \pi]$, then $\left| \int_{-1}^{0} (x+1) \prod_{k=1}^{n} (x - e^{i\phi_k}) dx \right| \geq \frac{1}{n+2}$.

Both conjectures are very important for the proofs of some famous conjectures, like Sendov's and Obreshkoff's ones. A possible connection between both conjectures appears [10].

The results related with the Conjecture1, we observe in Theorem1, Theorem 2. In Theorem 3 we generalize and prove some new results of this conjecture.

We can see the results of Theorem1 in [4, 5]. Such one of Theorem2 could be seen in [6].

The main results are relevant to Conjectue2, when n=1 and n=2.

Many authors use some modulus of some integrals in the complex plane for various estimates in their works. For example we can see how Bojanov and Rahman in [3] use this method. These estimates are explored for the localization of the zeros of some polynomials.

The results are useful in the (open) problems of [10-13].

2. Related results

Theorem 1: *Let* $k=1,2,...,n$, $n\in N$, $a_k,\varphi\in R$, $a_k\in[0,1],.\ \varphi\in\left[0,\frac{\pi}{2}\right]$. *Then the function*

$$\left|\int_0^{e^{i\varphi}} x\prod_{k=1}^{n}(x+a_k)dx\right|\geq\frac{1}{n+2}\quad\text{for}\quad n=1,2,3\,.$$

Theorem 2: *Let* $k\in N$, $a\in R$, $a\in[0,1]$, $a\in[0,1]$. *Then the function*

$$\left|\int_0^{i} x(x+a)^k dx\right|\geq\frac{1}{k+2}\,.$$

Theorem 3. *Let* $a,b\in R$, $0\leq a\leq b$. *Then the function* $f(a,b)=\left|\int_0^{i} x(x+a)(x+b)dx\right|$ *satisfies the inequality*

$$f(a,b)\leq\frac{1}{2}|a+i|\left|\int_0^{i}(x+b)dx\right|.$$

3. Main results

Theorem1. *If* $n=2$ *then the Conjecture 2 is true, i.e.* $\left|\int_0^1 t^m(t-a)dt\right|\geq\frac{1}{m+2}$, *where* $a=2\cos\alpha e^{i\alpha}$, $\alpha\in\left[\frac{\pi}{3},\frac{\pi}{2}\right]$.

Proof. $I=\int_0^1 t^m(t-a)dt=\int_0^1 t^{m+1}dt-a\int_0^1 t^m dt=$

$$=\left(\frac{t^{m+2}}{m+2}-\frac{at^{m+1}}{m+1}\right)\Bigg|_0^1=\frac{1}{m+2}\left(1-\frac{m+2}{m+1}a\right).$$

$$\left|1-\frac{m+2}{m+1}a\right|^2=\left(1-2k\cos^2\alpha\right)^2+\left(2k\sin\alpha\cos\alpha\right)^2=1-4k\cos^2\alpha+4k^2\cos^2\alpha$$

where $k=\frac{m+2}{m+1}$.

Then $|I| \geq \frac{1}{m+2}$, which confirms the theorem.

Theorem2. *If* $n=2$ *then the Conjecture 2 is true, i.e.* $\left|\int_0^1 t^m (t-a)^2 dt\right| \geq \frac{1}{m+3}$,

where $a = 2\cos\alpha e^{i\alpha}$, $\alpha \in \left[\frac{\pi}{3}, \frac{\pi}{2}\right]$.

Proof. $I = \int_0^1 t^m (t^2 - 2at + a^2)dt = \int_0^1 t^{m+2} dt - 2a\int_0^1 t^{m+1} dt + a^2 \int_0^1 t^m dt =$

$$= \frac{1}{m+3} - \frac{2a}{m+2} + \frac{a^2}{m+1} = \frac{1}{m+3}\left(1 - 2\frac{m+3}{m+2}a + \frac{m+3}{m+1}a^2\right).$$

If we put $k = \frac{m+3}{m+2}$, $k_1 = \frac{m+3}{m+1}$, we have:

$$I = \frac{1}{m+3}\left(1 - 2ka + k_1 a^2\right) = \frac{1}{m+3}\left(1 - 2ka + k^2 a^2 - k^2 a^2 + k_1 a^2\right) =$$

$$= \frac{1}{m+3}\left[(1-ka)^2 + (k_1 - k)a^2\right].$$

But according to the proof of Theorem 1

$$|1-ka|^2 = 1 - 4k\cos^2\alpha + 4k^2\cos^2\alpha .$$

$$\left|(k_1 - k)a^2\right| = \left[\frac{m+3}{m+1} - \left(\frac{m+3}{m+2}\right)^2\right].4\cos^2\alpha = \frac{(m+3).4\cos^2\alpha}{(m+1)(m+2)^2}$$

$$4k^2\cos^2\alpha - 4k\cos^2\alpha = 4\cos^2\alpha.k(k-1) = = \frac{(m+3).4\cos^2\alpha}{(m+2)^2} .$$

Then we obtain that

$$|1-ka|^2 \geq 1 + \frac{(m+3)}{(m+2)^2}4\cos^2\alpha \geq 1 + \frac{(m+3).4\cos^2\alpha}{(m+1)(m+2)^2} = 1 + \left|(k_1 - k)a^2\right| .$$

This

$$|I| \geq \frac{1}{m+3}\left[|1-ka|^2 - \left|(k_1 - k)a^2\right|\right] \geq$$

$$\geq \frac{1}{m+3}\left[1 + \left|(k_1 - k)a^2\right| - \left|(k_1 - k)a^2\right|\right] = \frac{1}{m+3},$$

which confirms the assertion.

Acknowledgment

This research was financed from University of Economics of Varna research grants No.19 2018- 04-27.

References

1. Georgiev G. H., Constructions of Rational Surfaces in the Three-dimensional Sphere, *AIP Conference Proceedings*, Applications of Mathematics inEngineering and Economics(AMEE'11), Editors George Venkov, RalitzaKovachevaand, VeselaPasheva, American Institute of Physics, Melville, NY, vol. 1410, pp. 213-220, 2011.
2. Georgiev G. H., *Geometric Transformations for Modeling of Curves and Surfaces*, Shaker Verlag, Aachen, 2012.
3. G. H. Georgiev, Shape Curvatures of Planar Rational Spirals, In: *Curves and Surfaces*, LNCS, Vol. 6920, Springer, pp. 270-279, 2012.
4. Soyanov T., Inequalities of the modulus of some complex integrals, *Theory and Practice of modern science*, VII international workshop, Moskwa, V.1, pp. 10-16, 2012.
5. Soyanov T., Some estimates of the modulus of some complex integrals using division of polynomials, IV international scientific conference «*Tendencies and prospects of development of modern scientific knowledge*», Moskwa, pp. 11-15, 2012.
6. Soyanov T., "Some estimates below the modulus of integrals in the complex plane", V-th international scientific conference *«Tendencies and prospects of development of modern scientific knowledge»*, Moskwa, V.1, pp. 12-16, 2012.
7. Bojanov B., Q. I. Rahman and J. Szynal, On a conjecture of Sendov about the critical points of a polynomial. "Math.Z" 190, pp. 281-285, 1985.
8. . Souroujon D. M, T. Stoyanov, About the Primitive Polynomials of Polynomials with Real Zeros. *Journal of Analysis and Applications*, Vol. 14, No 1, pp. 21-31, 2016.
9. Stoyanov T., Some new methods for the estimates of the modulus of some integrals of the unit circle in the complex plane. *Journal of Analysis and Applications*, Vol. 17, No.2, pp.119-130, 2019, ISSN:0972-5954.
10. Sourouon D., T. Zapryanova, On the relation between the number of real and complex zeros of polynomials of a certain kind, *AIP Conference proceeding*, vol. 2159, 030032, 2019.
11. Zapryanova T., D Souroujon, On the Iterates of Jackson Type Operator Gs,n, Mediterr. *J. Math*. Vol.13, pp. 5053-5061, 2016.
12. Zapryanova T., Best approximation and Moduli of Smoothness, *Pliska Studia Mathematica Bulgarica*, Vol.21, No.1, pp. 299-306, 2012,
13. Zapryanova T., Tachev G., Generalized Inverse Theorem for Schoenberg Operator, *JMMF*, Vol.1 Issue 2, pp. 11-16, 2012.

Some Estimates Which Are Needed for the Cardioid Interiorities

Todor Stoyanov Stoyanov

Department of Mathematics, University of Economics, Varna, Bulgaria
todstoyanov@yahoo.com

Abstract. In this paper we make some estimates, which are necessary for the localization of the zeros of some derivatives of some complex polynomials. Most of them appear cardioid interiorities.

Keywords. Local extremum, Global extremum, Hessian, Zeros.

1. Introduction

The localization of the zeros of the complex polynomials is very important area of the mathematics. The impossibility to find the zeros of any polynomials using the coefficients makes every statement here very significant. There exist many conjectures which are not proved, like Sendov's conjecture, Obreshkoff's conjecture. The assertions localize the zeros of the derivative of the any complex polynomial in some areas. Here we present some new results about the estimates, which are needed for the localization of the zeros of some polynomials in some areas, like cardioids or generalized cardioid interiors (Theorem2). These results could be applied in [13-16].

2. Related Results

Lemma 1. *If* $x, y \in [0,1]$, *then the function* $h(x,y) = x\sqrt{1-y^2} + y\sqrt{1-x^2} \leq 1$.

Proof. If we fix $y \in [0,1]$**,** then

$$t(x) = h(x,y) = x\sqrt{1-y^2} + y\sqrt{1-x^2}.$$

$$\frac{\partial t}{\partial x} = \sqrt{1-y^2} - \frac{xy}{\sqrt{1-x^2}} = \frac{\sqrt{1-x^2-y^2+x^2y^2} - xy}{\sqrt{1-x^2}},$$

where $x \neq 1$.

The function $l(x) = \sqrt{1-x^2-y^2+x^2y^2} - xy = 0$, when $1-x^2-y^2 = 0 \Leftrightarrow x = \sqrt{1-y^2}$. We know

$$l(x) > 0 \text{ with } x < \sqrt{1-y^2},\ l(x) < 0 \text{ with } x > \sqrt{1-y^2}.$$

$$l(0) = \sqrt{1-y^2} > 0,\quad l(1) = -y < 0.$$

Then $h(x,y) \le t\left(\sqrt{1-y^2}\right) = 1 - y^2 + y^2 = 1$, which confirms the lemma.

Theorem 1. *Let* $x, y, z \in [0,1]$. *Then the function* $f(x,y,z) = xy\sqrt{1-z^2} + yz\sqrt{1-x^2} + zx\sqrt{1-y^2}$ *satisfies* $f(x,y,z) \le \frac{2\sqrt{3}}{3}$.

Proof. Let us calculate

$$\frac{\partial f}{\partial x} = y\sqrt{1-z^2} + z\sqrt{1-y^2} - \frac{xyz}{\sqrt{1-x^2}}$$

$$\frac{\partial f}{\partial y} = x\sqrt{1-z^2} + z\sqrt{1-x^2} - \frac{xyz}{\sqrt{1-y^2}}$$

$$\frac{\partial f}{\partial z} = x\sqrt{1-y^2} + y\sqrt{1-x^2} - \frac{xyz}{\sqrt{1-z^2}}$$

We want $x \neq 1$, $y \neq 1$, $z \neq 1$. We need to solve the system

$$\begin{cases} y\sqrt{1-z^2} + z\sqrt{1-y^2} - \dfrac{xyz}{\sqrt{1-x^2}} = 0 \\ x\sqrt{1-z^2} + z\sqrt{1-x^2} - \dfrac{xyz}{\sqrt{1-y^2}} = 0 \\ y\sqrt{1-x^2} + x\sqrt{1-y^2} - \dfrac{xyz}{\sqrt{1-z^2}} = 0 \end{cases}$$

If $x \neq 0$, $y \neq 0$, $z \neq 0$, the system is

$$\begin{cases} \dfrac{\sqrt{1-z^2}}{z} + \dfrac{\sqrt{1-y^2}}{y} = \dfrac{x}{\sqrt{1-x^2}} \\ \dfrac{\sqrt{1-z^2}}{z} + \dfrac{\sqrt{1-x^2}}{x} = \dfrac{y}{\sqrt{1-y^2}} \\ \dfrac{\sqrt{1-x^2}}{x} + \dfrac{\sqrt{1-y^2}}{y} = \dfrac{z}{\sqrt{1-z^2}} \end{cases}$$

We put $\frac{\sqrt{1-x^2}}{x} = p$, $\frac{\sqrt{1-y^2}}{y} = q$, $\frac{\sqrt{1-z^2}}{z} = s$ $(*)$.

Then we have

$$\begin{cases} q + s = \dfrac{1}{p} & (1) \\ s + p = \dfrac{1}{q} & (2) \\ p + q = \dfrac{1}{s} & (3) \end{cases}$$

After the subtraction (1) – (2), we get $q + \frac{1}{q} = p + \frac{1}{p}$, i.e. $p = q$ or $p = \frac{1}{q}$.

Analogously $q = s$ or $q = \frac{1}{s}$, and $s = p$ or $s = \frac{1}{p}$.

Because of $(*)$, the only possibility is $p = q = s$ and therefore $p + p = \frac{1}{p}$, i.e. $p = \sqrt{\frac{1}{2}}$, i.e. $x = y = z = \sqrt{\frac{2}{3}}$.

The critical point is $M\left(\sqrt{\frac{2}{3}}, \sqrt{\frac{2}{3}}, \sqrt{\frac{2}{3}}\right)$.

We have:

$$\frac{\partial^2 f}{\partial x^2} = \frac{-yz}{(1-x^2)^{3/2}}, \quad \frac{\partial^2 f}{\partial y^2} = \frac{-zx}{(1-y^2)^{3/2}}, \quad \frac{\partial^2 f}{\partial z^2} = \frac{-xy}{(1-z^2)^{3/2}}$$

$$\frac{\partial^2 f}{\partial x \partial y} = \sqrt{1-z^2} - \frac{zy}{\sqrt{1-y^2}} - \frac{zx}{\sqrt{1-x^2}},$$

$$\frac{\partial^2 f}{\partial y \partial z} = \sqrt{1-x^2} - \frac{xy}{\sqrt{1-y^2}} - \frac{xz}{\sqrt{1-z^2}},$$

$$\frac{\partial^2 f}{\partial z \partial x} = \sqrt{1-y^2} - \frac{yx}{\sqrt{1-x^2}} - \frac{zy}{\sqrt{1-z^2}}.$$

Then for the matrix

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix},$$

we get

$$H(M) = \begin{pmatrix} -2\sqrt{3} & -\sqrt{3} & -\sqrt{3} \\ -\sqrt{3} & -2\sqrt{3} & -\sqrt{3} \\ -\sqrt{3} & -\sqrt{3} & -2\sqrt{3} \end{pmatrix} = +\sqrt{3}\begin{pmatrix} -2 & -1 & -1 \\ -1 & -2 & -1 \\ -1 & -1 & -2 \end{pmatrix}.$$

The determinants:

$$D_1 = |-2| = -2 < 0, \quad D_2 = \begin{vmatrix} -2 & -1 \\ -1 & -2 \end{vmatrix} = 3 > 0, \quad D_3 = \begin{vmatrix} -2 & -1 & -1 \\ -1 & -2 & -1 \\ -1 & -1 & -2 \end{vmatrix} = -4 < 0.$$

That means: M is a local maximum $f(M) = \frac{2\sqrt{3}}{3}$. On the plane z=0 we have r(x,y)=f(x,y,0)=xy≤ 1.

On the plane z=1 we have h(x,y)=f(x,y,1)= $x\sqrt{1-y^2} + y\sqrt{1-x^2} \leq 1$, according to the Lemma. Because f is a symmetric function and defined in the cube $[0,1]^3$, therefore $f \leq \frac{2\sqrt{3}}{3}$.

Corollary 1. *If* $\alpha, \beta, \gamma \in \left[0, \frac{\pi}{2}\right]$, *then the function*

$$g(\alpha, \beta, \gamma) = \sin\alpha \sin\beta \cos\gamma + \sin\alpha \sin\gamma \cos\beta + \sin\beta \sin\gamma \cos\alpha \leq \frac{2\sqrt{3}}{3}.$$

Proof. If in the conditions of Theorem 6 we put $x = \sin\alpha$, $y = \sin\beta$, $z = \sin\gamma$, and we obtain that

$$g(\alpha, \beta, \gamma) = f(x, y, z) \leq \frac{2\sqrt{3}}{3}.$$

Theorem 2. *If all the zeros of the polynomial* $p(z) \in \mathbb{C}[z]$ *are* $z_k, k = 1,2, \ldots, n$. *Then for each zero* v *of the third derivative* $p'''(z)$ *exists some* $k_0 \in \mathbb{N}$, $1 \leq k_0 \leq n$, *such that* $v \in \overline{C_q\left(\frac{z_{k_0}}{2}, \frac{|z_{k_0}|}{2}\right)}$, *where* $q = \frac{2\sqrt{3}}{3}$.

3. Main Results

Lemma 2. If $x, y \in R$ **,** then

$$\det\begin{pmatrix} x & y & y & \cdots & y \\ y & x & y & \cdots & y \\ y & y & x & \cdots & y \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y & y & y & \cdots & x \end{pmatrix} = (x-y)^{n-1}\left[x + (n-1)y\right].$$

Let us consider the function

$$f(x, y, z, t) = xyz\sqrt{1-t^2} + xyt\sqrt{1-z^2} + xzt\sqrt{1-y^2} + yzt\sqrt{1-x^2},$$

which is defined in $[0,1]^4$.

Let us calculate:

$$\frac{\partial f}{\partial x} = yz\sqrt{1-t^2} + yt\sqrt{1-z^2} + zt\sqrt{1-y^2} - \frac{xyzt}{\sqrt{1-x^2}},$$

$$\frac{\partial f}{\partial y} = xz\sqrt{1-t^2} + xt\sqrt{1-z^2} + zt\sqrt{1-x^2} - \frac{xyzt}{\sqrt{1-y^2}},$$

$$\frac{\partial f}{\partial z} = xy\sqrt{1-t^2} + xt\sqrt{1-y^2} + yt\sqrt{1-x^2} - \frac{xyzt}{\sqrt{1-z^2}},$$

$$\frac{\partial f}{\partial t} = xy\sqrt{1-z^2} + xz\sqrt{1-y^2} + yz\sqrt{1-x^2} - \frac{xyzt}{\sqrt{1-t^2}}.$$

If we put $\frac{\sqrt{1-x^2}}{x}=a$, $\frac{\sqrt{1-y^2}}{y}=b$, $\frac{\sqrt{1-z^2}}{z}=c$, $\frac{\sqrt{1-t^2}}{t}=d$, the we have:

$\frac{\partial f}{\partial x}=0$, $\frac{\partial f}{\partial y}=0$, $\frac{\partial f}{\partial z}=0$, $\frac{\partial f}{\partial t}=0$, is

$$\begin{cases} a+b+c=\frac{1}{d} & (1)\\ a+b+d=\frac{1}{c} & (2)\\ a+c+d=\frac{1}{b} & (3)\\ b+c+d=\frac{1}{a} & (4) \end{cases}.$$

We subtract (1) of (2) we get $c-d=\frac{1}{d}-\frac{1}{c}$, i.e. $(c-d)\left(1-\frac{1}{cd}\right)=0$.

If $c=\frac{1}{d}$, then $a+b=0$ which is impossible.

Then we get $a=b=c=d$, i.e. $3a=\frac{1}{a}$, i.e. $a=\sqrt{\frac{1}{3}}$, because f in $[0,1]^4$.

Therefore $x=y=z=t=\frac{\sqrt{3}}{2}$.

Therefore critical points is $M\left(\frac{\sqrt{3}}{2},\frac{\sqrt{3}}{2},\frac{\sqrt{3}}{2},\frac{\sqrt{3}}{2}\right)\in[0,1]^4$.

$$\frac{\partial^2 f}{\partial x\partial y}=z\sqrt{1-t^2}+t\sqrt{1-z^2}-\frac{yzt}{\sqrt{1-y^2}}-\frac{xzt}{\sqrt{1-x^2}}$$

$$\frac{\partial^2 f}{\partial x\partial z}=y\sqrt{1-t^2}+t\sqrt{1-y^2}-\frac{yzt}{\sqrt{1-z^2}}-\frac{xyt}{\sqrt{1-x^2}}$$

$$\frac{\partial^2 f}{\partial x\partial t}=z\sqrt{1-y^2}+y\sqrt{1-z^2}-\frac{yzt}{\sqrt{1-t^2}}-\frac{xyt}{\sqrt{1-x^2}}$$

$$\frac{\partial^2 f}{\partial x^2}=-\frac{yzt}{\left(1-x^2\right)^{3/2}}$$

Analog we calculate $\frac{\partial^2 f}{\partial y\partial z}$, $\frac{\partial^2 f}{\partial y\partial t}$, $\frac{\partial^2 f}{\partial z\partial t}$. Then the hessian

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial x \partial t} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} & \frac{\partial^2 f}{\partial y \partial t} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} & \frac{\partial^2 f}{\partial z \partial t} \\ \frac{\partial^2 f}{\partial t \partial x} & \frac{\partial^2 f}{\partial t \partial y} & \frac{\partial^2 f}{\partial t \partial z} & \frac{\partial^2 f}{\partial t^2} \end{pmatrix}$$

is

$$H(M) + \sqrt{3} = \begin{pmatrix} -3 & -1 & -1 & -1 \\ -1 & -3 & -1 & -1 \\ -1 & -1 & -3 & -1 \\ -1 & -1 & -1 & -3 \end{pmatrix}.$$

According to Lemma 1 we have:

$$\det(-3) = -3 < 0,$$

$$\det\begin{pmatrix} -3 & -1 \\ -1 & -3 \end{pmatrix} = 8 > 0,$$

$$\det\begin{pmatrix} -3 & -1 & -1 \\ -1 & -3 & -1 \\ -1 & -1 & -3 \end{pmatrix} = -20 < 0,$$

$$\det\begin{pmatrix} -3 & -1 & -1 & -1 \\ -1 & -3 & -1 & -1 \\ -1 & -1 & -3 & -1 \\ -1 & -1 & -1 & -3 \end{pmatrix} = 48 > 0.$$

That means: f has local maximum in $M - f(M) = \frac{3}{4}\sqrt{3}$.

On the plane $t = 0$ we have

$$f(x, y, z, 0) = xyz \le 1.$$

On the plane $t = 1$ we have

$$f(x, y, z, 1) = xy\sqrt{1 - z^2} + yz\sqrt{1 - x^2} + xt\sqrt{1 - y^2} \le \frac{2}{3}\sqrt{3}.$$

according to Lemma 2.

Because f is a symmetries function, which is defined in $[0,1]^4$, we can formulate the next theorem:

Theorem 3. If $f:[0,1]^4 \to R$,

$f(x,y,z,t) = xyz\sqrt{1-t^2} + xyt\sqrt{1-z^2} + xzt\sqrt{1-y^2} + yzt\sqrt{1-x^2}$,

then $f(x,y,z,t) \le \frac{3}{4}\sqrt{3}$.

Corollary 2. If $\alpha,\beta,\gamma,\delta \in \left[0,\frac{\pi}{2}\right]$, then the function

$$g(\alpha,\beta,\gamma,\delta) = \sin\alpha\sin\beta\sin\gamma\cos\delta + \sin\alpha\sin\beta\sin\delta\cos\gamma +$$
$$+ \sin\alpha\sin\gamma\sin\delta\cos\beta + \sin\beta\sin\gamma\sin\delta\cos\alpha \le \frac{3}{4}\sqrt{3}.$$

Proof. In the conditions of the Theorem we put $x = \sin\alpha$, $y = \sin\beta$, $z = \sin\gamma$, $t = \sin\delta$. Then we obtain $g(\alpha,\beta,\gamma,\delta) = f(x,y,z,t) \le \frac{3}{4}\sqrt{3}$.

Acknowledgment

This research was financed from University of Economics of Varna research grants No.19 2018- 04-27.

References

1. D. M. Souroujon, T. D.M. Simeunovic, On The Location Of The Zeros Of Polynomials, *Math. Morav*., 2 (1998), 91–96.
2. Stoyanov, About the Primitive Polynomials of Polynomials with Real Zeros. *Journal of Analysis and Applications*, Vol. 14 (2016). No 1, pp. 21-31.
3. F. Kittaneh, Bounds for the Zeros of Polynomials from Matrix Inequalities, *Arch. Math*., 81(5) (2003), 601–608.
4. H. Alzer, Bounds for the Zeros of Polynomials, Complex Variables, *Theory Appl.*, 8(2) (2001), 143–150
5. M. Dehmer, Schranken für die Nullstellen Komplexwertiger Polynome, *Die Wurzel*, 8 (2000), 182–185
6. M. Marden, Geometry of Polynomials, Mathematical Surveys. *Amer. Math. Soc*., 3 (1966)
7. N. Obreschkoff, Verteilung Und Berechnung Der Nullstellen Reeller Polynome, *Hochschulbücher Für Mathematik*. 55. Berlin, Veb Deutscher Verlag Der Wissenschaften, (1963)
8. P. Henrici, Applied And Computational Complex Analysis. Volume I, *Wiley Classics Library Edition*, (1988).
9. Q. I. Rahman, G. Schmeisser, Analytic Theory Of Polynomials. Critical Points, Zeros And Extremal Properties, *Clarendon Press*, (2002)
10. T. Stoyanov, "Some Extensions Of Rolle's And Gauss-Lucas Theorems", In: Second International Workshop, *Transform Methods And Special Functions* (Varna, August 23-30.1996).
11. T. Stoyanov, About The Zeros Of Some Entire Functions And Their Derivatives, *Journal Of The Australian Mathemanovtical Society* 68 (2000), 165-169.
12. T. Stoyanov, Some Localization of the Zeros of the Derivatives of a Complex Polynomial in the Disks or Cardioid Interiorities, *Int. J. of Math. Analysis* Vol. 9, 2015, no. 58, 2849-2855 HIKARI Ltd
13. T. Zapryanova, On the Iterates of Jackson Type Operator Gs,n, Mediterr. *J. Math*. 13 (2016) 5053-5061
14. T. Zapryanova, Best approximation and Moduli of Smoothness, *Pliska Studia Mathematica Bulgarica*, Vol.21, No1, (2012), 299-306.
15. D. Souroujon, T. Zapryanova, On the relation between the number of real and complex zeros of polynomials of a certain kind, *AIP Conference Proceeding*, 2159, (2019) 030032
16. Zapryanova T., Tachev G., Generalized Inverse Theorem for Schoenberg Operator, *JMMF*, 1(2), (2012), 11-16.

Asymptotic Methods for Researching of Differential Equations and Differential Equations' Systems Using a Quadratic Nonlinearity

Kostadin Sheyretski [1], Meglena Lazarova [2]

University of National and World Economy, Sofia, Bulgaria
[2] Faculty of Applied Mathematics and Informatics, Technical University of Sofia, Bulgaria
sheyretski@unwe.bg, meglena.laz@tu-sofia.bg

Abstract. Two applications of the small parameter's method in searching for an asymptotic solution of some differential equations with quadratic nonlinearity are considered. The solutions are obtained with accuracy of first approximation. In this paper we give the opportunities of studying the analytical methods of approximation. We are trying to find the opportunities to connect them with the modern computer technology.

Keywords. Differential equations, Asymptotic methods, Nonlinear dynamics.

1. Nature of the asymptotic solutions

Some useful methods for finding of the nonlinear differential equations are the so called *asymptothic methods* [1]. An interesting fact for these methods is that the function of the real variable x is decomposed into a series when $x \to x_0$ and this series is usually divergent. When the series is fixed to some of its members, then the asymptotic approximation is going to be more precise at $x \to x_0$. An important advantage is that the error resulting from this replacement is easily estimated. Let us consider the asymptotic approximation of the function $F(x)$ when $x \to x_0$ and also we have the functional row $f_n(x), \quad n = 1, 2, 3, \ldots$

The functional raw $f_n(x)$, given in the domain D is called an asymptotic raw if for each n in D the following equality holds true:

$$f_{n+1}(x) = o\left(f_n(x)\right), \text{ when } x \to x_0 \tag{1}$$

or we can write it in the following way:

$$\lim_{x \to x_0} \frac{f_{n+1}(x)}{f_n(x)} = 0 \cdot \tag{2}$$

Then we can make an asymptotic decomposition of the function $F(x)$ using the following asymptotic raw:

$$F(x) = \sum_{n=1}^{N} a_n f_n(x) + o\left(f_N(x)\right) \quad \text{when} \quad x \to x_0 . \tag{3}$$

The coefficients are easily found by consistent using of the following border crossing:

$$a_m = \lim_{x \to x_0} \frac{F(x) - \sum_{n=1}^{m-1} a_n f_n(x)}{f_m(x)} \quad \text{for} \quad m = 2,3,... \tag{4}$$

$$a_1 = \lim_{x \to x_0} \frac{F(x)}{f_1(x)}. \tag{5}$$

In the next examples we illustrate how to apply the one of the asymptotic methods called the method of the small parameter [2, 3]. In the both examples given below differential equations with quadratic nonlinearity are considered. This determines the specificity of the solution.

2. A case of quadratic nonlinearity in a second-order differential equation

Let us apply the same method for an equation with a quadratic nonlinearity:

$$\ddot{x} + \omega^2 x + \varepsilon \omega^2 x^2 = 0, \tag{6}$$

as we consider the case when $|\varepsilon| \ll 1$. We denote the first derivative of time by dot over x. The initial conditions are as follows:

$$x(0) = L, L \ll 1,$$
$$\dot{x}(0) = 0. \tag{7}$$

We search the solution of the equation (6) in the form:

$$x = C \cos \Omega t + \varepsilon x_1. \tag{8}$$

After differentiation of (8), substituting in (6) and after some calculations we get the following expression:

$$C(\omega^2 - \Omega^2) \cos \Omega t + \varepsilon \left(\ddot{x}_1 + \omega^2 x_1 + \frac{C^2 \omega^2}{2} + \frac{C^2 \omega^2}{2} \cos 2\Omega t \right) = 0. \tag{9}$$

The equation (9) can be solved considering the following expressions:

$$\omega^2 = \Omega^2, \tag{10}$$
$$\ddot{x}_1 + \omega^2 x_1 + \frac{C^2 \omega^2}{2} + \frac{C^2 \omega^2}{2} \cos 2\omega t = 0. \tag{11}$$

The equation (11) has the following solution:

$$x_1 = -\frac{C^2}{2} + \frac{C^2}{6} \cos 2\omega t. \tag{12}$$

Considering the initial conditions (7) we get:

$$C - \frac{\varepsilon C^2}{3} = L,$$

For the amplitude is obtained:

$$C = L + \frac{\varepsilon L^2}{3}. \tag{13}$$

So the final solution of equation (6) for given initial conditions (7) is:

$$x = L \left(1 + \frac{\varepsilon L}{3}\right) \cos \omega t - \frac{\varepsilon L^2}{2} + \frac{\varepsilon L^2}{3} \cos 2\omega t. \tag{14}$$

3. Quadratic nonlinearity of two connected oscillators

Let ε is a small parameter and consider the following system of equations:

$$\ddot{q}_1 + \nu_1^2 q_1 + \varepsilon \nu_1^2 q_1^2 - \lambda_1^2 q_2 = 0,$$

$$\ddot{q}_2 + \nu_2^2 q_2 + \varepsilon \nu_2^2 q_2^2 - \lambda_2^2 q_1 = 0. \quad (15)$$

We search the system's solutions for (15) in the following form:

$$q_1 = a_1 \cos(\Omega t + \varphi) + \varepsilon b_1 \cos(2\Omega t + 2\varphi) + \varepsilon \delta_1, \quad (16)$$

$$q_2 = a_2 \cos(\Omega t + \varphi) + \varepsilon b_2 \cos(2\Omega t + 2\varphi) + \varepsilon \delta_2, \quad (17)$$

where Ω is the oscillation frequency of the connected oscillators and φ is their phase.

When we substitute (16) and (17) in the system (15) we obtain:

$$\left(-\Omega^2 a_1 + \nu_1^2 a_1 - \lambda_1^2 a_2\right)\cos(\Omega t + \varphi) + \varepsilon\left(-4\Omega^2 b_1 + \nu_1^2 b_1 + \frac{\nu_1^2 a_1^2}{2} - \lambda_1^2 b_2\right)\cos(2\Omega t + 2\varphi) +$$

$$+\varepsilon\left(\nu_1^2 \delta_1 - \lambda_1^2 \delta_2 + \frac{\nu_1^2 a_1^2}{2}\right) = 0 \quad (18)$$

$$\left(-\Omega^2 a_2 + \nu_1^2 a_2 - \lambda_2^2 a_1\right)\cos(\Omega t + \varphi) + \varepsilon\left(-4\Omega^2 b_2 + \nu_2^2 b_2 + \frac{\nu_2^2 a_2^2}{2} - \lambda_2^2 b_1\right)\cos(2\Omega t + 2\varphi) +$$

$$+\varepsilon\left(\nu_2^2 \delta_2 - \lambda_2^2 \delta_1 + \frac{\nu_2^2 a_2^2}{2}\right) = 0 \quad (19)$$

To satisfy the identities (16) and (17) for every value of the trigonometric functions is needed and sufficient the following six equations to be fulfilled:

$$\left(-\Omega^2 + \nu_1^2\right) a_1 - \lambda_1^2 a_2 = 0, \quad (20)$$

$$-\lambda_2^2 a_1 + \left(-\Omega^2 + \nu_2^2\right) a_2 = 0, \quad (21)$$

$$\left(-4\Omega^2 + \nu_1^2\right) b_1 - \lambda_1^2 b_2 = -\frac{\nu_1^2 a_1^2}{2}, \quad (22)$$

$$-\lambda_2^2 b_1 + \left(-4\Omega^2 + \nu_2^2\right) b_2 = -\frac{\nu_2^2 a_2^2}{2}, \quad (23)$$

$$\nu_1^2 \delta_1 - \lambda_1^2 \delta_2 = -\frac{\nu_1^2 a_1^2}{2}, \quad (24)$$

$$-\lambda_2^2 \delta_1 + \nu_2^2 \delta_2 = -\frac{\nu_2^2 a_2^2}{2}. \quad (25)$$

We solve the equation's system (20) - (21) using the following method:

We put $a_1 = \sigma a_2$, where σ is the amplitude distribution coefficient. Then from the first two equations we can straight away determine the values of Ω and σ:

$$\sigma = \frac{\lambda_1^2}{\nu_1^2 - \Omega^2} = \frac{\nu_2^2 - \Omega^2}{\lambda_2^2}, \tag{26}$$

$$\left(\nu_1^2 - \Omega^2\right)\left(\nu_2^2 - \Omega^2\right) - \lambda_1^2\lambda_2^2 = 0. \tag{27}$$

The equation (25) determines the normal frequencies of the system $\Omega_{1,2}$. For each of these normal frequencies only one decision corresponds to $\sigma_{1,2}$. The numerical value of the amplitude a_2 is determined by the initial conditions.

To solve the system of equations (22) - (23) we take one of the two values determined by the equation (27) for the normal frequency of the system and the corresponding to this normal frequency value of the amplitude distribution coefficient.

$$\left(-4\Omega^2 + \nu_1^2\right)b_1 - \lambda_1^2 b_2 = -\frac{\nu_1^2\sigma^2 a_2^2}{2},$$

$$-\lambda_2^2 b_1 + \left(-4\Omega^2 + \nu_2^2\right)b_2 = -\frac{\nu_2^2 a_2^2}{2}.$$

The final system's solution is given by:

$$b_1^{1,2} = -\frac{a_2^2}{2}\frac{\left(-4\Omega_{1,2}^2 + \nu_2^2\right)\nu_1^2\sigma_{1,2}^2 + \lambda_1^2\nu_2^2}{\left(-4\Omega_{1,2}^2 + \nu_1^2\right)\left(-4\Omega_{1,2}^2 + \nu_2^2\right) - \lambda_1^2\lambda_2^2}, \tag{28}$$

$$b_2^{1,2} = -\frac{a_2^2}{2}\frac{\left(-4\Omega_{1,2}^2 + \nu_1^2\right)\nu_2^2 + \lambda_2^2\nu_1^2\sigma_{1,2}^2}{\left(-4\Omega_{1,2}^2 + \nu_1^2\right)\left(-4\Omega_{1,2}^2 + \nu_2^2\right) - \lambda_1^2\lambda_2^2}. \tag{29}$$

Similarly to the previous system we also search the solutions of the system's equations (24)-(25):

$$\nu_1^2\delta_1 - \lambda_1^2\delta_2 = -\frac{\nu_1^2\sigma^2 a_2^2}{2},$$

$$-\lambda_2^2\delta_1 + \nu_2^2\delta_2 = -\frac{\nu_2^2 a_2^2}{2}.$$

It is easy to see that the solution is:

$$\delta_1^{1,2} = -\frac{a_2^2}{2}\frac{\left(\nu_1^2\nu_2^2\sigma_{1,2}^2 + \nu_2^2\lambda_1^2\right)}{\nu_1^2\nu_2^2 - \lambda_1^2\lambda_2^2}, \tag{30}$$

$$\delta_2^{1,2} = -\frac{a_2^2}{2}\frac{\left(\nu_1^2\nu_2^2 + \nu_1^2\sigma_{1,2}^2\lambda_2^2\right)}{\nu_1^2\nu_2^2 - \lambda_1^2\lambda_2^2}. \tag{31}$$

We obtained two types of solutions according to the specific value of the normal frequency. The solutions do not obey to the principle of the superposition due to the nonlinear character of the given equations.

4. Conclusion

The obtained solutions are accurate to the first approximation of the small parameter. The presented methodology can be generalized to approximations of a higher order but the calculations can be done mainly by computer. Thus the analytical method gives a more general idea of the system's behavior but there is a loss of adequacy for longer periods of time. On the other hand if they are combined with appropriate numerical methods then the accuracy of the analytical solution which is obtained can improve the accuracy of the obtained analytical solution. This may lead to the possibility of making a good judgment when the asymptotic order can be broken and how to examine its behavior in the limits of applicability.

Acknowledgement

The second author is totally supported by Grant DN 12/11/20 Dec. 2017 of the Ministry of Education and Science of Bulgaria.

References

1. Murray J. D.. Lectures on Nonlinear Differential-Equation Models in Biology, Oxford U. P., 1977.
2. Hayashi C. Nonlinear oscillations in physical systems. McGrow-Hill book company, New York, 1964.
3. Naifeh, A. Introduction to Perturbation Techniques. John Wiley&Sons, New York, 1981.
4. Jeliazkov I., Oscillations and waves , UP „ St. Kliment Ohridski", Sofia, 2000.
5. Lichtenberg A., M. Liberman. Regular and stochastic motion. Springer-Verlag, New York, 1983.

A Comparative Numerical Analysis for Finding the Exact Solution for the Zero Coupon Bond's Price in the Classical Vasicek Model Influenced by the Impact of the Market Price of Risk

Meglena Lazarova[1], Sylvi-Maria Gurova[2]

[1] Faculty of Applied Mathematics and Informatics, Technical University of Sofia, Bulgaria
[2] Institute of Information and Communication Technologies - BAS
meglena.laz@tu-sofia.bg, smgurova@parallel.bas.bg

Abstract. In this paper we consider the zero coupon bond's price in the classical Vasicek model. Based on a partial differential equation of the second order a common solution for the bond's price with the influence of a market price of risk is obtained. We call this model – a Vasicek model with a risk factor. Two assumptions for the market price of risk are considered. The first suggests that the market price of risk takes only positive values and the second one only negative values. Using these two assumptions and replacing them into the exact solution of the zero coupon bond's price, we obtain two models for finding of the bonds' prices exact solutions. Two numerical experiments are also introduced. They are made by using fixed values of some of the parameters when other of them are changing in time. Based on the models' results, some useful conclusions are made.

Keywords. Finance modeling, Zero coupon bond's price, Market price of risk, Stochastic differential equations, Partial differential equations.

1. Introduction

Vasicek's model is a mathematical model which describes the random movement of the interest rates, see Vasicek O. (1977), [6]. In particular it is a stochastic investment model which represents the movement of the interest rate by a diffusion process. The main purpose in such model is to find the zero coupon bond's price at time *t* with maturity T. As the bond's price can changes in time quite accidently, it can be interpreted as a stochastic process. In this paper a classical Vasicek model for estimating the price of the zero coupon is analyzed. There is a parameter in the model which influences the price of the zero coupon bond called a market risk. An interesting question that arises is how does this market risk affects the final price of the zero coupon bond when it takes positive or negative values. Based on this assumptions we consider two financial models where the zero coupon bond's price is determined by using the apparatus of both stochastic and partial differential equations. The connection between them is obvious from the Fayman-Kac's theorem and it turned out that is a good way to determine the final bond's price, see Oksendal B. (1998), [3]. Also we introduce a numerical comparative analysis made by a specific boundary condition and give some essential results. In section 2 we intorduce the classical Vasicek model where the general formula of the market price of risk and the value of the zero coupon bond are given. Based on the assumptions that the market price of risk can be either positive or negative, we calculate and present two formulas of the exact solutions for the zero coupon bond's price. A numerical analysis and some comparative characteristics are made in Section 3. Some concluding remarks are given in Section 4.

2. Some main points in the classical Vasicek model

The classical Vasicek model is one of the earliest and the simplest term structure models based on the assumptions that the interest rates follow a random movement. In finance, the main application of the classical Vasicek model is concerned with the determining of the zero coupon bond's price, using the fact that the interest rate is following a stochastic differential equation, see Nazil M. N. (2009), [7].

Let r_0, , $\alpha > 0$, $\sigma > 0$ and μ are real constants. In the classical Vasicek model the interest rate is given by:

(1) $dr(t) = \alpha(\mu - r)dt + \sigma dW_t$ with an initial condition $r(0) = r_0$,

where

r - the instantaneous short rate interest

α - the speed of mean reversion

μ - the long-run expected value for the interest rate r

σ - the instantaneous standard deviation of the interest rate r

W_t - a standard Wiener process with mean 0 and standard deviation 1

The stochastic process used by Vasicek is also known as the Ornstein-Uhlenbesk process. The first term in the stochastic process proposed by Vasicek pulls the short rate r back towards μ. So μ can be thought of as the long-run level of the short rate. When the short rate r is above μ, the first term trends to pull r downward since α is assumed to be a positive one. When the short rate r is below μ then r trends to drift upward. The impact of the mean reversion is to create realistic interest rate cycles with the level of α determing the length and the violence of rises and falls in interest rates. Also there are some financial problems that are interested in the constant $\sigma^2 / 2\alpha$. This constant is known in the literature as a long term variance. The meaning of this constant is that the future expected values for the interest rate r may be regrouped into a long-term period of time with this value of the variance.

Let $P(r,t,T)$ be the price value of a zero coupon bond at time t and maturity T. The price depends from the interest rate movement and also the market value of the price reflects to the market expectation of the interest rate in the future. By Ito's lemma we obtain the following movement in the price of a zero coupon bond:

(2) $$dP(r,t,T) = P_r dr + \frac{1}{2} P_{rr} (dr)^2 + P_t dt,$$

where $P_r = \frac{\partial P(r,t,T)}{\partial r}$, $P_t = \frac{\partial P(r,t,T)}{\partial t}$ and $P_{rr} = \frac{\partial^2 P(r,t,T)}{\partial r^2}$.

Taking into account that $dr(t) = \alpha(\mu - r)dt + \sigma dW_t$ and substituting in (2) we obtain the following equation for the zero coupon bond's price:

(3) $$dP = P_r[\alpha(\mu - r)dt + \sigma dW_t] + \frac{1}{2} P_{rr}\sigma^2 dt + P_t dt$$

The calculations are obtained thanks to the following differential table:

X	dW_t	dt
dW_t	dt	0
dt	0	0

i.e $(dW)^2 = dt$, $dW_t dt = 0$ and $dt.dt = 0,$ see Stoyanov J. (1978), [5].

In the further exposition of the model is extremely important to find a formula for the zero coupon bond's price *P(r, t, T).* The realization of this task is achieved by the assumption of that the bond market is non-arbitrary. Let on this market an investor forms a portfolio with a quantity *W* which includes a unit quantity of a zero coupon bond 1 with maturity T_1 and a quantity *w* of a zero coupon bond 2 with a maturity T_2. Then the portfolio's value is given by:

(4) $$W = P_1 + w.P_2,$$

where P_1 is the zero coupon bond's price of the first portfolio's bond and P_2 is the zero coupon bond's price of the second portfolio's bond. Then the change of the portfolio's value in time is given by:

(5) $$dW = dP_1 + w.dP_2,$$

We write the Ito's formula separately for the differentials dP_1, dP_2 and obtain the following partial differential equations:

$$dP_1 = P_{1_r}\,dr + \frac{1}{2}P_{1_{rr}}(dr)^2 + P_{1_t}dt = P_{1_r}[\alpha(\mu - r)dt + \sigma dW_t] + \frac{1}{2}P_{1_{rr}}\sigma^2 dt + P_{1_t}dt$$

$$dP_2 = P_{2_r}\,dr + \frac{1}{2}P_{2_{rr}}(dr)^2 + P_{2_t}dt = P_{2_r}\left[\alpha(\mu - r)dt + \sigma dW_t\right] + \frac{1}{2}P_{2_{rr}}\sigma^2 dt + P_{2_t}dt$$

After substituting the obtained values of dP_1 and dP_2 in the stochastic differential equation (5) we obtain the following equation:

(6) $$dW = \left[\frac{1}{2}P_{1_{rr}}\sigma^2 + P_{1_r}\alpha(\mu - r) + P_{1_t} + wP_{2_t}\right]dt + w\left[\frac{1}{2}P_{2_{rr}}\sigma^2 + P_{2_r}\alpha(\mu - r)\right]dt + \left(P_{1_r} + wP_{2_r}\right)\sigma dW_t$$

If the coefficient in front of the differential dW_t is a zero i.e. $P_{1_r} + w.P_{2_r} = 0$ or $w = -\frac{P_{1_r}}{P_{2_r}}$, then we say that on the market is forming a riskless portfolio.

Once the risk has been removed the change of the portfolio's value in time is given by:

(7) $$dW = \left(rP_1 + rwP_2\right)dt\,.$$

By equating the right sides of the equations (6) and (7) and substituting by $w = -\frac{P_{1_r}}{P_{2_r}}$ we obtain the following equation:

(8) $$rP_1 - \frac{P_{1_r}}{P_{2_r}} P_2 r = \frac{1}{2} P_{1_{rr}} \sigma^2 + P_{1_r} \alpha(\mu - r) + P_{1_t} - \frac{P_{1_r}}{P_{2_r}} . P_{2_t} - \frac{1}{2} \sigma^2 \frac{P_{1_r}}{P_{2_r}} . P_{2_{rr}} - \frac{P_{1_r}}{P_{2_r}} . P_{2_r} \alpha(\mu - r)$$

After simplifying and dividing by $P_{1_r} . \sigma$ we obtain the following equality:

(9) $$\frac{\frac{1}{2}\sigma^2 P_{1_{rr}} + \alpha(\mu - r) P_{1_r} + P_{1_t} - rP_1}{\sigma P_{1_r}} = \frac{\frac{1}{2}\sigma^2 P_{2_{rr}} + \alpha(\mu - r) P_{2_r} + P_{2_t} - rP_2}{\sigma P_{2_r}} = -\lambda$$

i.e. the market price of risk is given by:

$$-\lambda = \frac{\frac{1}{2}\sigma^2 P_{rr} + \alpha(\mu - r) P_r + P_t - rP}{\sigma P_r} .$$

The market price of risk λ is assumed to be a constant. From the last equation we obtain the partial differential equation of the zero coupon bond's price for the classical Vasicek model of the form:

(10) $$\frac{1}{2}\sigma^2 P_{rr} + [\alpha(\mu - r) + \lambda\sigma] P_r + P_t - rP = 0$$

with the following boundary condition:

(11) $$P(r, T, T) = 1$$

i.e the zero coupon bond's price at maturity equals its principal amount 1.

The zero coupon bond's in the classical Vasicek model is analytically found and is given by the following formula, see Donald R. van Deventer et. al (1997) , [1]:

(12) $$P(r, t, T) = P(r, \tau) = e^{-rF(\tau) - G(\tau)},$$

where $F(t,T) = F(\tau) = \frac{1}{\alpha}\left(1 - e^{-\alpha\tau}\right)$ and $G(t,T) = G(\tau) = \left[\mu + \frac{\lambda\sigma}{\alpha} - \frac{\sigma^2}{2\alpha^2}\right]\left[\tau - F(\tau)\right] + \frac{\sigma^2}{4\alpha} F^2(\tau)$

are functions of the remaining time to maturity $\tau = T - t$, see Donald R. van Deventer et. al (1997), [1].

The stochastic process proposed by Vasicek allows to calculate the expected value and the variance of the short rate at any time in the future s from the perspective of the current time t. Denoting the short rate at time t by $r(t)$ then the expected value of the short rate at future time s is as follows:

(13) $$E_t\left[r(s)\right] = \mu + \left[r(t) - \mu\right] e^{-\alpha(s-t)}.$$

The standard deviation of the potential values of the interest rate r around this mean value is:

(14) $$\text{Standard deviation}_t\left[r(s)\right] = \sqrt{\frac{\sigma^2}{2\alpha}\left[1 - e^{-2\alpha(s-t)}\right]}.$$

It is known that the interest rate $r(s)$ at future time s is normally distributed. So there is a positive probability that $r(s)$ can be negative. This is inconsistent with a no-arbitrage economy in the special sense that consumers hold an option to hold cash instead of investing at negative interest rates. The magnitude of this theoretical problem with the Vasicek model depends on the level of the interest rates and the chosen parameters.

Theoretically the parameter $\lambda,$ given in formula (9) is represented in with a negative value, see Donald R. van Deventer et al. (1997), [1]. In this paper we take a positive value of the parameter λ and on this base we make a comparative analysis between the two zero coupon's bond prices in the case of a negative market price of risk and a positive one.

In the case when the market price of risk is negative one, see Donald the following parabolic differential equation is obtained, see Donald R. van Deventer et. al (1997), [1]:

$$\sigma^2.P_{rr}+\left[\alpha(\mu-r)+\lambda\sigma\right]P_r-rP+P_t=0 \tag{15}$$

with a boundary condition (11).

When the portfolio is a riskless one, then theoretically $\lambda=0.$

Such model is called *a model without a risk factor,* see Lazarova and Gurova (2018), [10]. The main purpose in this paper is that we are interested in a model influenced by a risk factor.

Respectively, when the market price of risk is positive one then the parabolic differential equation is as follows, see Kwok Y. K., [8] and Buckova Z., et. all (2016), [9]:

$$\sigma^2.P_{rr}+\left[\alpha(\mu-r)-\lambda\sigma\right]P_r-rP+P_t=0 \tag{16}$$

again with a boundary condition (11).

In order to describe the model for finding of the zero coupon's bond's price in the presence of a market risk factor correctly, we use the marginal condition *P (r, T, T) = 1*. The meaning of this condition is that in the year of the maturity the price of the zero coupon's bond reaches 1.

The exact solution for equation (15) which gives the zero coupon bond's price in the classical Vasicek model influenced by the impact of the market price of risk is the following:

$$P(r,t,T)=\exp\left\{-r.F(\tau)-(\tau-F(\tau))\left[-\frac{\sigma^2}{2\alpha^2}+\mu+\frac{\lambda\sigma}{\alpha}\right]-\frac{\sigma^2F^2(\tau)}{4\alpha}\right\}. \tag{17}$$

The exact solution corresponding to the equation (16) is given by:

$$P(r,t,T)=\exp\left\{-r.F(\tau)-(\tau-F(\tau))\left[-\frac{\sigma^2}{2\alpha^2}+\mu-\frac{\lambda\sigma}{\alpha}\right]-\frac{\sigma^2F^2(\tau)}{4\alpha}\right\}. \tag{18}$$

In the following Section 3 we will make a comparative analysis between the two exact solutions (17) and (18) which are solutions of the parabolic partial differential equations (15) and (16). The main purpose is to determine what is happening with the zero coupon bond's price in the two cases given above by using different values of the model's parameters. This is in the presence of market risk so that the investor buys more from them.

3. Numerical analysis and comparative characteristics

In section 2 we considered the classical Vasicek model and the analytical obtaining of the zero coupon bond's price for it. Depending on two proposed assumptions for the parameter describing the market price of risk λ we obtained two Cauchy problems (15) and (16) with one and the same boundary condition (11). Also the exact solutions of these problems are given.

The essential part in this paper is that the parameter which describes the market price of risk is different from zero i.e $\lambda \neq 0$. It is taken to be a different constant for each year until the maturity year *T*, see Sobhani A. and Milev M. (2018), [4]. When λ is a function of time, then the interest rate and the volatility parameters depend on time and such model reflects more realistic on the random movement of the asset prices in financial market, see Gzyl H. et. al. (2017), [2].

Using the exact solutions given in formulas (17), (18) and the software Mathematika 11 we find the zero coupon's bond price for each year before reaching the maturity T. The calculations are made by fixing the parameters - α, μ, σ, T and changing the parameters $\lambda, r, t.$

a. First numerical experiment

This numerical experiment uses the data in the table below. The zero coupon bonds' prices for the both cases – when the market price of risk is a positive and negative one are given for different periods of time before reaching the maturity year $T = 5$. The prices are obtained by using the following numerical values (Table 1).

Table 1: Specific values determing the zero coupon bond's price in the classical Vasicek model with a positive and a negative market price of risk.

	α	μ	σ	λ	r	t (for each years)	T (a maturity year)	Price
Vasicek model with a risk factor „+λ“	0.20	0.10	0.05	0.05	0.010	0	5	0.8468
	0.20	0.10	0.05	0.10	0.015	1	5	0.8873
	0.20	0.10	0.05	0.15	0.020	2	5	0.9191
	0.20	0.10	0.05	0.20	0.025	3	5	0.9452
	0.20	0.10	0.05	0.25	0.030	4	5	0.9701
	0.20	0.10	0.05	0.30	0.035	5	5	1.00
Vasicek model with a risk factor „-λ“	0.20	0.10	0.05	0.05	0.010	0	5	0.8087
	0.20	0.10	0.05	0.10	0.015	1	5	0.8337
	0.20	0.10	0.05	0.15	0.020	2	5	0.8692
	0.20	0.10	0.05	0.20	0.025	3	5	0.9126
	0.20	0.10	0.05	0.25	0.030	4	5	0.9588
	0.20	0.10	0.05	0.30	0.035	5	5	1.00

It is seen from Table 1 that the zero coupon bond's price for the two cases shows a slight price decrease in time when we have the case that λ is negative.

Using the data from Table 1 and the software Matematika 11 we obtain Figure 1. The yellow surface that is seen on the figure represents the solution of the equation (17) and the green one the solution of the equation (18). Graphically we give an illustration of the two

formulas solutions by showing their differences. The contours of the two graphs could be better seen in Figure 2 which represents the change in the values of the zero coupon bond's prices.

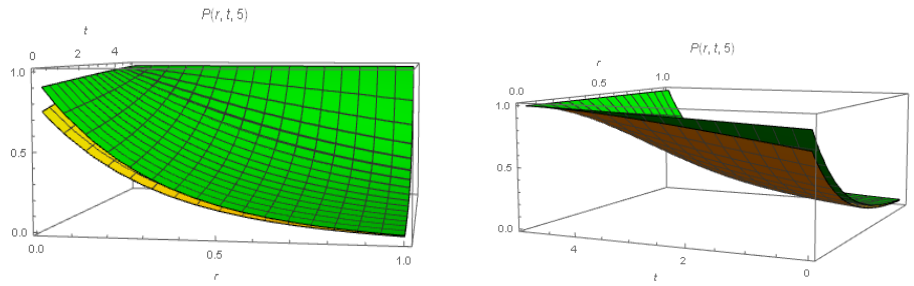


Fig. 1. A comparative graphic characteristic showing the difference in zero coupon bonds' prices between the models for a period of 5 years.

The graphic and the table below show the difference in the comparative characteristics

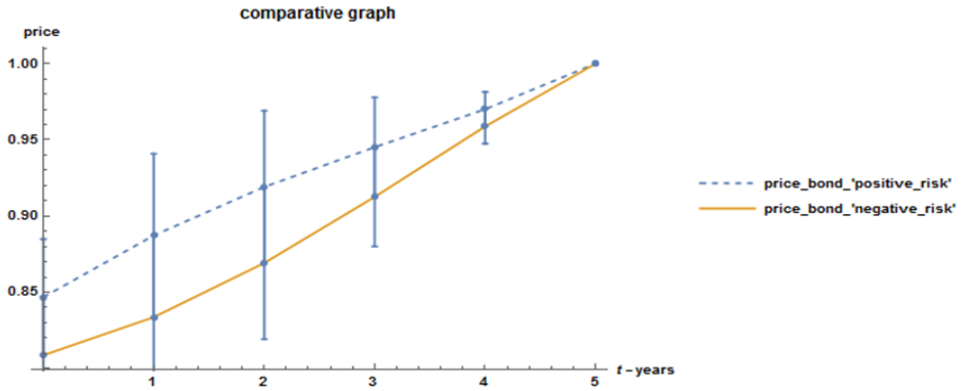


Fig. 2. A comparative graph with given deviations for the first experiment.

These differences in the prices at different time periods are reflected in Table 2. There is an increase in the price difference until the second year and then there is a tendency to decrease until it reaches a year of maturity. This increase / decrease is also visible in Figure 2.

Table 2. The deviation of the values of the bonds' prices for different periods of time.

Years:	0	1	2	3	4	5
Error:	0.0381	0.0536	0.0499	0.0326	0.0113	0

It can be seen from Table 1 that the zero coupon bonds' prices are calculated in each individual year before reaching the maturity year using formula (17) are higher than those obtained by formula (18). The difference in prices is clearly expressed by Figure 2, Figure 3 and Table 2. Based on the sample used for the parameters describing the models, we can say that it is more profitable for the buyer to buy more bonds which prices are calculated by formula (18**).**

b. Second numerical experiment

This numerical experiment uses the data given in the table below. The difference is in the set of the numerical values of the parameters α, μ, σ, T where the values of the variables λ, t, r are changing. Making a comparison between the column "Price" from Table 1 and the column "Price" from Table 3, we can see that the zero coupon bond's price in the classical Vasicek model influenced by the impact of the market price of risk is much lower than in the first experiment.

Table 3. Specific values determing the zero coupon bonds' prices in the classical Vasicek model with a positive and a negative market price of risk.

	α	μ	σ	λ	r	t (for each years)	T (a maturity year)	Price
Vasicek model with a risk factor „+λ“	0.30	0.25	0.10	0.05	0.010	0	10	0.2644
	0.30	0.25	0.10	0.10	0.015	2	10	0.3974
	0.30	0.25	0.10	0.15	0.020	4	10	0.5571
	0.30	0.25	0.10	0.20	0.025	6	10	0.7283
	0.30	0.25	0.10	0.25	0.030	8	10	0.8877
	0.30	0.25	0.10	0.30	0.035	10	10	1.00
Vasicek model with a risk factor „-λ“	0.30	0.25	0.10	0.05	0.010	0	10	0.2106
	0.30	0.25	0.10	0.10	0.015	2	10	0.2854
	0.30	0.25	0.10	0.15	0.020	4	10	0.4038
	0.30	0.25	0.10	0.20	0.025	6	10	0.5829
	0.30	0.25	0.10	0.25	0.030	8	10	0.8172
	0.30	0.25	0.10	0.30	0.035	10	10	1.00

Using the data from Table 3 and the software Matematika 11 we obtain Figure 3. The blue surface represents the solution of equation (18) and the red one represents the solution of the equation (17). The considered figure represents the graphical comparison of the solutions obtained in solving of the two formulas where their differences are shown.

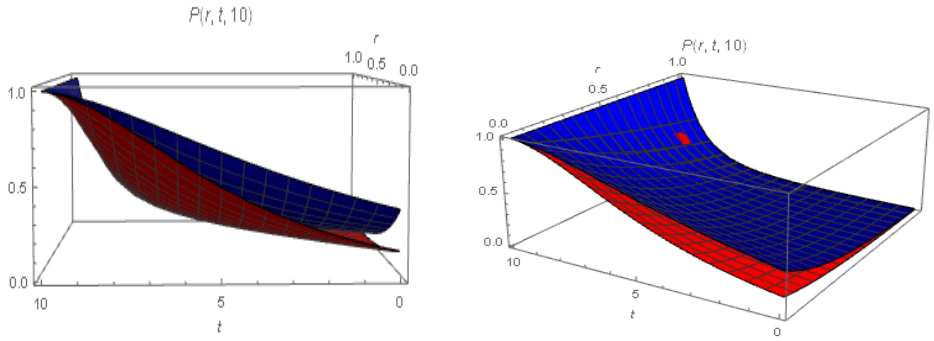


Fig. 3. A comparative graph representing the two Vasicek models with different market prices of risk given from different angles for a time period of 10 years.

The significant contrast from the first experiment is much "brighter" expressed in Figure 4. The graph represents the price difference of the two models which are calculated by

using the data from Table 3. The price differences that depend on the time periods are given in Table 4. Until the fourth year it is seen that there is an increase in the price difference. After that until the year of maturity a dicrease is reached and also seen. We found a similar relationship in the first experiment. This increase / decrease is also detectable in Figure 4.

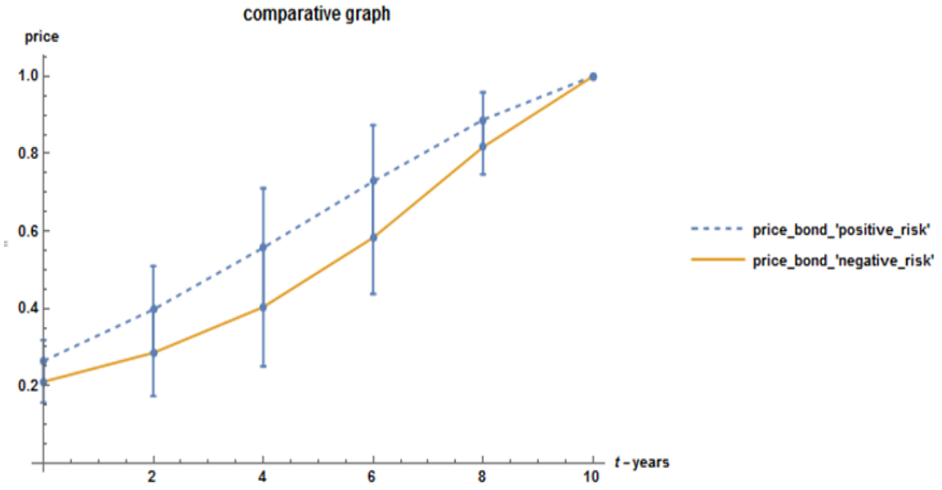


Fig. 4. A comparative graph with given deviations for the second experiment.

Table 4. The deviation of the value of the bond's price for different periods of time.

Years:	0	2	4	6	8	10
Error:	0.0538	0.1120	0.1533	0.1454	0.0705	0

The conclusions that can be drawn from the second numerical experiment are similar to the first one. Here we can notice that when the parameters λ, r, t are increasing, the bonds' prices are also increasing respectively. But the prices in the second experiment are lower than these obtained in the first experiment. It can be seen from Table 1 and Table 3.

4. Concluding remarks

In this paper the zero coupon bond's price in the classical Vasicek model is introduced. We take into account the influence of the market price of risk over the bond's price and we called the obtained model - a Vasicek model with a risk factor. Two assumptions for this model and more specifically for the market price of risk are considered. The first suggests that the market price of risk takes only positive values and the second one only negative values. Based on these two assumptions we give two numerical comparative experiments for different values of the model's parameters for selected periods of time. Some good profitable conclusions for the bond buyer are made.

5. Acknowledgement

The first author is supported by the project "Stochastic and simulation models in the field of medicine, social sciences and dynamic system" funded by the National Science Fund of Ministry of Education and Science of Bulgaria, Contract № DN 12/11/20 Dec. 2017).

The second author is partially supported by the financial fund allocated to the Sofia University "St. Kliment Ohridski", the Grant No: 80-10-116/16.04.2020 and supported by the Grant No: BG05M2OP001-1.002 "Building and development of centers of competence", financed by the Science and Education for Smart Growth Operational Program (2014-2020), project BG05M2OP001-1.002-0002-C 01 "Digitalization of the Economy in a Big Data Environment".

References

1. Donald van Deventer R., Imai K., Financial Risk Analytics – A term structure model approach for banking, insurance and investment management. IRWIN professional Publishing, USA (1997).
2. Gzyl H., Milev M., Tagliani A. Discontinuous payoff option pricing by Mellin transform: A probabilistic approach. Finance Research Letters, **20,** 281-288 (2017).
3. Oksendal, B., Stochastic differential equations – an addition with application. Springer- Verlag. Fifth edition, Berlin Heidelberg (1998).
4. Sobhani A., Milev M., A numerical method for pricing discrete double barrier option by Legendre multiwavelet. Journal of Computational and Applied Mathematics, **328**, 355-364 (2018).
5. Stoyanov, J. Stochstic processes – Theory and Applications, Sofia (1998).
6. Vasicek, O., An equilibrium characterization of the term structure. J. Financial Economics. **5** (1977).
7. Nazir, M. N., Short rates and bond prices in one-factor models, U.U.D.M. Project Report (2009).
8. Kwok Y. K., Interest Rate models and Bond Pricing, Chapter 7, https://www.math.ust.hk/~maykwok/courses/ma571/03_04/P_Ch07.pdf
9. Buckova Z., Stehlikova B., Sevcovic D., Numerical analytical methods for bond pricing in short rate convergence models of interest rates, 2016. https://arxiv.org/pdf/1607.04968.pdf
10. Lazarova M., Gurova S-M, Numerical comparative analysis on the classical Vasicek model for determing the zero coupon bond's price", MATTEX 2018, Conference proceeding, volume 1, pp. 121-135, ISSN 1314-3921.

Network Models Generation Approach for BDZ Travel Route Planning in Sofia Regional Area

Miglena Ivanova

University of National and World Economy, Sofia, Bulgaria
ivanova_mg@abv.bg

Abstract. A method for network models generation approach for passenger railway transport travel route planning in Sofia regional area, using the SPSS program, is presented in the report. A method using the SPSS program is illustrated, for defining the diversions quadrates, for the duration of operations with non-zero full reserve, from their average values (mathematical expectations) of exemplary network models for travels by BDZ in Sofia regional area. An analysis of the calculated diversions quadrates, for the duration of operations with non-zero full reserve, from their average values (mathematical expectations) is made, and as a result, symmetrical alignments for the diversions of certain network model are obtained.

Keywords. Network model, BDZ travel route planning, Squared deviations, Duration of operations with non-zero full reserve, From their average values (mathematical expectations), Symmetrical alignments, SPSS.

1. Network models for railway passenger transport travel in Sofia regional area

The network model is a graphic image of planned process, which shows the reciprocal relation and the sequence of its included operations (works). From a mathematical point of view, the network model is an oriented finite graph, the vertices of which depict events, and the ribs - elementary operations (works).

In the network models, each rib is matched to the time required to perform the respective operations. Operation (work) - this is a process that is done to achieve a certain result and connects one event to another. Event -the fact that the outcome of one or more operations, necessary and sufficient for the commencement of subsequent operations. Route - this is a sequence of operations in the network model in which the final event of an operation coincides with the initial event of the next post operation. Full path – this is a route in the network model, which beginning coincides with the starting event and the end – with the final event [1].

This report presents an approach for generating network models when planning a travel route with railway passenger transport in Sofia regional area (fig.1). An important indicator for passenger transport planning is the amount of transported passengers [2]. According to the new "Train movement schedule" (http://www.bdz.bg/bg/) for 2019 of the national railway carrier "BDZ Passenger Services" Ltd. (from 8.12.2018), the sections with the highest number of passengers on workdays (in the morning hours) are on ***Pernik - Sofia*** route.

This is the main factor, determining the choice of BDZ travel route in the present study: route ***Pernik - Pernik railway yard - Metal - Daskalovo - Dragichevo - Vladaya - Gorna Banya - Zaharna Fabrika - Sofia*** from Sofia regional area (fig.1). A route for travelling with

BDZ from Pernik station – to Sofia station can be generated by the "BDZ Passenger Services" Ltd electronic platform (http://www.bdz.bg/bg/), as shown in fig.2.

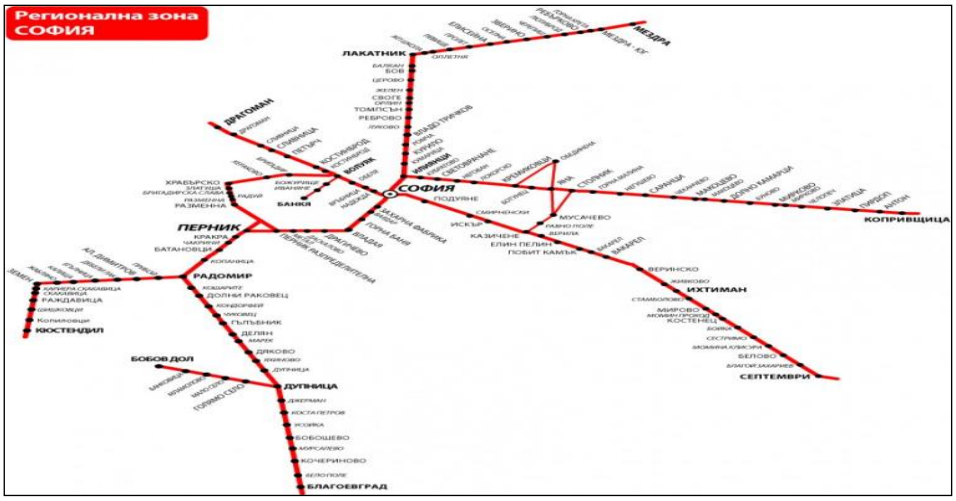


Fig.1 Map of Sofia regional area railway network

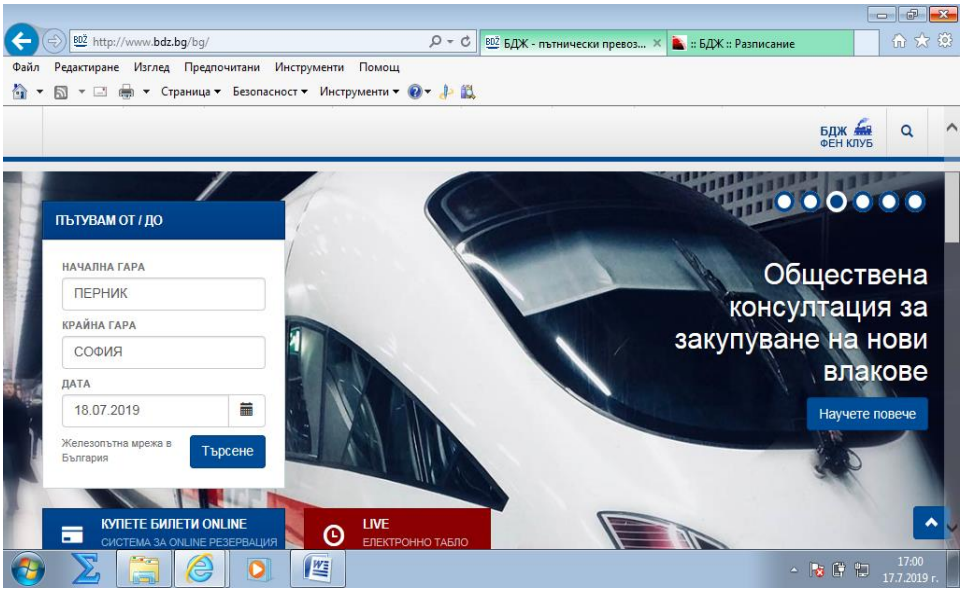


Fig.2. A route for travelling with BDZ from Pernik station - to Sofia station

To illustrate the network models generation approach for BDZ travel route, two network models (Model *A* and Model *B*) will be examined for BDZ travelling on the following destination: ***Pernik - Pernik railway yard - Metal - Daskalovo - Dragichevo - Vladaya - Gorna Banya - Zaharna Fabrika - Sofia***. Three options for train travel from Pernik station to Sofia station, in the morning hours on a workday (18.07.2019), are included in each of the two

network models. Travel guidelines (generated by the "BDZ Passenger Services" Ltd electronic platform) for network model *A* and network model *B* are shown respectively in table 1 and table 2. The three options included in network model *A* are with departure times: 7:00, 7:20 and 7:40 (table 1), and the options for network model *B* are with departure times: 8:15, 8:43 and 9:03 (table 2). The codes for building events of network model *A* and *B* are given in *Code* columns of table 1 and table 2.

Table 1. Travel guidelines for train travel from Pernik station to Sofia station /Model A/

Option A1					
Train: Blagoevgrad - Sofia			**Category: Fast**		
Departing			**Arriving**		
Code	*Time*	*Station/Stop*	*Code*	*Time*	*Station/Stop*
1	7:00	Pernik	2	7:04	Pernik railway yard
3	7:05	Pernik railway yard	4	7:10	Metal - stop
4	7:10	Metal - stop	5	7:12	Daskalovo
5	7:12	Daskalovo	6	7:15	Dragichevo
7	7:16	Dragichevo	11	7:31	Gorna Banya
14	7:36	Gorna Banya	17	7:44	Zaharna Fabrika
18	7:45	Zaharna Fabrika	21	7:51	Sofia
Option A2					
Train: Pernik - Sofia			**Category: Sub**		
Departing			**Arriving**		
Code	*Time*	*Station/stop*	*Code*	*Time*	*Station/stop*
8	7:20	Pernik	9	7:24	Pernik railway yard
10	7:27	Pernik railway yard	12	7:32	Metal - stop
12	7:32	Metal - stop	13	7:34	Daskalovo
13	7:34	Daskalovo	14	7:36	Dragichevo
15	7:37	Dragichevo	17	7:44	Vladaya
19	7:47	Vladaya	24	7:56	Gorna Banya
25	7:57	Gorna Banya	27	8:06	Zaharna Fabrika
28	8:07	Zaharna Fabrika	30	8:13	Sofia
Option A3					
Train: Pernik - Sofia			**Category: Sub**		
Departing			**Arriving**		
Code	*Time*	*Station/Stop*	*Code*	*Time*	*Station/Stop*
16	7:40	Pernik	17	7:44	Pernik railway yard
18	7:45	Pernik railway yard	20	7:50	Metal - stop
20	7:50	Metal - stop	22	7:52	Daskalovo
22	7:52	Daskalovo	23	7:55	Dragichevo
24	7:56	Dragichevo	26	8:03	Vladaya
29	8:08	Vladaya	31	8:18	Gorna Banya
32	8:19	Gorna Banya	33	8:27	Zaharna Fabrika
34	8:28	Zaharna Fabrika	35	8:34	Sofia

Both network models *A* and *B* (table 3 and table 4) can be built with the data from table 1 and table 2 for the codes and duration (in minutes) of each operation (i, j) for each of the three options from the *travel guidelines for train travel from Pernik station to Sofia station*.

Table 2. Travel guidelines for train travel from Pernik station to Sofia station /Model B/

Option B1					
Train: Pernik - Sofia			**Category: Sub**		
Departing			**Arriving**		
Code	*Time*	*Station/Stop*	*Code*	*Time*	*Station/Stop*
1	8:15	Pernik	2	8:19	Pernik railway yard
3	8:21	Pernik railway yard	4	8:26	Metal - stop
4	8:26	Metal - stop	5	8:28	Daskalovo
5	8:28	Daskalovo	6	8:31	Dragichevo
7	8:37	Dragichevo	9	8:44	Vladaya
10	8:45	Vladaya	14	8:54	Gorna Banya
15	8:55	Gorna Banya	16	9:03	Zaharna Fabrika
17	9:04	Zaharna Fabrika	20	9:10	Sofia
Option B2					
Train: Petrich - Sofia			**Category:Fast**		
Departing			**Arriving**		
Code	*Time*	*Station/Stop*	*Code*	*Time*	*Station/Stop*
8	8:43	Pernik	11	8:47	Pernik railway yard
12	8:48	Pernik railway yard	13	8:53	Metal - stop
13	8:53	Metal - stop	15	8:55	Daskalovo
15	8:55	Daskalovo	21	9:13	Gorna Banya
23	9:17	Gorna Banya	26	9:25	Zaharna Fabrika
27	9:26	Zaharna Fabrika	30	9:33	Sofia
Option B3					
Train: Kyustendil - Sofia			**Category: Fast**		
Departing			**Arriving**		
Code	*Time*	*Station/Stop*	*Code*	*Time*	*Station/Stop*
16	9:03	Pernik	18	9:07	Pernik railway yard
19	9:08	Pernik railway yard	21	9:13	Metal - stop
21	9:13	Metal - stop	22	9:16	Daskalovo
22	9:16	Daskalovo	24	9:19	Dragichevo
25	9:20	Dragichevo	28	9:27	Vladaya
29	9:28	Vladaya	31	9:38	Gorna Banya
32	9:39	Gorna Banya	33	9:48	Zaharna Fabrika
34	9:49	Zaharna Fabrika	35	9:55	Sofia

To illustrate the network models generation approach for BDZ travel route in Sofia regional area, a full reserve $R_p(i, j)$ of operations is determined, which is one of the main parameters of network models and its determination allows an analysis of the presented network models *A* and *B.*

The operations' codes, durations in minutes, and full reserve for network model *A* and *B* are generated with the SPSS program, respectively in table 3 and table 4.

Table 3 is generated by SPSS program with the functions *Analyze* $\Rightarrow$ *Reports* $\Rightarrow$ *Case Summaries* $\Rightarrow$ *Summarize Cases* $\rightarrow$ *Variables:* ***Kode.Model.A*** ***t.i.j.Model.A***, *Grouping Variable(s):* ***R.p.i.j.Model.A***, $\vee$ *Display cases* $\Rightarrow$ *Options* $\rightarrow$ *Title:Case Summaries,* $\vee$ *Subheadings for totals*, where ***Kode.Model.A*** is the list of operations codes of network model *A*, ***t.i.j.AB*** is the time corresponding to each operation in network model *A*, and ***R.p.i.j.Model.A*** is the full reserve corresponding to each model *A* operation. Table 4 is generated by SPSS program analogically.

Table 3

Case Summaries

			Kode.Model.A	t.i.j.Model.A
R.p.i.j.Model.A	0	1	1-2	4
		2	2-3	1
		3	3-4	5
		4	4-5	2
		5	5-6	3
		6	6-7	1
		7	7-11	15
		8	11-14	5
		9	14-15	1
		10	14-17	8
		11	15-17	7
		12	17-18	1
		13	17-19	3
		14	18-20	5
		15	19-24	9
		16	20-22	2
		17	22-23	3
		18	23-24	1
		19	24-26	7
		20	26-29	5
		21	29-31	10
		22	31-32	1
		23	32-33	8
		24	33-34	1
		25	34-35	6
	20	1	1-8	0
		2	8-9	4
		3	9-10	3
		4	10-12	5
		5	12-13	2
		6	13-14	2
	21	1	24-25	1
		2	25-27	9
		3	27-28	1
		4	28-30	6
		5	30-35	0
	40	1	1-16	0
		2	16-17	4
	43	1	18-21	6
		2	21-35	0

Table 4

Case Summaries

			Kode.Model.B	t.i.j.Model.B
R.p.i.j.Model.B	0	1	1-2	4
		2	2-3	2
		3	3-4	5
		4	4-5	2
		5	5-6	3
		6	6-7	6
		7	7-9	7
		8	9-10	1
		9	10-14	9
		10	14-15	1
		11	15-16	8
		12	15-21	18
		13	16-18	4
		14	18-19	1
		15	19-21	5
		16	21-22	3
		17	22-24	3
		18	24-25	1
		19	25-28	7
		20	28-29	1
		21	29-31	10
		22	31-32	1
		23	32-33	9
		24	33-34	1
		25	34-35	6
	22	1	21-23	4
		2	23-26	8
		3	26-27	1
		4	27-30	7
		5	30-35	0
	28	1	1-8	0
		2	8-11	4
		3	11-12	1
		4	12-13	5
		5	13-15	2
	45	1	16-17	1
		2	17-20	6
		3	20-35	0

From table 3 and table 4, it can be determined that both network models *A* and *B*, for BDZ travel on ***Pernik - Sofia*** route from the Sofia regional area (in the morning hours), have equal number of events 35. The network model *A* has *fifteen operations* with zero full reserve: *six operations* are with non-zero full reserve $R_p(i,j)=20$, *five operations* are with non-zero full reserve $R_p(i,j)=21$, *two operations* are with non-zero full reserve $R_p(i,j)=40$, and *two operations* are with non-zero full reserve $R_p(i,j)=43$. Network model *B* has *thirteen operations* with non-zero full reserve: *five operations* are with non-zero full reserve $R_p(i,j)=22$, *five operations are with non-zero full reserve* $R_p(i,j)=28$, and *three operations* are with non-zero full reserve $R_p(i,j)=45$.

2. A method using the SPSS program for defining the squared deviations, for the duration of operations with non-zero full reserve, from their average values (mathematical expectations) of network models for travels by BDZ in Sofia regional area

In this part of the present study, a method using the SPSS program is presented, for defining the squared deviations, for the duration of operations with non-zero full reserve, from their average values (mathematical expectations) of examined network models *A* and *B* for travels by BDZ in Sofia regional area.

The values $x_1, x_2, ..., x_n$ of a random variable X differ from one another and deviate from the mathematical expectation EX in one direction or another. The random variable $X - EX$ is called deviation (of X from EX). Part of the deviations $x_i - EX$, $i = 1, 2, ..., n$, is positive and another part is negative. The random variable $\left(X - EX\right)^2$ is called squared deviation (of X from EX), and is more suitable for examination, as its value is without regard to its sign. The number of values of squared deviations $\left(x_i - EX\right)^2$, $i = 1, 2, ..., n$ are as many as the values of the random variable X. The mathematical expectation of a squared deviation $E\left(X - EX\right)^2 = \sum_{i=1}^{n}\left(x_i - EX\right)^2 p_i$ is called dispersion of the random variable X [3 - 8].

In order to define the squared deviation $\left(x_i - EX\right)^2 ModelA$, $i = 1, 2, ..., 15$ and $\left(x_i - EX\right)^2 ModelB$, $i = 1, 2, ..., 13$ for the duration of the non-zero reserve operations from their average values (mathematical expectations) of the considered network models *A* and *B*, for BDZ travel in Sofia regional area, with SPSS program, two quantitative variables $X.Model.A$ and $X.Model.B$ should be defined on the SPSS program's page *Variable View*, and afterwards the values of these variables should be input to *Data View* page. The values for the duration in minutes, for each operation (i, j) with non-zero full reserve of the network model A, are put in the $X.Model.A$ column, and in $X.Model.B$ column - the values for the duration in minutes, for each operation (i, j) with non-zero full reserve of the network model B.

The average value (arithmetic mean) for the duration of non-zero full reserve operations of *A* and *B* network models can be generated by the SPSS program [9] using the functions *Analyze* $\Rightarrow$ *Descriptive Statistics* $\Rightarrow$ *Descriptives* $\rightarrow$ *Variable(s):**X.Model.A (X.Model.B)*** $\Rightarrow$ *Descriptives:Options* $\vee$ *Mean, Display Order* $\bullet$ *Variable List* (table 5 and table 6) and is calculated using the formula [9,10] (for unweighted arithmetic mean and weighted arithmetic mean for statistical population):

$$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} \qquad (1)$$

where $\overline{X}$ is the average value (arithmetic mean) for duration of non-zero full reserve operations of the network models *A* and *B*, x_i are the values for the duration of non-zero full reserve operations of the network model *A* ($i = 1, 2, ..., 15$) and network model *B* (

$i = 1,2,...,13$), N is the number of operations with non-zero full reserve for each of the network models (f_i, $i = 1,2,...,k$ are the frequencies (number of repetitions) of x_i, k is the groups number, a $\sum_{i=1}^{k} f_i = N$).

The mathematical expectation of a random variable X analogically to the average value (arithmetic mean) is an average weighted value. [3,4]. If the relevant probabilities p_i are taken instead of the frequencies of the average value (arithmetic mean), the formula for the mathematical expectation is derived $EX = \sum_{i=1}^{n} x_i p_i$ [3 - 8] of the random variable X :

$$\overline{X} = EX = \sum_{i=1}^{N} x_i p_i = \sum_{i=1}^{k} x_i p_i \qquad (2)$$

The network model *A* has an average value (mathematical expectation) $\overline{X} = EX = 2,87$ (table 5), and the network model *B* has an average value (mathematical expectation) $\overline{X} = EX = 3,00$ (table 6).

Table 5

Descriptive Statistics

	N	Mean
X.Model.A	15	2,87
Valid N (listwise)	15	

Table 6

Descriptive Statistics

	N	Mean
X.Model.B	13	3,00
Valid N (listwise)	13	

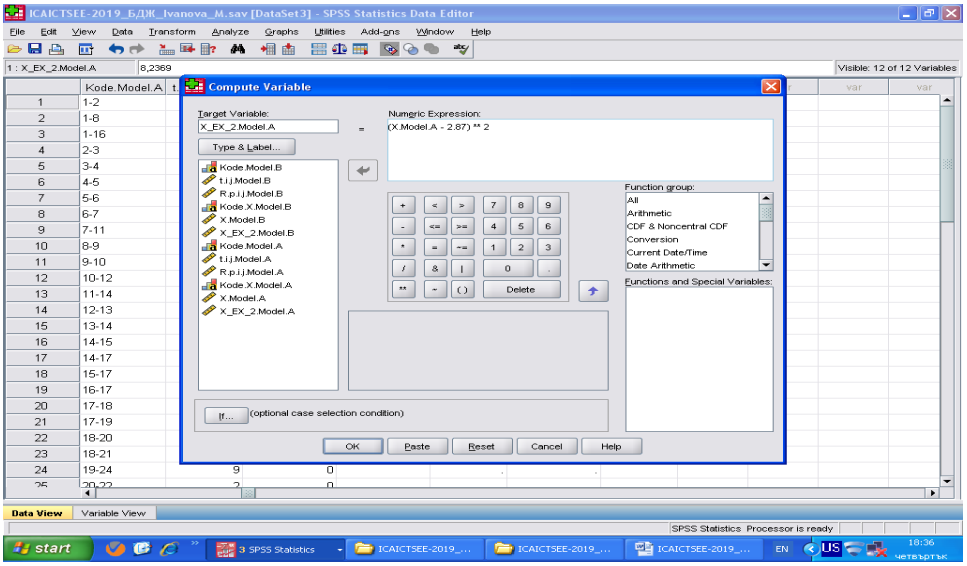


Fig. 3. Calculation of the squared deviation for duration of operations with non-zero full reserve from their average values (mathematical expectations) for the network model A by the SPSS program.

The squared deviation $\left(x_i - EX\right)^2 ModelA$ for the duration of the operations with non-zero full reserve from their average values (mathematical expectations) $\overline{X} = EX = \sum_{i=1}^{N} x_i p_i$ of the network model *A* represented by the variable $X\_EX\_2.Model.A$ can be calculated with SPSS program by choosing the functions *Transform* $\Rightarrow$ *Compute Variable*, as $X\_EX\_2.Model.A$ is put in the *Target Variable* field, and $\left(X.Model.A - 2{,}87\right)**2$ is put in the *Numeric Expression* field , as shown on fig. 3. The result is generated in the column of the variable $X\_EX\_2.Model.A$ on the *Data View* page. Analogically the squared deviation $\left(x_i - EX\right)^2 ModelB$ for the duration of the operations with non-zero full reserve from their average values (mathematical expectations) of network model *B* with the functions *Target Variable*: $X\_EX\_2.Model.B \Rightarrow$ *Numeric Expression:* $\left(X.Model.B - 3{,}00\right)**2$.

Table 7

Case Summaries

				Kode.X.Model.A	X_EX_2.Model.A	p.Model.A
X.Model.A	0	1		1-8	8,24	1/15
		2		1-16	8,24	1/15
		3		21-35	8,24	1/15
		4		30-35	8,24	1/15
		Total	N	4	4	4
			Mean		8,2369	
	1	1		24-25	3,50	1/15
		2		27-28	3,50	1/15
		Total	N	2	2	2
			Mean		3,4969	
	2	1		12-13	,76	1/15
		2		13-14	,76	1/15
		Total	N	2	2	2
			Mean		,7569	
	3	1		9-10	,02	1/15
		Total	N	1	1	1
			Mean		,0169	
	4	1		8-9	1,28	1/15
		2		16-17	1,28	1/15
		Total	N	2	2	2
			Mean		1,2769	
	5	1		10-12	4,54	1/15
		Total	N	1	1	1
			Mean		4,5369	
	6	1		18-21	9,80	1/15
		2		28-30	9,80	1/15
		Total	N	2	2	2
			Mean		9,7969	
	9	1		25-27	37,58	1/15
		Total	N	1	1	1
			Mean		37,5769	
	Total		N	15	15	15
			Mean		7,0489	

Table 8

Case Summaries

				Kode.X.Model.B	X_EX_2.Model.B	p.Model.B
X.Model.B	0	1		1-8	9,00	1/13
		2		20-35	9,00	1/13
		3		30-35	9,00	1/13
		Total	N	3	3	3
			Mean		9,0000	
	1	1		11-12	4,00	1/13
		2		16-17	4,00	1/13
		3		26-27	4,00	1/13
		Total	N	3	3	3
			Mean		4,0000	
	2	1		13-15	1,00	1/13
		Total	N	1	1	1
			Mean		1,0000	
	4	1		8-11	1,00	1/13
		2		21-23	1,00	1/13
		Total	N	2	2	2
			Mean		1,0000	
	5	1		12-13	4,00	1/13
		Total	N	1	1	1
			Mean		4,0000	
	6	1		17-20	9,00	1/13
		Total	N	1	1	1
			Mean		9,0000	
	7	1		27-30	16,00	1/13
		Total	N	1	1	1
			Mean		16,0000	
	8	1		23-26	25,00	1/13
		Total	N	1	1	1
			Mean		25,0000	
	Total		N	13	13	13
			Mean		7,3846	

The codes of operations with non-zero full reserve, the duration of operations with non-zero full reserve in minutes, calculated squared deviation for the duration of operations with non-zero full reserve from their average values (mathematical expectations), and the probability of operations with non-zero full reserve for network model *A* and *B*, are generated using the SPSS program respectively in table 7 and table 8.

Table 7 is generated by the SPSS program using the functions *Analyze* $\Rightarrow$ *Reports* $\Rightarrow$ *Case Summaries* $\Rightarrow$ *Summarize Cases* $\rightarrow$ *Variables:* ***Kode.X.Model.A*** ***X_EX_2.Model.A*** ***p.Model.A***, *Grouping Variable(s):* ***X.Model.A***, $\vee$ *Display cases* $\Rightarrow$ *Summary Report:Statistics* $\rightarrow$ *Cell Statistics:* *Number of cases* *Mean* $\Rightarrow$ *Options* $\rightarrow$ *Title:Case Summaries,* $\vee$ *Subheadings for totals*, where ***Kode.X.Model.A*** is the list of codes for the operations with non-zero full reserve of network model *A*, ***X_EX_2.Model.A*** are squared deviation for the duration of operations with non-zero full reserve from their average values (mathematical expectations) of network model *A*, ***p.Model.A*** is the probability of operations with non-zero full reserve for network model *A*, and ***X.Model.A*** is the time corresponding to each operation with non-zero full reserve of network model *A*. Table 8 is analogically generated with the SPSS program.

Analysing table 7 and table 8 data about the network models *A* and *B*, the following conclusions can be made:

One operation with duration $t(9,10)\,ModelA = 3$ from the section *Pernik railway yard* in direction ***Pernik - Sofia*** has the smallest squared deviation $(x_i - EX)^2\,ModelA = 0,02$ for the duration of operations with non-zero full reserve from their average values (mathematical expectations) for the network model *A*. For the network model *B, three operations* have the smallest squared deviation $(x_i - EX)^2\,ModelB = 1,00$: *one operation* with duration $t(13,15)\,ModelB = 2$ from the section *Metal - Daskalovo* in direction ***Pernik - Sofia***, and *two operations* with duration $t(i,j)\,ModelB = 4$, as one of them is the *operation* with duration $t(8,11)\,ModelB = 4$ from the section *Pernik - Pernik railway yard* in direction ***Pernik - Sofia***, and the other is *operation* with duration $t(21,23)\,ModelB = 4$ from the section *Gorna Banya* in direction ***Pernik - Sofia***. The operations with longest duration are the ones with the biggest squared deviation for the duration of operations with non-zero full reserve from their average values (mathematical expectations) of network models *A* and *B*, and they are from the section *Gorna Banya - Zaharna Fabrika* in direction ***Pernik - Sofia.*** For model *A* this is the *operation* with duration $t(25,27)\,ModelA = 9$, having a squared deviation $(x_i - EX)^2\,ModelA = 37,58$, and for model *B* this is the operation with duration $t(23,26)\,ModelB = 8$, having a squared deviation $(x_i - EX)^2\,ModelB = 25,00$.

Using the illustrated, in the present paper, approach with the SPSS program, an **average value (mathematical expectation) of the quadrate of deviation** [3,4] (for the duration of operations with non-zero full reserve from their average values) is generated in table 7 and table 8, for network models *A* and *B* for BDZ travel in Sofia regional area. Network model *A* has an **average value (mathematical expectation) of the squared deviation** (for the duration of operations with non-zero full reserve from their average values) $E(X - EX)^2 = 7,0489$, which is smaller than **the average value (mathematical expectation) of the squared deviation** $E(X - EX)^2 = 7,3846$ of network model *B*.

Using the fact, that both model *A* and *B* are generated for the route ***Pernik - Sofia*** in the morning hours, we can **make an analysis comparing the operations with equal duration** $t(12,13)ModelA = t(13,14)ModelA = t(13,15)ModelB = 2$ and $t(8,9)ModelA = t(16,17)ModelA = t(8,11)ModelB = t(21,23)ModelB = 4$, $t(24,25)ModelA = t(27,28)ModelA = t(11,12)ModelB = t(16,17)ModelB =$ $= t(26,27)ModelB = 1$ and $t(10,12)ModelA = t(12,13)ModelB = 5$, $t(1,8)ModelA = t(1,16)ModelA = t(21,35)ModelA = t(30,35)ModelA =$ $= t(1,8)ModelB = t(20,35)ModelB = t(30,35)ModelB = 0$ and $t(18,21)ModelA = t(28,30)ModelA = t(17,20)ModelB = 6$ from both network models using the obtained squared deviations for the duration of operations with non-zero full reserve from their average values (mathematical expectations). Operations of model *A* with duration $t(i,j)ModelA = 2$ have **squared deviation** $(x_i - EX)^2 ModelA = 0,76$, **less than** the squared deviation of model *B* operations with duration $t(i,j)ModelB = 2$ and $t(i,j)ModelB = 4$, which have **equal squared deviations** $(x_i - EX)^2 ModelB = 1,00$, **less than the squared deviations** $(x_i - EX)^2 ModelA = 1,28$ of model *A* operations with duration $t(i,j)ModelA = 4$. Model *A* operations with duration $t(i,j)ModelA = 1$ have **squared deviations** $(x_i - EX)^2 ModelA = 3,50$, **less than** the squared deviations of network model *B* operations with duration $t(i,j)ModelB = 1$ and $t(i,j)ModelB = 5$, which have **equal squared deviations** $(x_i - EX)^2 ModelB = 4,00$, **less than the squared deviations** $(x_i - EX)^2 ModelA = 4,54$ of model *A* operations with duration $t(i,j)ModelA = 5$. Model *A* operations with duration $t(i,j)ModelA = 0$ have **squared deviations** $(x_i - EX)^2 ModelA = 8,24$, **less than** the squared deviations of network model B operations with duration $t(i,j)ModelB = 0$ and $t(i,j)ModelB = 6$, which have **equal squared deviations** $(x_i - EX)^2 ModelB = 9,00$, **less than the squared deviations** $(x_i - EX)^2 ModelA = 9,80$ of model *A* operations with duration $t(i,j)ModelB = 6$.

From the analysis of the squared deviation for operations duration $t(i,j) = 2$ and $t(i,j) = 4$, $t(i,j) = 1$ and $t(i,j) = 5$, $t(i,j) = 0$ and $t(i,j) = 6$ for both network models *A* and *B,* we can presume that the pairs of values 2 and 4, 1 and 5, 0 and 6 for the operations' durations, are symmetrically distributed on the real axis relative to **number** 3.

For model *A*, this **number** is the operation with duration $t(9,10)ModelA = 3$ with the smallest squared deviation, and fort the model *B* this number is the **average value (mathematical expectation)** $\overline{X}.ModelB = EX.ModelB = 3,00$ for the duration of operations with non-zero reserve.

Using the statement that *the distribution of a discrete random variable* X *is symmetrical relative to the number* c*, if the values of* X *are symmetrically located on the real axis regarding the point* $x = c$ *and the probabilities, correspondent to the symmetrical pair of values are equal* [5], we can find symmetrical distributions with a centre of symmetry $x = 3$, for operations with duration 2 and 4, 1 and 5, 0 and 6, and with non-zero full reserve from the network model *B* that has **equal squared deviation from their average values (mathematical expectations).** From the mathematical expectation property, of symmetrically distributed random variable, stating that *if* X *is a symmetrically distributed random variable regarding the* $x = c$*, then* $EX = c$ [5], follows that the distributions for the model *B* operations' durations to find, and with centre of symmetry $x = 3$, have mathematical expectation $EX = 3$. Hence, for the symmetrical distribution of the network model *B* to find, the *operations* with duration 2 and 4 would have equal squared deviations $(x_i - EX)^2 = 1,00$, *operations* with duration 1 and 5 would have equal squared deviations $(x_i - EX)^2 = 4,00$, *operations with duration* 0 and 6 would have equal squared deviations $(x_i - EX)^2 = 9,00$.

To find the symmetrical distributions for the squared deviation for the duration of operations with non-zero full reserve from their average values (mathematical expectation), from table 4 we can define the network model *B* full paths including the operations with equal squared deviation. These full paths are three: $L_1 = (1,8,11,12,13,15,16,17,20,35)$, $L_2 = (1,8,11,12,13,15,16,18,19,21,23,26,27,30,35)$ and $L_3 = (1,8,11,12,13,15,21,23,26,27,30,35)$. The first full path $L_1 = (1,8,11,12,13,15,16,17,20,35)$ passes through operations with duration 2 and 4, 1 and 5, 0 and 6. The other two full paths $L_2 = (1,8,11,12,13,15,16,18,19,21,23,26,27,30,35)$ and $L_3 = (1,8,11,12,13,15,21,23,26,27,30,35)$ pass through operations with duration 2 and 4, 1 and 5 (they include operations with duration 0, but do not pass through an operation with duration 6). Using the defined full paths we can find two symmetrical distributions for network model *B*, where we can include (total) ten of the eleven operations with duration 2 and 4, 1 and 5, 0 and 6, and with non-zero full reserve from the network model B, that have **equal squared deviation from their average values (mathematical expectations)**. From the found symmetrical distributions, the second does not include operations with duration $t(1,8)ModelB = t(30,35)ModelB = 0$, because the two full paths L_2 and L_3 do not pass through an operation with duration 6, but the operation with duration $t(1,8)ModelB = 0$ is included in the first symmetrical distribution.

The first found symmetrical distribution for the squared deviations for the duration of operations with non-zero full reserve from their average values (mathematical expectation) of network model *B* (table 9) includes operations with duration $t(13,15)ModelB=2$ and $t(8,11)ModelB=4$, $t(11,12)ModelB=t(16,17)ModelB=1$ and $t(12,13)ModelB=5$, $t(1,8)ModelB=t(20,35)ModelB=0$ and $t(17,20)ModelB=6$, through which the full path of network model *B* $L_1=(1,8,11,12,13,15,16,17,20,35)$ passes. Table 9 is generated be SPSS program with the functions *Analyze* $\Rightarrow$ *Reports* $\Rightarrow$ *Case Summaries* $\Rightarrow$ *Summarize Cases* $\rightarrow$ *Variables:**p.L1.Model.B X.1.L1.Model.B X.2.L1.Model.B X.3.L1.Model.B X.4.L1.Model.B*, *Grouping Variable(s):* ***X_EX_2.L1.Model.B***, $\vee$ *Display cases* $\Rightarrow$ *Options* $\rightarrow$ *Title:Case Summaries,* $\vee$ *Subheadings for totals.*

Table 9

Case Summaries

			p.L1.Model.B	X.1.L1.Model. B	X.2.L1.Model. B	X.3.L1.Model. B	X.4.L1.Model. B
X_EX_2.L1.Model.B	1,00	1	1/6	t (13,15)=2	t (13,15)=2	t (13,15)=2	t (13,15)=2
		2	1/6	t (8,11)=4	t (8,11)=4	t (8,11)=4	t (8,11)=4
	4,00	1	1/6	t (11,12)=1	t (16,17)=1	t (11,12)=1	t (16,17)=1
		2	1/6	t (12,13)=5	t (12,13)=5	t (12,13)=5	t (12,13)=5
	9,00	1	1/6	t (1,8)=0	t (1,8)=0	t (20,35)=0	t (20,35)=0
		2	1/6	t (17,20)=6	t (17,20)=6	t (17,20)=6	t (17,20)=6

All operations combinations of network model *B* full path L_1, which have duration $t(13,15)ModelB=2$ and $t(8,11)ModelB=4$, $t(11,12)ModelB=t(16,17)ModelB=1$ and $t(12,13)ModelB=5$, $t(1,8)ModelB=t(20,35)ModelB=0$ and $t(17,20)ModelB=6$, and which have **equal squared deviations**, respectively $(x_i-EX)^2=1,00$, $(x_i-EX)^2=4,00$ and $(x_i-EX)^2=9,00$ are given **in columns** *X.1.L1.Model.B, X.2.L1.Model.B, X.3.L1.Model.B, X.4.L1.Model.B* of table 9.

The second found symmetrical distribution for the squared deviations for the duration of operations with non-zero full reserve from their average values (mathematical expectation) of network model *B* (table 10) includes operations with duration $t(13,15)ModelB=2$ and $t(8,11)ModelB=t(21,23)ModelB=4$, $t(11,12)ModelB=t(26,27)ModelB=1$ and $t(12,13)ModelB=5$, through which the full paths of model B

$L_2=(1,8,11,12,13,15,16,18,19,21,23,26,27,30,35)$ and

$L_3=(1,8,11,12,13,15,21,23,26,27,30,35)$ pass.

Table 10

Case Summaries

			p.L2.L3. Model.B	X.1.L2.L3. Model.B	X.2.L2.L3. Model.B	X.3.L2.L3. Model.B	X.4.L2.L3. Model.B
X_EX_2.L2.L3.Model.B	1,00	1	1/4	t (13,15)=2	t (13,15)=2	t (13,15)=2	t (13,15)=2
		2	1/4	t (8,11)=4	t (21,23)=4	t (21,23)=4	t (8,11)=4
	4,00	1	1/4	t (11,12)=1	t (11,12)=1	t (26,27)=1	t (26,27)=1
		2	1/4	t (12,13)=5	t (12,13)=5	t (12,13)=5	t (12,13)=5

Analogically to table 9, table 10 is generated by SPSS program. All operations combinations of network model *B* full paths L_2 and L_3, which have duration $t(13,15)ModelB = 2$ and $t(8,11)ModelB = t(21,23)ModelB = 4$, $t(11,12)ModelB = t(26,27)ModelB = 1$ and $t(12,13)ModelB = 5$, and which have equal squared deviations, respectively $(x_i - EX)^2 = 1,00$, $(x_i - EX)^2 = 4,00$ are given **in columns** *X.1.L2.L3.Model.B, X.2.L2.L3.Model.B, X.3.L2.L3.Model.B, X.4.L2.L3.Model.B* of table 10.

3. Conclusion

There are three main results obtained in this paper.

First is the original method for generating network models when planning a route with passenger railway transport in Sofia regional area, illustrated in this paper.

Second is the presented approach for defining the squared deviations for the duration of operations with non-zero full reserve from their average values (mathematical expectations), of network models for BDZ travel in Sofia regional area, using the SPSS program.

Third is the analysis of calculated squared deviation for the duration of operations with non-zero full reserve from their average values (mathematical expectations) and found symmetrical distributions for squared deviations of certain network model.

References

1. Balashevich V., Matematicheskiye metody v upravlenii proizvodstvom, Izdatelstvo „Vysheyshaya shkola", Minsk, 1976
2. Bakalova V., T. Minkov, Prognozirane i planirane na transportnata dejnost, Sofia, Izdatelski kompleks - UNSS, 2011.
3. Sugarev Z., S. Kamenarov, Teoriya na veroyatnostite, Sofia, Izdatelstvo „Nauka i izkustvo", 1979.
4. Ekimov T. S., Vissha matematika vtora chast, Pernik, 1998.
5. Uzunov M., Rakovodstvo po teoriya na veroyatnostite, Sofia, Universitetsko izdatelstvo „Stopanstvo", 1998.
6. Todorov D., K. Nikolov, Matematika, Sofia, Universitetska pechatnitsa - UNSS, 2009.
7. Ivanova M., M. J. Ivanova, I. Ivanov, Matematika II primerni testove za samopodgotovka na studenti v ikonomicheskite visshi uchebni zavedeniya, Sofia, Izdatelski kompleks – UNSS, 2019.
8. Bozhinov N., Mathematics Part Two, Sofia, UNWE Publishing Complex, 2013.
9. Goev V., Statisticheska obrabotka i analiz na informatsiyata ot sotsiologicheski, marketingovi i politicheski izsledvaniya sas SPSS, Universitetska pechatnitsa - UNSS, Sofia, 1996
10. Kaloyanov T., V. Petrov, Statistika, Sofia, Izdatelski kompleks – UNSS, 2014.

Application of Modern Web Solutions in the Scientific Periodical Journals

Dimitar Dimitrov, Hristo Yankov

Todor Kableshkov University of Transport – Sofia, Bulgaria
ddimitrov@vtu.bg, hryan@vtu.bg

Abstract. The document examines current trends in web solutions in scientific periodical journals. There are not many periodicals in Bulgaria, and the available ones cover a relatively narrow field of knowledge, which makes it difficult to publish good articles. The purpose of this journal is to promote the achievements of Bulgarian scientists and to publish material from foreign authors of interest to the professional community in the field of mechanics, transport and communications. The application of the project is related to a site that is based on modern web technologies. The presented algorithm in the article determines the technological consistency of the process of registering and receiving reports for scientific forums. The organization of scientific forums at Todor Kableshkov University of Transport is a traditional task with many years of experience, which was a serious prerequisite for very clearly defined and systematized technological and organizational processes. The information system includes two main modules: general information module and administrative module. The system is built on the basis of modern web technologies. The language used is PHP and the database is the widely used MySQL.

Keywords. Scientific, Periodical journal, Web solutions, Website, PHP programming, MYSQL database, Mechanics, Transport, Communications.

1. Introduction

There are not many periodical scientific journals in Bulgaria. The available ones cover a relatively narrow field of knowledge, which makes publishing of good articles difficult. This explains why the number of their readers is too small. Very few articles are published in a foreign language, so the number of the readers decreases even more. The aim of this journal is to unite the interests of more scientists due to widening the subject area and the bilingual publishing. The journal is going to make popular Bulgarian scientists' achievements and to publish materials by foreign authors that are interesting for the professional community.

The purpose of this magazine is to promote the achievements of Bulgarian scientists and to publish material from foreign authors of interest to the professional community in the field of mechanics, transport and communications. The application of the project is related to a site that is based on modern web technologies.

2. General Concepts

The use of web-based systems in the field of research and publication of scientific production has a long tradition, and these systems are being upgraded/updated on the basis of the latest information technologies and tools for their presentation and publication. Conducting scientific forums is an organizational and planning task aimed at gathering and analyzing the current achievements of scientific potential in a given field, on the one hand, and on the other,

to ensure adequate planning of material and human resources related to their conduct and publication.

The organization of scientific forums at Todor Kableshkov University of Transport is a traditional task with many years of experience, which was a serious prerequisite for very clearly defined and systematized technological and organizational processes.

The information system includes two main modules:

General information module - in which, through a system of screens, general information about the organization of scientific forums, program and organizational committees, as well as the publication of materials (full text, abstract, bibliography, etc.) is presented.

Administrative module - for processing of incoming reports on scientific forums, reports/ reference, reviews of reports and correspondence with authors, etc.

The basic algorithm of the process is presented in Fig. 1.

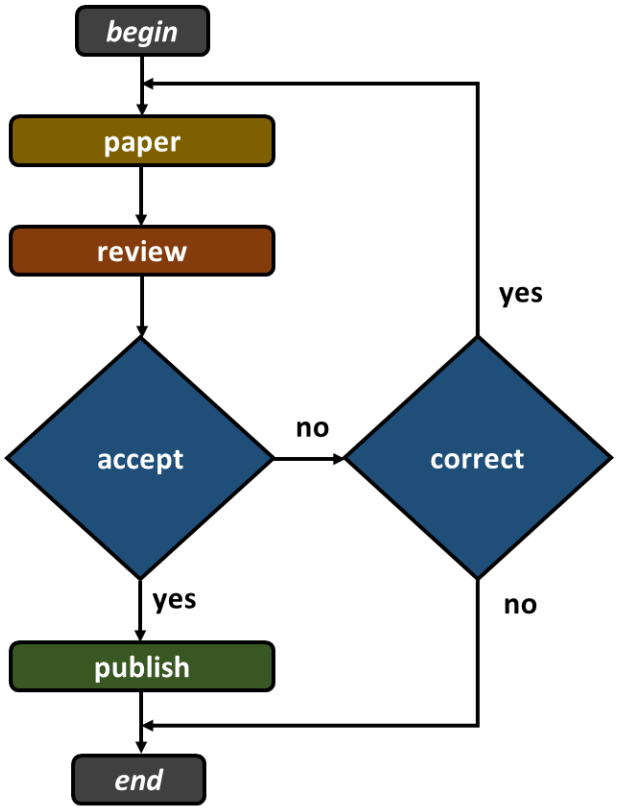


Fig. 1. Algorithm for accepting, reviewing and publishing scientific reports.

The business logic of the scientific production recruitment process is part of the advanced functionality of the algorithm and determines the technological consistency of the process of registering and receiving reports for scientific forums. This model aims to test and improve the quality of scientific output of the participants in the relevant scientific forum.

3. Scientific trends

Trends of this journal include the following scientific and applied fields:

- Technology, Organization and Management of Transport
- Safety and Reliability of Transport
- Economic Issues of Transport
- Trends and Innovations in Transport Education
- Engineering Logistics and Building Equipment
- Transport Equipment
- Mechanics, Mechatronics and Mathematics
- Transport Infrastructure
- Transport and Ecology
- Electric Power Systems and Equipment in Transport
- Telecommunications and Signaling Equipment, Automation Systems in Transport
- Electronics, Automation and Mechanization
- Materials Science and Technologies
- Architecture and Civil Engineering
- Maritime Transport and Technologies

These scientific fields can be dynamically supplemented in the journal system, including that they can be edited and supplemented.

The participants should register and submit their abstracts and papers only at the journal website. The papers are published after a confirming review. The papers typically can be up to 6 pages and should be sent as doc-files keeping the specific parameters

4. MTC – AJ Web Project

Mechanics Transport Communications - Academic Journal (MTC-AJ) is an applied web project for a scientific editions of Todor Kableshkov University of Transport – Sofia (from Fig. 2. to Fig. 9) [1, 2, 3].

The aim of this journal is to unite the interests of more scientists due to widening the subject area and the bilingual publishing. The journal is going to make popular Bulgarian scientists' achievements and to publish materials by foreign authors that are interesting for the professional community.

The journal starts as an electronic publication issued semi-annually. Its printed version is going to be deposited in main national libraries. The printed version of the journal will be available under certain conditions. At the moment we are working on making the subscribing possible.

All the suggestions for publishing are welcome, but the editors will approve of only the ones whose scientific and scientific-engineering quality has been claimed by two reviewers having academic ranks (external and internal to the university) and appointed by the editorial board.

The International Editorial Board consists of recognized scientists from Bulgaria, Russia, Austria, Spain, Germany, Serbia, Slovak Republic, Poland, Romania and Turkey.

The scientific Journal “MECHANICS TRANSPORT COMMUNICATIONS” is indexed in the following worldwide databases, such as:

- ROAD Directory of Open Access Scholarly Resources (https://road.issn.org/);
- eLIBRARY.RU (https://elibrary.ru/);
- SIS – Scientific Indexing Service (http://www.sindexs.org);
- Google Science (https://scholar.google.com/);
- J-Gate (https://jgateplus.com/home/);
- ERIHPLUS – European Reference Index for the Humanities and Social Science (https://dbh.nsd.uib.no/publiseringskanaler/erihplus/).

Fig. 2 MTC-AJ Homepage

The participants should register and submit their abstracts and papers only at the conference website. The papers are published after a confirming review. The papers /including main text and references/ can be up to 6 pages and should be sent as doc-files.

The decision to accept or refuse to publish a proposed material is made on the basis of its scientific value, criteria for publication in the journal and in accordance with the principles of academic morality, its ethical principles and rules, and the editorial policy of the journal.

The journal does not accept materials that:

• are not in accordance with its scientific policy and publishing criteria;

• contain unacceptable and pseudoscientific theories and claims;

• express political and personal biases of the authors;

• include texts that offend individuals and institutions or exhibit negative personal sentiments (offensive qualifications on a scientific, religious, political or personal basis).

The authors are directly responsible for the accuracy of their scientific publications and must adhere to the citation rules indicated.

It is unacceptable:

• plagiarism in all its forms;

• falsifying data and results in the material presented;

• co-authoring persons who have no real input into the research and into the writing of the relevant text;

• use of unscientific arguments;

• deliberate omission of contributions from other authors in the field of study;

• publication of offensive material, as well as material that does not contain essential, new or additional reasoning.

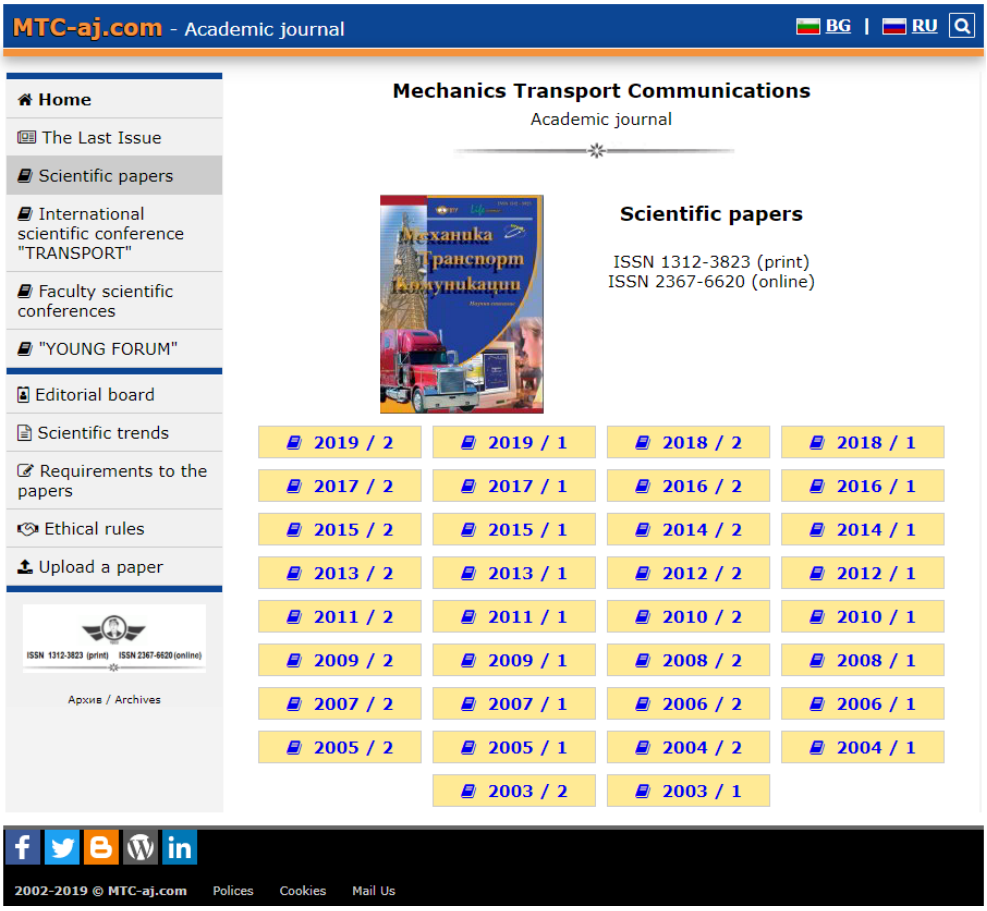


Fig. 3. MTC-AJ Scientific papers

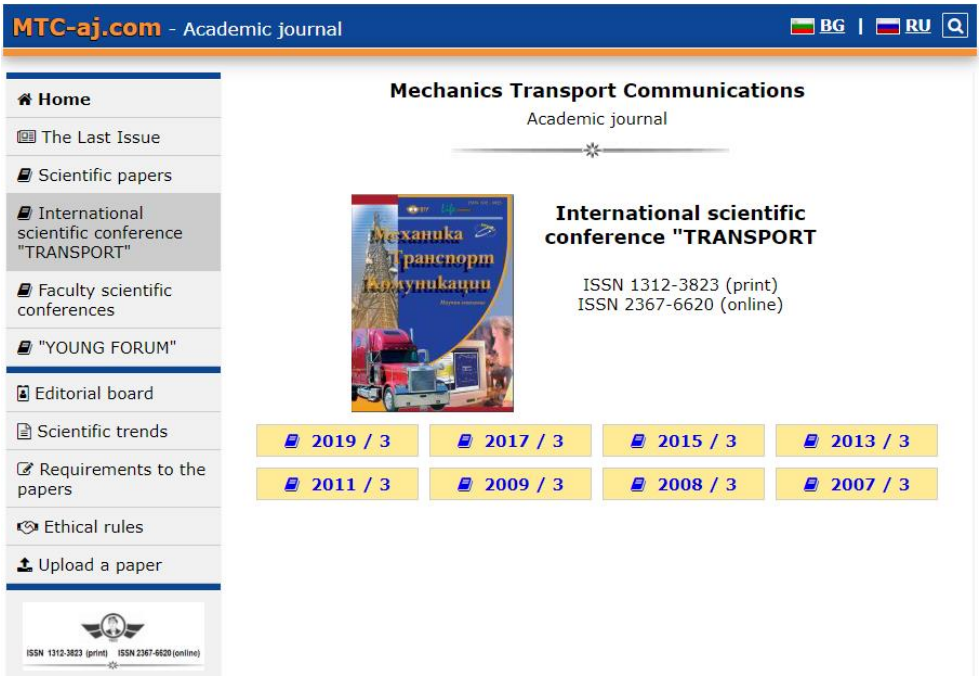


Fig. 4 MTC-AJ International Scientific Conference “TRANSPORT”

Fig. 5. MTC-AJ Faculty Scientific Conferences

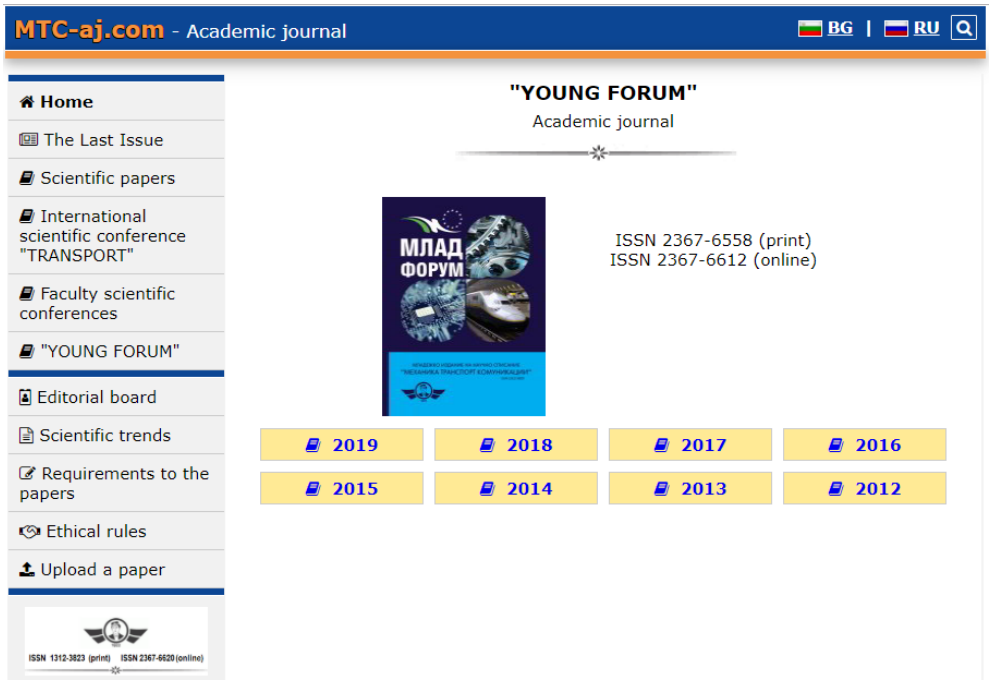
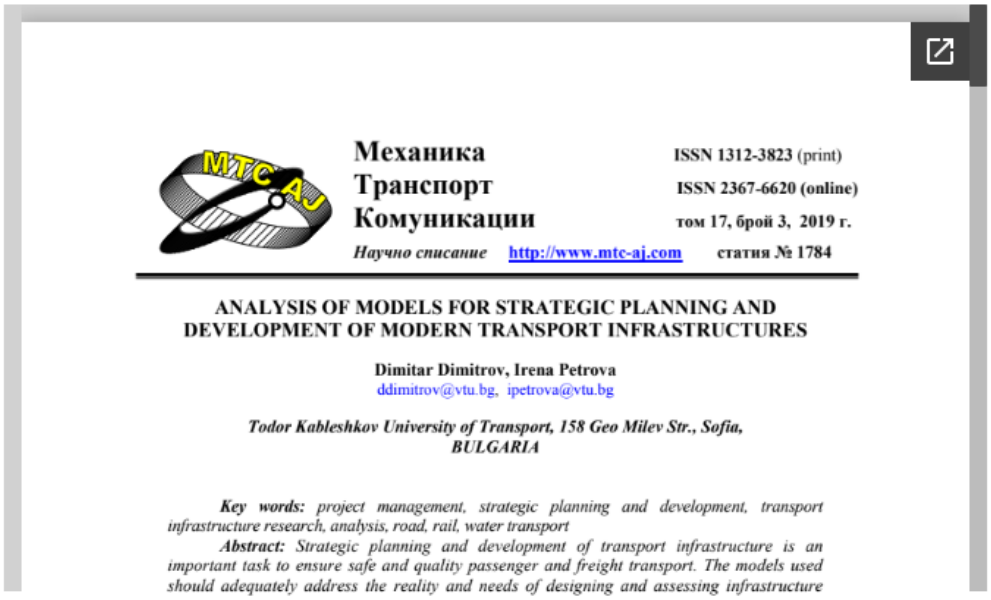


Fig. 6. MTC-AJ "Young Forum"

Fig. 7. MTC-AJ Abstract View



Механика
Транспорт
Комуникации

ISSN 1312-3823 (print)
ISSN 2367-6620 (online)
том 17, брой 3, 2019 г.
Научно списание http://www.mtc-aj.com статия № 1784

ANALYSIS OF MODELS FOR STRATEGIC PLANNING AND DEVELOPMENT OF MODERN TRANSPORT INFRASTRUCTURES

Dimitar Dimitrov, Irena Petrova
ddimitrov@vtu.bg, ipetrova@vtu.bg

Todor Kableshkov University of Transport, 158 Geo Milev Str., Sofia, BULGARIA

Key words: *project management, strategic planning and development, transport infrastructure research, analysis, road, rail, water transport*

Abstract: *Strategic planning and development of transport infrastructure is an important task to ensure safe and quality passenger and freight transport. The models used should adequately address the reality and needs of designing and assessing infrastructure*

open/download as PDF

управление на проекти | стратегическо планиране и развитие | транспортна инфраструктура | проучване | анализ | автомобилен | железопътен | воден транспорт | project management | strategic planning and development | transport infrastructure research | analysis | road | Dimitar Dimitrov | Irena Petrova

Fig. 8. MTC-AJ Full Text View

BIBLIOGRAPHY

[1.] Antov A., Transportno modelirane 101 - prakticheskо rakovodstvo, Sofiya, 2017, http://www.infrarch.com/101/?nav=0
(*[1.] Антов А., Транспортно моделиране 101 - практическо ръководство, София, 2017,* http://www.infrarch.com/101/?nav=0)

[2.] Varadinova-Milkova Yu., "Modeli za ustoychivo razvitie na zhelezopatnata infrastruktura", VTU "Todor Kableshkov", Sofiya, 2014.
(*[2.] Варадинова-Милкова Ю., "Модели за устойчиво развитие на железопътната инфраструктура", ВТУ "Тодор Каблешков", София, 2014.*)

...
...
...

[14.] Valkov R. and Krumov Z. 2010 Infrastructure Project Implementation Within Crisis Conditions, Collection Z 15, Int. Sci. Conf. Resolving Crisis Situations (Zilina Slovakia) 789–94

[15.] Oughton Edward J., Usher Will, Tyler Peter and Hall Jim W, Infrastructure as a complex adaptive system, Complexity, Volume 2018, Article ID 3427826, 11 pages, https://doi.org/10.1155/2018/3427826

Fig. 9. MTC-AJ Bibliography View

5. Conclusion

This publication presented the implemented project of Web-oriented information system for organizing and conducting scientific forums, as well as publishing of scientific literature at Todor Kableshkov University of Transport [4, 5, 6].

The functional characteristics of the system needed for the whole process of organizing and managing information flows and publishing of scientific publications were defined.

The system is built on the basis of modern web technologies. The language used is PHP and the database is the widely used MySQL [7, 8]. The many years of work of the system has proven its high degree of stability and robustness, which fully meets the initial requirements for automation of the periodical scientific literature process.

References

1. Mechanics Transport Communications - Academic Journal Website (https://www.mtc-aj.com/)
2. Todor Kableshkov University of Transport Website (https://www.vtu.bg/)
3. International scientific conference "TRANSPORT" Website (https://conference.vtu.bg/)
4. Dimitrov D., Model for Building Integrated Information Systems for University Management, Todor Kableshkov Publishing House, ISBN 978-954-12-0248-7, Sofia, 2018, https://ddimitrov.vtu.bg/miiisuu.pdf
5. Dimitrov D., Laboratory base for training and experimental research in the field of information and internet technologies at Todor Kableshkov University of Transport, Scientific journal Mechanics Transport Communications, ISSN 1312-3823, Issue 1, IV-1 IV-6, 2015, Art. No. 1107, 2015, https://mtc-aj.com/library/1107.pdf
6. Dimitrov D., Iliev H., Designing and Building a Web-Oriented System for Organizing and Holding Scientific Forums, XXI International Transport Conference 2013, Varna, 2013, https://mtc-aj.com/library/823.pdf
7. https://www.php.net/
8. https://www.mysql.com/

Two Different Teams - One Goal

Svetlana Vasileva, Veneta Egova, Mariyana Panayotova

NMHS, Dobrich, Bulgaria
svetlanaeli@abv.bg, eti_ned@abv.bg, pmg_gabrovska@abv.bg

Abstract. Usually we perceive information technology as computer information technology. But the information (and communication) technology is much more than the computer technology. The paper analyzes the participation of two different school teams of mathematical high school in seemingly quite different competitions for students who were held in Bulgaria in the summer 2019. The team named RoboDobrich, composed of students aged 13-16 years with interests in the field of computer sciences and information technology participated in the finals of the competition program "Robotics for Bulgaria". And the team named "Albena" composed of students aged 17-18 years with very different interests of robotics (biology, chemistry, geography, physics) participated in the first International Competition for students "Karst - the Last White Spot on the planet Earth". Both teams prepared with great diligence on the part of the children, they had mentors and teachers from different fields and at the competitions they "played" as one team combining knowledge and different skills. And, of course, preparing with a huge ambition to meet the challenges of the corresponding competition, they learned a lot and were extremely satisfied after the end.

Keywords. Information technology, Lego robotics, Karst, Systematic approach, Integrating knowledge and skills, Team work.

1. Introduction

„Emotional journey! There is no other such competition! ... A strict schedule,... young volunteers supporting the teams. An exceptional competition that gave us the opportunity to show our skills in presentation, logical thinking, teamwork, dealing with critical situations, working under pressure, problem solving, communication skills, programming and all this with a lot of fun. The Robotics for Bulgaria Festival was a sea of emotions for us, anxiety, tension, smiles, adrenaline, confidence, happiness, and most importantly a stimulus and inspiration for new successes in the next competition. At one point we were desperate because our robot blocked, but there were people who encouraged us and rejoiced in our success! They appreciated our team game, which is the most important thing in a team. We succeeded and appreciated the skills of the team members to communicate with each other, to help and support each other in the competition process. SAP Labs Bulgaria and the Workshop for Civic Initiatives Foundation (WCIF) did everything possible to make us feel wonderful and 3 months after the competition we can't wait for the next visit!" – This is how the captain of the RoboDobrich team, participated in the competition "Robotics for Bulgaria", 2018/2019 season in the interim report, answers the question "Tell us more about the participation in the Robotics Festival". [1].

"Through the regulations of this first and unique international competition, we offer participants creative freedom to express knowledge and skills about karst, to demonstrate in direct contact with a karst environment. We appeal to all teachers who are not indifferent to karst - help your students who are attracted to the challenges of karst and field research, to join a team and prepare to participate in this competition… Karst has something to teach us. But

we can and we must pay tribute karst ... Together we can more!" - From the invitation to participate in the first International Competition for students " Karst - the Last White Spot on the planet Earth".

The regulations and content of this International Student Competition on Karst are primarily in line with the systemic nature of karst, which allows [3]:

To unite in a team some students with skills and preferences for different scientific fields, but with a lasting interest in karst and emotionally motivated to work in an attractive environment. The members of the team jointly use the acquired knowledge in the different subjects, make sense of the connections between what has been learned and make attempts to apply it to solve a practical case in the real conditions of unknown karst territory;

To apply research methods in solving specific tasks in natural field conditions;

To motivate teachers in different subjects and with different professional specialization to work and experiment with options for interdisciplinary connections through karst. All this is checked and evaluated in a "competitive environment in solving specific practical tasks in an unknown karst territory." [3]

In this work we will do parallel between these two competitions and the participations of the two teams that were formed by the correspondingly team mentors - teachers in the mathematical high school in Dobrich city. We will also analyze and summarize the conclusions of the mentors of the two teams: Students "develop skills in teamwork, not only to think as part of a whole, but also to enjoy such collective work" [4]; they acquire a lot of new knowledge, while having fun and playing; and last but not least, they receive recognition and feel winners after the competitions.

2. Rules for Team work and Team Thinking

In both competitions, the teams are prepared under the guidance of two mentors and participate and are evaluated as a team. As early as 2018, working on a project for a student scientific conference, students showed and demonstrated their attitude and statement about teamwork: "think together as a whole better than alone" [4].

In the description of Task Teamwork 1 [5], the robotics team unanimously identified the following **Core Values** [5] and [6]:

- **Discovery**,

- **Integration** (pooling of knowledge and skills),

- **Teamwork** and **solving** complex things while **having fun** are the values we profess and at the same time, they are a necessary and sufficient condition for the existence of a team;

- A team with such qualities is always **successful**;

- **Overcoming** the **difficulties** of tasks and missions, we **find** a way to each other and **control** ourselves;

- **Diversity of opinions** is the road to the right decision.

- The **problems** for the team are **challenge** and **fun**, **self-discovery** and **inspiration**.

The competition "Robotics for Bulgaria" is held in three stages (Task 1, Task 2 and Task 3), stretched over time in 5 months. As for each task / stage, the team has about a month to perform the necessary work and upload the correspondingly reports to the website of "Robotics for Bulgaria". Then, within two weeks, the reviewers evaluated the submissions Task Team, Task Project, Task Robot Design and Task Robot Game. Analyzing the shortcomings pointed out by the reviewers and their recommendations for improvements, the team tries to improve the work and achieve higher results at the next stage. In fig. 1 it is shown footage from the participation of the team of the festival "Robotics for Bulgaria".

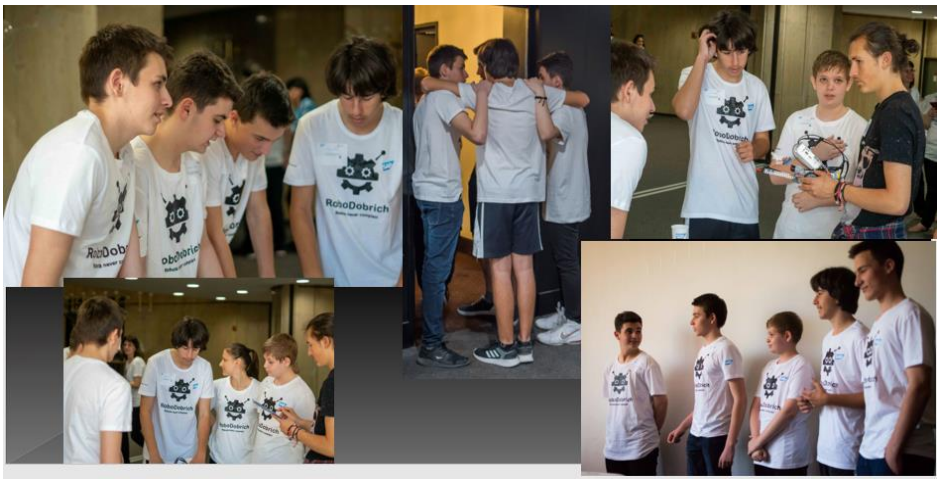


Fig. 1 Team work of RoboDobrich at the finals of "Robotics for Bulgaria"

The International Student Competition "Karst - The Last White Spot on Planet Earth" is also held in three rounds, but in three consecutive days. The other difference from the robotics competition, where reviews are received directly, is that here the feedback on how the team coped with the tasks is in the form of informal statements, comments or explanations from the jury. And to a certain extent, the team collecting these individual pieces of information had the opportunity to draw the appropriate conclusions and fill in the gaps until the next stage.

The general rules regarding teamwork in the Student Contest on Karst are the following: team members have preferences for different subjects and scientific fields, but are united by continued interest in karst. They work in a team and complement, develop and upgrade their knowledge and skills. In addition to theoretical knowledge of karst based on its systemic nature [7], team members must have practical skills acquired through observations and research in a karst area under the guidance of a mentor. Each team is prepared by a team of teachers in different subjects, including a mentor (a researcher from a scientific / academic institution or an expert from the administration of a protected karst area) [2], [3] and [8].

Proof that the successful team is a successful combination and integration of the individual knowledge and skills of the competitors are the additional competitive events - there is 1 representative nominated by each team. „The aim is to provide the members of the student teams with additional opportunities for individual manifestations of abilities and skills in three areas: Sports competition: The challenges of karst; Outdoor art studio: The attractiveness of karst (Nature as a sculptor); Leaders in Information and Communication technologies: Solving karst-related tasks that are subject to further evaluation." [2] and [3].

Fig. 2 shows some moments from the team work of the team „Albena" in the 3 stages of the Student Competition on Karst.

If we go back to the Robotics team, to the question from the interim report "What additional activities have been carried out by the students from the team regarding teamwork, presentation skills, activities outside the team?" [1] the answer was as follows:

Most of the students knew each other from math and physics competitions. Five of the students were from 8a, the two girls were from 8b, the two youngest were from 7a. It can be said that the children knew each other even before the formation of the team, and the enthusiasm was so great that overcoming the difficulties and the joy of successfully completing

the missions united the team very much. They relied on each other from the very beginning - they distributed by list who will provide what consumables for completing the field [2] and [3].

Fig. 2 Team work in the karst territory

Together they prepared the club room, and before that they asked the director of the High School to give the team one of the free rooms. With great enthusiasm they cleared and tidied up the office. The other thing that socialized them was playing basketball and table tennis. To develop presentation skills, the mentors had to help giving them ideas on how to present each other, proving that they really became a team.

3. New Knowledge and New Skills

Preparation for both competitions included the acquisition of new knowledge and skills. In the robotics team, this was done concurrently in two ways:

- Imperceptibly - when the students were trying to understand why the robot does not perform a certain mission [2], [9] or performs it incorrectly;

- Obligatory - during the work on the Terms of Reference project [2], [3], which in the 2018/2019 Season required the teams to solve a certain human physical or social problem in the field of space exploration and to propose a solution.

Preparation of research projects by the robotics team. Three of the team members have a strong interest in physics. The decision of the team the project to consider the problem of the movement of the inhabitants of the future colonies on planets with low gravity and in particular on the Moon (movement in the open environment) was made by the physicists in the team and the others unanimously supported the idea. The team considered that the solution to this problem would be important for the faster and calmer movement of the inhabitants of these colonies between the modules. Additional literature was used to prepare the project and films were watched and discussed on Youtube [10].

Preparation of the team to the Student Contest on Karst. Due to the specifics of the karst and its systematic nature [8], the Albena team had to acquire new knowledge (in Chemistry, Geography, Physics, etc.) and skills continuously and at an accelerated pace from the beginning of the preparation for the competition, during each competition round. Until the last day. The competition rules included [8]:

- First round of competition: Test of theoretical knowledge about karst and karst systems. It is held in a hall and includes individual solving of tests by all members of the student teams of the teams.

- Second competition round: Field observations, research and experiments on a given and traced route in karst terrain to test practical skills in solving research problems. Field work includes: Descriptions of observed processes / situations; Work with maps and orientation in the area, incl. in a cave; Work in a natural laboratory (field trials and experiments); Data collection from demonstrated field research methods and monitoring; Solving set practical tasks. The route is prepared in advance, with marked control points, signposts and information boards, observation points, points for field research and experiments (equipped with research equipment and a specialist to control and support the analysis and experiments), a natural laboratory for analysis of questions), etc. The descriptions made by the competitors and the results collected by them and processed from the field research and experiments made up the working (data) bases for the next third competition round. The teams also receive additional materials and research data on the karst in the area of the competition, provided by the organizers.

- Third round: Solving a given practical case (problem, situation) in the area of the competition. It is held in two parts: Preparation of the decision: independent work of the student teams of the teams (in the office); Public defense (in a hall). The teams receive on paper a description of a problem situation and related specific tasks for which to offer optimal solutions.

Teams work in a team and have a fixed time available. Appropriate working conditions are provided - computer equipment and the necessary tools. The teams prepare team-coordinated author's ideas and interpretations of the given practical case. The team decision is presented and defended publicly, and for this purpose freely chosen forms and means can be used, incl. multimedia presentation.

This extremely busy and full of impressions and knowledge and skills competition was in fact training in real conditions and learning through empathy, because there was one main advantage stemming from the systemic nature of karst, which requires interdisciplinarity, both in its research and training for it [8].

Therefore, in the formation and preparation of teams it is necessary to include students with different scientific preferences and abilities, united by interest in karst and research and experiments in an attractive environment, as well as teachers in different subjects and with different professional specializations.

The second advantage is that the competition creates a real opportunity to test, "in competitive conditions, on different options for interdisciplinary links - one of the most current modern educational problems." [8].

4. Results of the training and participation

In the interim report [1] of RoboDobrich team about the quality results of the activity in the 2018/2019 season it was noted that from an absolute beginner the team became extremely successful through: Discovery, Integration of knowledge and skills, teamwork and solving complex things, having fun; finding a way to each other and mastering oneself; color

of opinions as a way to the right decision; approaching problems such as challenge, self-discovery and inspiration. In other words: the **core values** formulated in the initial stage gradually **led** the team to a successful participation in the finals of the competition. Quantitatively, this can be described as follows:

- Ranking after Task 1 - 23rd place in Division B;
- Ranking after Assignment 2 - 13th place in Division B;
- Ranking after submission of Assignment 3 - 7 place and
- Ranking for the finals of the competition in Division B.

The honors of the finals in the category "Team game" (2nd place) and in the category "Robot design" (3rd place).

To demonstrate the results of the team that participated in the Student Competition on Karst, we will cite the analysis of the results of the performance of the teams in the various competitive events, made by the organizers in [8].

The results of the first competition round (theoretical knowledge and skills for team solving practical tasks in the office) „are relatively weak in all five teams“ - an average of 57.7 points out of 120 possible. Even the best performing team of Burgas has only 75.1 points. In the team solving of practical tasks “the results of the teams are very close” [8], as the team of Teteven has a slight advantage - 33 points with an average score of 29.2 points (at a maximum possible 45 points).

In the field work the best results were achieved by the teams of Latvia and Teteven (49 points each) with an average score of 41.8 points (at 60 possible).

In solving and defending the practical case, the presentation of the teams according to the pre-set 3 criteria are the following:

1. Theoretical knowledge: with a maximum possible 50 points, three teams have a score of over 40 points (Burgas, Latvia, Dobrich) - 42.2, 41.4 and 40.8 points, respectively.

2. Originality and technical skills in the chosen form of presentation: with a maximum possible 20 points, the teams have an average score of 14.85 points. “All showed their creative skills” [8] and the difference between the first team (Burgas - 17.9 points) and the fifth - Yablanitsa - 9.3 points) in the ranking according to this criterion is only 8.6 points.

3. Originality and persuasiveness in public presentation and defense: “even with this criterion, the difference in the evaluations of the teams is small, because they captivated the audience with their enthusiasm and persuasiveness in defending their decisions” [8]. The average score is 15.5 points out of 20 possible, and the difference between the first team (Dobrich - 18.9 points) and the fifth (Teteven - 10.3 points) is again 8.6 points.

In the final ranking in the third round the first 3 teams have almost the same results: 71.2 points (team Burgas), 70.9 points (team Latvia) and 70.0 points (team Dobrich) with an average score of 63.5 points (at 90 possible).

In the final ranking of the competition the average result of the teams is 185.2 points out of 300 possible. The first three teams (Burgas, Dobrich, Latvia) have an average score of 199 points, and “the difference between them is insignificant - only 14 points - proof of a very equal performance” [8].

The conclusions about the results achieved by the Albena team are as follows: During the first two stages, where there are still elements of learning, the team performs solidly and well and is still learning. In the third stage, all the knowledge and skills acquired until the last moment, including the soft skills and team thinking possessed by the team members, proved to be brilliant and the defense of the case decision was highly praised by the jury. Another proof of the huge contribution of the additional knowledge and skills of the team is the first place in the additional competitive event Leaders in Information and Communication

technologies, where conditionally one of the team members won first place with 9.3 points. The attention to detail (which is important not only in research in the natural sciences, but also in the technical sciences) also contributed to the wonderful performance not only of the Karst team, but also, as described above, of the robotics team.

Moreover, if we compare the results in the individual competitions, the team "Albena" is the absolute winner, given that all five team members have interests mainly in biology and chemistry (and not in Information and Communication technologies): Sports competition - The challenges of karst (team Latvia): 9 points; Outdoor art studio: The attractiveness of karst - Nature as a sculptor (team Yablanitsa): 9 points; Leaders in Information and Communication technologies: - Solving karst-related tasks that are subject to further evaluation (team Dobrich): 9.3 points.

"Unlike other international student competitions and Olympiads held in school subjects, this competition is for karst, which is not a discipline in the education system." [8] Therefore, participation in the competition requires from the students not only interest and motivation, but also serious and continuous preparation in extracurricular activities. Another aggravating moment should not be missed - the competition is held in extracurricular time, in the middle of summer - the time of vacations and personal commitments.

At the same time, there is a great interest among students in karst and caves, which are very attractive, attractive and challenging for this age. Thus, the team "Albena" was composed not of the best students in science in high school, but of good enough and motivated students who had no other commitments during the competition. But, as noted by the organizers and the jury, "even with insufficient preparation, the five teams achieved a decent performance. Each team had an advantage in some of the competitive events, and the total sum of demonstrated knowledge, skills and team effectiveness was reflected in the final ranking" [8] "In this competition, **the emphasis is not on "fighting for medals", but on demonstrating the ability to apply knowledge and skills in a real karst environment. Moreover, leading is not the personal performance in a competitive environment, but teamwork - the success of the team depends on the "cooperation" between its members (leading is the synergy: team decisions -> gathering knowledge and skills of individual members**). The preparation and participation in the competition, stimulate the all-round development and enrichment of the young person, as well as his/her physical endurance and mental stability, will, purposefulness, spatial orientation, technical skills, etc." [8].

The results for the team that took part in the finals of "Robotics for Bulgaria" are similar. In fig. 3 it is shown the news, which announces the competition for funding robotics teams in schools where no robotics teams have been formed yet. The ad reads [5]: "

The program supports teachers by providing specialized training, online support, equipment for the activities of a team - a set of Lego with a programmable module - Mindstorm [9] and a competitive field. By supporting teachers to mentor a school robotics team, the program encourages students' interest in science and technology and gives them the opportunity to work in a team as true researchers… Divided into school robotics teams, students design and program robots that perform socially significant missions. As part of the initiative, students develop research projects related to environmental protection. The added value of the initiative is the skills acquired by children for teamwork. The initiative stimulates students' interest in science and technology and provides an opportunity for learning through experience" [5].

Therefore, with great joy and satisfaction, we noted that in the photo of fig. 3 [11] the photographer from the competition captured the robot created by the RoboDobrich team!

Fig. 4 [12] shows a moment from the finals of "Robotics for Bulgaria". In the photo, the photographer captured the team during a break between the sections of the competition.

Fig. 3 The page from the website of the WCIF with published photo of the robot designed and constructed by the RoboDobrich team

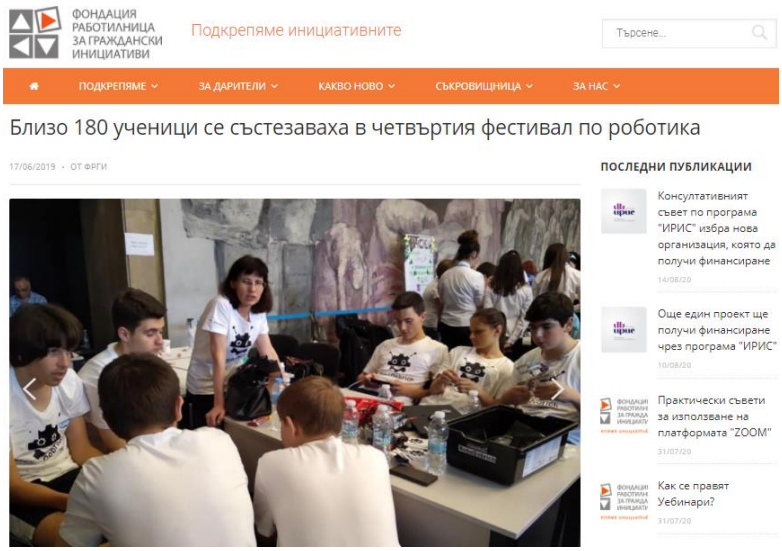


Fig. 4 The website of the WCIF with published album with selected photos from the finals of the competition “Robotics for Bulgaria”

5. Conclusion

In the role of mentors of the student teams, our observations on the work of the student teams during the preparation and during the presentation of the students at the contests are:

- The pupils need to learn on their own and in a team, to overcome difficulties on their own and in a team, and to be satisfied, they need to be convinced that the team has succeeded, not that an adult has helped them. We were with them all the time, recording almost every experience and we had a lot of photos collected. We only intervened when they were

experiencing negative emotions (to calm them down) or when we had to organize ourselves to decide something or when they asked us for our opinion.

- The participation of teachers as mentors of teams is something completely new, unexpected and different. In this sense, it is a kind of challenge, because we had to act as a back of the team, to motivate and unite the students, to unite different personalities, characters, leaders and performers. So we found ourselves in a role that put us outside the comfort zone, which is quite challenging.

- The field of action of the teams was especially interesting. For all of us, this is a completely new role, a new mission. Because we had to communicate with each other in another environment, to learn together, to explore the unknown, to try different approaches, to have fun at the same time and most importantly - to truly enjoy with pure childish enthusiasm every successful mission and successfully completed task This affected very well everyone in the teams.

- Regarding "Robotics for Bulgaria": For us, all this is like a paradox - it is these experiences that detach us from the dynamics of the hectic and monotonous daily life, filled with a lot of stress and tension, which brings us closer to ourselves purely human. This convinced us once again that one of the missions of robots has been to always bring us back to our human face in this technology, electronic age. So each of us, along with everything we learn about robots and missions, can discover something new about ourselves. We have seen that here, as in any other field, the beginning is not easy, the road is long, and success requires a lot of effort, time and full dedication to the cause.

- Regarding the International Competition for students "Karst - the Last White Spot on the planet Earth": Preparation for participation in the competition changes the stereotype created by the traditional way of teaching – “through karst the real world begins to be perceived by students as cyclical, made up of systems in which man and his society are directly involved”. [8] Humans are an integral part of the natural karst systems in which they live, so they are obliged to comply with its laws of development. Through the competition the students as a future citizens learn to draw conclusions about the effective use of karst resources. The deeper they get into the essence of (karst) systems, the more effectively, they will be able to analyze and evaluate the real and potential transformations in the globally changing modern world.

- Both students and mentors were absolute outsiders at the beginning. Nevertheless, in the finals, apart from the satisfaction of the wonderful emotions and experiences, both teams proved to be extremely successful, because the most important are team action, team integrated knowledge and thinking, systematic approach and consideration and perception of things in all their interconnectedness. And last but not least: by competitive element and the integration of knowledge and skills are acquired qualitatively new knowledge and skills.

References

1. Vasileva, S., V. Egova, Interim Technical Report, Training program for mentors „Robotics for Bulgaria“, 2019.
2. International Competition for students "Karst - the Last White Spot on the planet Earth", 2019, (http://prokarstterra.bas.bg/sci-competition/)
3. International Karst Competition for Students "The Last White Spot" on Planet Earth, 2019, (http://prokarstterra.bas.bg/sci-competition/about.html)
4. Vasileva, S., A. Kondov, Some Ideas about Student Projects in the Subjects Informatics and Information Technologies in Secondary School, 8th International Conference On Application Of Information And Communication Technology And Statistics In Economy And Education (ICAICTSEE – 2018), October 18-20TH, 2018, UNWE, Sofia, Bulgaria 199-207, 2020.
5. Robotics for Bulgaria, Issue 2018/2019, 2018.

6. Program for training mentors, Guidelines Robotics for Bulgaria, 2019.
7. Stefanov, P., D. Michova, M. Silyanova, Geography, Textbook for 8th grade, K&M, Sofia, 1995. (http://prokarstterra.bas.bg/sci-competition/bg/handbook.html)
8. Stefanov, P., D. Stefanova, 2020, The First International Competition for Students "Karst - the Last „White Spot" on the Planet Earth" – Results and Perspectives, (http://sociobrains.com/MANUAL_DIR/SocioBrains/Issue%2066,%20February%202020/Natural%20Geography/11_%20Petar%20Stefanov,%20Dilyana%20Stefanova.pdf)
9. How to program EV3, 2018, (https://elearning.legoeducation.com/courses/how-to-program).
10. Electromagnetic boots, (https://www.instructables.com/id/Electromagnetic-Boots-and-Gloves, https://www.youtube.com/watch?v=uRKHCRR9-C4)
11. Workshop for Civic Initiatives Foundation, 2019, (https://frgi.bg/bg/articles/kakvo-novo/programa-robotika-za-balgariya-tarsi-motivirani-uchiteli)
12. Nearly 180 students competed in the fourth robotics festival, 2019, (https://frgi.bg/bg/articles/kakvo-novo/blizo-180-uchenitsi-se-sastezavaha-v-chetvartiya-festival-po-robotika)