



INTERNATIONAL CONFERENCE INFORMATION AND COMMUNICATION TECHNOLOGIES IN BUSINESS AND EDUCATION



CONFERENCE PROCEEDINGS



Publishing house "Science and economics"
University of Economics - Varna

International Conference
INFORMATION AND COMMUNICATION
TECHNOLOGIES IN
BUSINESS AND EDUCATION

Conference proceedings

**INFORMATION AND COMMUNICATION
TECHNOLOGIES IN
BUSINESS AND EDUCATION**

**Proceedings of the International Conference
dedicated to the
50th anniversary
of the Department of Informatics**

2019

University publishing house “Science and economics”
University of Economics – Varna

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Permission for use must always be obtained from University publishing house “Science and economics”. Please contact izdatel@ue-varna.bg

Published reports have not been edited or adjusted. The authors are solely responsible for the content, originality and errors of their own fault.

With the support of ADASTRA, 411 Marketing Business Services.

© University publishing house “Science and economics”,
University of Economics – Varna, 2019.

ISBN 978-954-21-1004-0

PROGRAMME COMMITTEE

Abdel-Badeeh M. Salem, Egypt

Aijaz A. Shaikh, Finland

Aleksandar Dimov, Bulgaria

Alexander Feldman, USA

Alexander Sergeev, Russia

Antoanela Naaji, Romania

Avram Eskenazi, Bulgaria

Dariusz Mrozek, Poland

Dimitrina Polimirova, Bulgaria

Dmitriy Berg, Russia

Elena Nechita, Romania

Emil Denchev, Bulgaria

Georgi Dimitrov, Bulgaria

Ivan Popchev, Bulgaria

Jeanne Schreurs, Belgium

Julian Vasilev, Bulgaria

Kameliya Stefanova, Bulgaria

Krasimir Shishmanov, Bulgaria

Maciej Czaplewski, Poland

Maciej Pondel, Poland

Maksim Medvedev, Russia

Maria Hristova, Bulgaria

Marian Cristescu, Romania

Marina Medvedeva, Russia

Michael Voskoglou, Greece

Nadezhda Filipova, Bulgaria

Nikola Nikolov, Ireland

Niv Ahituv, Israel

Pavel Petrov, Bulgaria

Rami Malkawi, Jordan

Sara Paiva, Portugal

Silvia Parusheva, Bulgaria

Snezhana Sulova, Bulgaria

Todorka Atanasova, Bulgaria

Valentin Kisimov, Bulgaria
Valery Trofimov, Russia
Veselin Popov, Bulgaria
Violeta Kraeva, Bulgaria
Vladimir Sulov, Bulgaria
Vladimir Zanev, USA
Zdzislaw Polkowski, Poland

ORGANIZING COMMITTEE

Chairman: Prof. Vladimir Sulov, PhD

Members:

Prof. Avram Eskenazi, PhD
Prof. Julian Vasilev, PhD
Assoc. Prof. Todorka Atanasova, PhD
Assoc. Prof. Silvia Parusheva, PhD
Assoc. Prof. Snezhana Sulova, PhD
Assoc. Prof. Pavel Petrov, PhD

Secretary: Tijen Talib

FOREWORD

These proceedings contain the papers of the International Conference “Information and Communication Technologies in Business and Education” which took place at the University of Economics – Varna, Bulgaria, 18 October 2019.

The international scientific conference is dedicated to the **50th anniversary of the Department of Informatics at the University of Economics – Varna**. The conference is also dedicated to the 100th anniversary of the University. The included papers describe recent scientific and practical developments in the field of information and communication technologies, information systems, and their applications in business and education.

The papers in the Proceedings are peer reviewed and are checked for plagiarism.

C O N T E N T S

1. Kamelia Stefanova, Dorina Kabakchieva Challenges and Perspectives of Digital Transformation	13
2. Julian Vasilev, Marian Cristescu Approaches for information sharing from manufacturing logistics with downstream supply chain partners	24
3. Abdel-Badeeh M. Salem Computational Intelligence in Smart Education and Learning	30
4. Marian Niedźwiedziński Proposal of the Multi-aspect Analysis of ICT's Needs in Management	41
5. Krasimir Shishmanov Digital Transformation in Bulgaria	52
6. Tatyana Makarchuk, Valery Trofimov, Svetlana Demchenko Modeling the life cycle of e-learning course using Moodle Cloud LMS	62
7. Vanya Lazarova Exploring the functionality of Bulgaria eGovernment site	72
8. Veselin Popov, Petya Emilova Big data and information security	83
9. Luben Boyanov Approaches for enhancing digitalization and digital transformation in supply chain management	91

10. Monika Tsaneva	
Challenges of GDPR compliance in consumer financing companies	103
11. Todorka Atanasova	
Main factors influencing digitization in construction companies	116
12. Silvia Parusheva	
Digitalization and digital transformation in construction – benefits and challenges	126
13. Snezhana Sulova	
The Usage of Data Lake for Business Intelligence Data Analysis	135
14. Pavel Petrov	
Web security technologies used in banks of Estonia, Latvia and Lithuania	145
15. Tanka Milkova	
ABC analysis of inventory in MS Excel	155
16. Ivan Kuyumdzhiev, Radka Nacheva	
Choosing Storage Devices with Consideration of Backup and Restore Performance in MS SQL SERVER	166
17. Yanka Aleksandrova	
Predicting Students Performance in Moodle Platforms Using Machine Learning Algorithms	177
18. Latinka Todoranova, Bonimir Penchev	
M-learning Applications	188
19. Boris Bankov	
The Impact of Social Media on Video Game Communities and the Gaming Industry	198

20. Velina Koleva, Svetoslav Ivanov	
Social media and the recruitment of IT professionals in Bulgarian IT companies	209
21. Deyan Mihaylov	
Modeling and Simulation of Some Functions of Two Independent Random Variables	225
22. Oumar Sy	
Using Association Rules as Semantics for Domain Ontologies	233
23. Ina Stanoeva	
Social media possibilities for improving arts marketing performance	247
24. Iskren Tairov, Vladislav Vasilev	
Perspectives on mobile devices adoption in healthcare sector	257
25. Angelin Lalev	
Deep Neural Networks for Detection of Credit Card Fraud	263
26. Miglena Stoyanova	
Priorities for Digital Transformation in Property Management	275
27. Mihail Radev	
Organizational Variants of IT Department at the University	284
28. Mariya Armyanova	
IoT problems and design patterns which are appropriate to solve them	291
29. Yavor Christov	
Corporate Big Data and Hybrid Integration Platforms	306

30. Ivan Belev	
Making financial data readable using Inline XBRL	314
31. Ivan Belev	
eXtensible Business Reporting Language (XBRL) – a technical overview	322
32. Anna Sobczak	
Entrepreneurship in the education of university students	331
33. Mihaela Markova	
E-commerce in the process of digital transformation: preparedness, requirements and technological tendencies	343
34. Petya Petrova	
Basic Soft Computing Methods In User Profile Modeling	353
35. Petar Dimitrov	
Methods for storing authentication data – a historical review	369
36. Svetoslav Ivanov	
Development stages of starting software company, problems and approaches for software development	382
37. Antonio Hadzhikolev	
The role of the smart city concept in the process of urban transformation	397
38. Oumar Sy	
Ontology-based materialized views for reverse engineering: Toward a set of conceptual guidelines	410

CHALLENGES AND PERSPECTIVES OF DIGITAL TRANSFORMATION

Kamelia Stefanova¹, Dorina Kabakchieva²

***¹ University of National and World Economy/ITC Department,
Sofia, Bulgaria, kstefanova@unwe.bg***

***² University of National and World Economy/ITC Department,
Sofia, Bulgaria, dkabakchieva@unwe.bg***

Abstract

Internet Economy, New Economy, Web Economy, and Connected Economy are all new ICT terms referring to the digitalization and digital transformation of companies/organizations, and defining the process of creating business value through digitally organized links between people, machines and companies, creating new related business models, products and services. Digital transformation involves also using Internet-of-Things (IoT), Artificial Intelligence (AI) and Business Analytics, among other advanced information technologies. The main purpose of this paper is to present the challenges and perspectives of digital transformation, including scope, strategy and enterprise architecture, and to point out its importance for business innovation and competitiveness.

Keywords: *digitalization of economy, digital transformation, enterprise architecture for digital transformation.*

INTRODUCTION

“Digitalization” is a complex ICT term which is currently undergoing its development. If “business” refers to the activities of a company, then digitalization of the business is a process in which ICT technologies reduce the costs of the company operational activities through computer collaboration of business processes, storing, sharing and analysing data, changing behaviour of customers and suppliers, and the organization of entire industrial connectivity of this business. Digitalization of individual business processes leads to the digitalization of economy, which is sometimes referred to as Internet Economy, New Economy, Web Economy, or Connected Economy. These terms define the process of creating business value through digitally organized links between people, machines and companies, creating new related business models, products and services.

The digitalization of economy also includes embedding sensors in equipment for more efficient management, creation of processes based on services and collaborative design for increased quality in the shortest time.

The main purpose of this paper is to present the challenges and perspectives of digital transformation, and its importance for business innovation and competitiveness. The digital transformation scope is described in the first paper section, and then the digital transformation strategy is explained in the second section. Some of the most relevant good practices concerning the enterprise architecture for digital transformation are discussed in the third section.

1. DIDITAL TRANSFORMATION SCOPE

Digital transformation process becomes a focus for exploration of many academic, research and applied works. The scope of such investigations is expanding dramatically, and the number of areas encompassed is growing fast. For the success of studies in the field of digital transformation, it is very important they to be positioned correctly in order characters of concrete directions to be defined. Different aspects of digital transformation manifestation could be found in the literature (Mariam. Ismail, Mohamed Khater, 2017). The most relevant perspectives to this paper are presented on Figure 1.

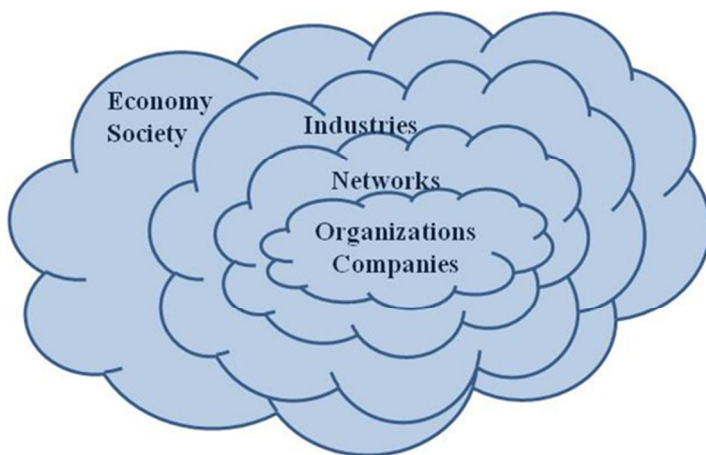


Figure 1. Aspects of Digital Transformation Process

The reason for figuring out these aspects of Digital Transformation (DT) Process is to understand how critical is to consider each of them to be developed in accordance with the trends and requirements of progression of the others. It is impossible to identify the beginning point of the DT Process. On one hand, each aspect is influencing the others, and on the other hand, it is dependent on them.

- The Modern development level of economy and society is totally influenced by globalization. Globalization process itself is passing through different stages. Globalization 3.0 is reported to be originating from 1989 (Peter Vanham, 2019). Today we are in Globalization 4.0 that has been started since 2008. The economy nowadays is characterized by very dynamic and deep changes, extremely high competition, strong customization, and the main cornerstones for succeeding and even surviving in it require implementation of embedded technologies, innovation in products and services, advanced knowledge and open mindsets. The new economy and society models are based on a different paradigm for connected/circular economy where streams of resources - materials, labor, energy, new information interact within an intelligent environment that creates restorative, regenerative and more productive economy and transparent society.

- The Industry aspect is characterized with very deep changes that influence the way industries operate and develop their production processes and services. New concepts for industries performance have been introduced. Popularity of “Industry 4.0” is not any more only on political and strategy papers. Digital transformation has changed the known boundaries within whole industrial sectors and between them. The nature of productions and factories became totally different with new business processes models based on introduction of innovative technologies. New terms like “smart factories”, “intelligent manufacturing” have appeared in order to describe the trends of penetrating digital transformation changes within industries. The production life cycles design requires new approaches that to support optimization models, customer orientation and competitors’ analysis.

- Networks have become a very critical part of the digital transformation process. Networks themselves are going through dynamic

changes encompassing new technologies and services for enlarging single entities performance. Today extended enterprises and digital ecosystems development are the new phenomena that break all the traditional models for linking businesses and people. And not only, new communication and execution models are getting introduced in order to create innovative models for doing business, exchanging information and connecting different participants within the global environment.

- Single businesses – companies and organizations are facing the most critical challenges and need for transformation they even experienced. Companies from all sizes are equally influenced by digital transformation processes. The directions for changes cover multiple aspects - digital technologies: Cloud computing, Big Data, Internet of Things; innovative business models; service and customer-oriented processes; global dynamic competition and increased risks. It is reported that about 90% of businesses in Europe and in the USA are supposed to strategically change their businesses based on implementation of digital technologies (Hess et al. 2016).

Following the above considerations, it is important to focus on the definition of digital economy. Different formulations could be found in scientific and research papers. The one that covers the broad dimensions of this phenomenon (Deloitte, 20019) is stating that new economic activities are appearing, and their results are based on billions of everyday online transactions connecting businesses, processes, people, devices, and data. It means that the heart of digital economy is the connectivity which ensures the development of connected economy where each entity could be linked with all the others in the global world.

2. DIGITAL TRANSFORMATION STRATEGY

The challenges for implementing innovations make businesses define their own strategies for development. Digital transformation is a very complex process that requires a clear vision and elaborated strategy for guiding the successful steps. The role of such a strategy is to plan the activities, to calculate the budget, to ensure objectives and to minimize the risks. Unfortunately, some surveys report that many businesses start with technological changes introduction without building a holistic plan. Forrester says that only 27% of businesses have a coherent digital strategy.

Digital Transformation Strategies (DTS) could cover different structures and different number of steps to be followed. Some publications are focusing on organizational, management, cultural aspects; others are pointing out the importance of new technologies and relevant skills for working with them; others are stressing on the changes related to customers and markets (Vitaliy Zhovtyuk, 2018; Carla Rudder, 2019). We consider that all these aspects should be covered by the DTS and suggest the following steps presented on Figure 2.

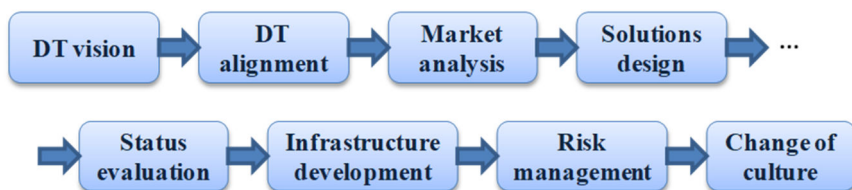


Figure 2. Steps of Digital Transformation Process

Step 1: Creation of a Digital Transformation clear vision

The main corner stone that would make this step successful is when figuring out the vision for DT is not to focus on the existing problems and how to be solved with innovative technologies, but to define new objectives that will better position the business in the future. On the other hand, these objectives should not be abstract ones. This all means that the strategic vision should address short term goals, based on the current available resources, but considering how to increase the competitive advantage, which are the critical gaps within the existing structure and what to be further steps for improvement and optimization.

Step 2: Digital Transformation alignment with business goals

Once the clear objectives are settled another critical step requires a lot of efforts. These objectives should be aligned within the whole organizational structure from top down. The new digital strategy should be well communicated outlining concrete outcomes coming from new digital processes. A mistake linking digital transformation only with the implementation of technology should be avoided. All aspects of business

functions need to be described with functional changes, with new budget, clear mission and plans for future.

Step 3: Markets, competitors and customers analysis

This step is focusing on the need for development of relevant topical strategy. It is impossible to start any transformation activity without deep analysis of what are the competitors advantages, the current customer needs, the industry changes, the market orientations, the political priorities, and technology innovative solutions.

Step 4: New solutions design

Technically and functionally all the business solutions will be designed and developed within this step. The critical point here is not to miss the user and employee digital expertise aspect. Clear understanding of what is the target customers' digital experience and what is the employees' digital knowledge and skills are the main pillars for successful design. This step requires an intelligent attitude to the whole digital environment design process, considering customer preferences how to complete the services easy and fast, how to have an access anytime from anywhere and employees needs how to fulfill more effectively and efficiently their tasks. The focus should not be only on innovations but how these innovations could provide better experience.

Step 5: Analysis and evaluation of the current situation

The four first steps tasks are devoted to the definition of what and how businesses would like to achieve with the digital transformation. Such objectives could not be reached if the current situation is not analyzed and evaluated according to the stated objectives for digital transformation. As management theory is saying – to begin a journey we need to know where we are, to identify the main gaps to get to the objectives, which are the most critical and how to be fulfilled. This means that the activities here should involve – analysis and assessment of current functionalities, processes to be automated, technologies to be updated, digital tools to be developed and changed, priorities for investment and capacity to be built.

Step 6: Infrastructure implementation

At this step all the questions related to the digital transformation of the infrastructure must be settled and solved. This step will close when all hardware and software components are implemented, when all applications, tools and data are integrated and tested successfully. Critical requirement here is to appoint and work with digitally competent professionals which further training and qualification will be an important task to ensure future business digital development.

Step 7: Management of the risk

Digital transformation introduces a lot of risks related to the new disruptive business models implementation, to the fast and deep changes of existing processes and solutions, to the scalable technological improvement, etc. These characteristics of the digital changes require specific actions related to the design of risk management and controls.

Step 8: Cultural change

We should not forget that by its incremental substance, the digital transformation process is people change. To ensure successful results, business management should consider the cultural aspects of transformation, to support smooth embracement of new ways of performing, to identify the proper communication of changes and successful strategy for training.

Business environment today is characterized with very high competition and the challenges for survival put the critical need for digital transformation that surely require implementation of a strategy to establish objectives encompassing new opportunities for business, new products and serviced for customers introducing new channels and platforms.

3. ENTERPRISE ARCHITECTURE FOR DIGITAL TRANSFORMATION

Digital transformation requires a new enterprise architecture. In order to experience digital transformation, traditional companies should continually update their enterprise architecture and eventually use different approaches for designing and managing it. McKinsey experts call

this model “perpetual evolution” (Bossert and Laartz, 2017), requiring from managers to comprehensively view their digital technologies and capabilities, and at the same time to manage them in a new manner, with less interdependencies and an emphasis on speed.

Bossert and Laartz (2017) define six main differences between traditional enterprise architecture and the one designed and managed for “perpetual evolution” (Figure 3). In the traditional model, business processes are centred around products and services, and a single operating model is used. The “perpetual evolution” model implements customer-centred approach and multiple operating models. As far as the business applications are concerned, the traditional model relies on interdependencies while the “perpetual evolution” model emphasises on the need for decoupling. Centralization is typical for the IT integration platform and the infrastructure services in the traditional enterprise architecture. Contrary to that, in the “perpetual evolution” enterprise architecture, lightweight connections are used in the IT integration platform and the software developers and IT operations jointly build new products and features. There is also a difference in the management of information and communication technologies – in the traditional model they are managed as precious asset, while in the “perpetual evolution” model – as commodity.

Digital transformation requires a different model for managing enterprise architecture.

Traditional model	Elements of enterprise architecture	Perpetual evolution model
Focus on product- or service-centered processes	Business operations	Focus on customer-centric journeys
Reliance on one operating model	Business capabilities	Use of multiple operating models (working at 2 speeds)
Emphasis on interdependency	Business applications	Emphasis on decoupling applications
Use of heavyweight bus ¹	IT integration platform	Use of lightweight connections
Software development managed centrally	Infrastructure services	Software developers and IT operations jointly build new products and features (DevOps)
Managed as precious asset	Information and communication technology	Managed as commodity
		Technology stack

¹Connection layer that contains most of the business logic (or rules of computing).

McKinsey&Company

Figure 3. McKinsey comparison of traditional and “perpetual evolution” enterprise architecture (Source: Bossert and Laartz, 2017)

Lambert and Rosen (2018) also discuss the strong dependence of successful IT delivery on business strategy and the necessity of a new, agile and innovative-driven architecture for digital transformation. They emphasize on the importance of providing customer-driven processes and digital capabilities. Five key areas are identified by IDC experts in which companies should strengthen their capabilities in order to achieve successful digital transformation, including “leadership, omni-experience, information, operating model and worksource” (Lambert & Rosen, 2018). Leadership and worksource are the main factors of transformation in every organization, while omni-experience, information and operating model are the enablers of new digital products, services and engagement across the ecosystem.

A recent survey (Blumberg, Bossert & Sokalski, 2018) points out some important good practices for facilitating digital transformation, contributing to better management of the increasing technological complexity resulting from the numerous digital applications. The first advice concerns involving high-level executives in discussions of business strategy, especially related to technologies, in order to better understand business needs and invest in continuously improving enterprise architecture. Emphasizing on long-term planning is another peculiar characteristic feature of digital leaders. Putting extra efforts in strategic planning usually leads to higher added value, more sustainable business solutions and wider recognition within the company. Focusing on business outcomes is of utmost importance when striving for digital transformation as it usually requires changing business models by implementing advanced technologies. Therefore, good collaboration between IT experts and businesspeople should be established in order to achieve higher benefits. Better interaction between business and IT could be achieved by evaluating the IT function based on the business capabilities deployed rather than the number of implemented technology applications. Another important peculiarity is related to developing and retaining highly skilled and motivated professionals. Talent in digital leaders is very often attracted by offering exiting work, good opportunities to learn and grow and well-structured career paths.

CONCLUSION

When defining the scope of digital transformation, it is important to consider the different aspects of the digital transformation process, including organization/company, networks, industries and society/economy, all of them influencing and depending on each other. Facing the challenges for implementing innovations in the digitalizing business environment, companies need to define new business strategies in order to succeed. Digital transformation is a very complex process that requires a clear vision and elaborated strategy for guiding the successful steps. In order to experience digital transformation, businesses also need to implement a different model for managing enterprise architecture.

REFERENCES

1. BLUMBERG, S., BOSSERT, O., and SOKALSKI, I. (2018). *Five enterprise-architecture practices that add value to digital transformations*. McKinsey&Company website article. Available from: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/five-enterprise-architecture-practices-that-add-value-to-digital-transformations> [Accessed 05/09/2019]
2. BOSSERT, O., LAARTZ, J. (2017). Perpetual evolution the management approach required for digital transformation. McKinsey&Company website article. Available from: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/perpetual-evolution-the-management-approach-required-for-digital-transformation> [Accessed 05/09/2019]
3. DELOITTE (2019). What is digital economy? Available from: <https://www2.deloitte.com/mt/en/pages/technology/articles/mt-what-is-digital-economy.html> [Accessed 23/08/2019]
4. ISMAIL, M, KHATER, M., ZAKI, M. (2017). Digital Business Transformation and Strategy. Cambridge Service Alliance 2017. Available from: <https://pdfs.semanticscholar.org/35b2/bd95aeb8e6ca0ea0652d0d7fc697eda3362a.pdf> [Accessed 10/08/2019]
5. LAMBERT, D., ROSEN, M. (2018). From Business Strategies to Successful IT Delivery. CIO, IDG Communications, Inc. Available from: <https://www.cio.com/article/3246686/new-realities-require-new-architecture.html> [Accessed 03/09/2019]
6. RUDDER, C. (2019). Digital transformation strategy: 7 key pieces. Available from: <https://enterpriseproject.com/article/2019/8/digital-transformation-strategy-7-pieces>, [Accessed 01/09/2019]
7. VANHAM, P. (2019). A brief history of globalization. World Economic Forum, Available from: <https://www.weforum.org/agenda/2019/01/how-globalization-4-0-fits-into-the-history-of-globalization/> [Accessed 18/08/2019]
8. ZHOVITYUK, V. (2018). How to Develop a Digital Transformation Strategy: Step-by-Step. Medium Business. Available from: <https://medium.com/@VitaliyZhovtyuk/how-to-develop-a-digital-transformation-strategy-step-by-step-guide-f8e4bd5719b6> [Accessed 23/08/2019]

APPROACHES FOR INFORMATION SHARING FROM MANUFACTURING LOGISTICS WITH DOWNSTREAM SUPPLY CHAIN PARTNERS

Julian Vasilev¹, Marian Cristescu²

¹ *University of Economics Varna/Department of Informatics,
Varna, Bulgaria, vasilev@ue-varna.bg*

² *“Lucian Blag” University of Sibiu/Faculty of Economic Sciences,
Sibiu, Romania, marian.cristescu@ulbsibiu.ro*

Abstract

In some cases, manufacturing enterprises cannot accept orders from corporate customers because their manufacturing capacity is busy. To solve this problem this article proposes some approaches for information sharing with downstream partners of supply chains. The main approach is sending an XML file with free capacity by days from a manufacturing enterprise to its customers. In this case customers who send orders to the manufacturer are sure that their orders will be accepted and fulfilled. The practical implication of the proposed method is for manufacturing enterprises who have a lot of orders and sometimes reject some of the orders due to the lack of free manufacturing capacity.

Keywords: *ERP systems, XML file.*

INTRODUCTION

In terms of logistical information, the spark of information sharing does not come from the entity that provides the information but from the enterprise that needs logistical information. For example, employees of a supermarket may be interested and find producers of onduline, to test the products they offer and to stop producers who produce quality onduline. Before the onset of the summer season, the employees of the supermarket can send requests to the manufacturers of onduline. In the best case, manufacturers execute orders (if the production capacity is free). But, it is possible that producers do not have the production capacity to allow a large batch to be produced in short terms, or to accept orders from other supermarkets and refuse production. All mentioned features indicate that the planning of production in the short term (in order to cover critical points in demand) is almost impossible.

1. SHARING INFORMATION THROUGH XML FILES

In order to assist joint logistics, the supermarket may require production capacity information from the production plant (number of sheets onduline which can be produced in one day) and a calendar schedule for the planned production by day, according to the orders made by customers (Fig. 1).

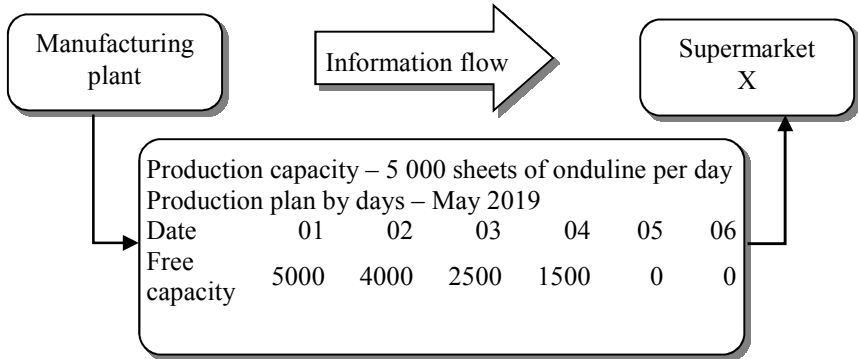


Figure 1. Information flow between a production enterprise and a corporate client

In the presence of such information, shared by the manufacturing company with the supermarket (at the request of the supermarket), the supermarket can decide whether to send an order to the manufacturer or not to send an order at all, to wait for an answer that will certainly be negative (due to the fact that the manufacturing facilities are insufficient or occupied). Fig. 1 shows that on 05.05.2019 and 06.05.2019 the manufacturing enterprise cannot accept requests, since its manufacturing facilities are busy.

One of the possible approaches to communication is by providing an XML file from the dealer's manufacturer (Fig. 2).

```

<?Xml Version= "1.0" Encoding=" Cyrilic -1251 "? >
- <Enterprise Capacity="5000">
</Enterprise>
- <Plan Item=green onduline sheet">
<day1>01.05.2019</day1>
<free1>5000</free1>
<day2> 02.05.2019</day2>
<free2> 4000 </free2>
<day3> 03.05.2019</day3>
<free3> 2500 </free3>
<day4> 04.05.2019</day4>
<free4> 1500 </free4>
<day5> 05.05.2019</day5>
<free5> 0 </free5>
<day6> 06.06.2019</day6>
<free6> 0 </free6>
</Plan>
</Xml>

```

Figure 2. XML file structure that is being prepared by the manufacturer and sent to the merchant

The XML file structure is presented in fig. 2. The XML file can be sent daily by the manufacturer to the supermarket (the merchant). A possible approach for information sharing is presented. Other known approaches to implement the communication are: (1) using the SOAP Protocol, (2) web services (Stoyanova, 2015a) and (3) communication by port.

The usefulness of sharing information is for both partners in supply chain. The manufacturer receives orders that are fulfilled. The trader sends an order which is accepted by the manufacturer (it is not rejected due to reserved manufacturing capacity for other orders). The salesperson waits for an answer whether the order will be executed. The probability of being executed is greater because the trader has up-to-date information on production capacity and free production capacity. The investments for the manufacturer and the trader are in terms of expansion

of the used software (the ERP system of the manufacturer and the ERP system of the customer). The functionality of the software application is extended, data could be extracted, transformed and loaded to databases in use (Kuyumdzhiev and Bankov, 2017). According to some authors, such investments are an important factor for the application of the implemented software (Stoyanova, 2015b). There are benefits for both customers and suppliers of the integrated relationship between them, namely: cost reductions, improving the effectiveness of actions, strengthening confidence among countries, adding value (Milusheva, 2013). The process of generating an XML file from an ERP system requires minimal effort of system designers and programmers.

The process of reading XML file from an ERP system takes more resources (time and people work) because an input filter is required (an analyzer for the correctness of the entered data in the XML file). After the initial validation of the XML file is checked, the data is used by the merchant. Again, we have several options – (1) Manually examining the file and making a decision, (2) automated input of the data from the XML file into the trader's information system and the removal of advice and (3) automatic transfer of data from the XML file (between TRADER) without human participation. Obviously, the quickest approach to data exchange is the following – automatically generating an XML document, automatically sending the XML document to the merchant (or to multiple vendors), and automatically accepting data from the XML document.

The sharing of order information (between the ordering and the organisation accepting orders) is a prerequisite for the execution of transactions between the two parties. As noted, in the e-commerce systems, where core business activities are carried out through dynamic online systems (Sulova, 2016), an order from a customer to a supplier is entered the provider's webpage. The supplier can provide the order to the customer as an electronic document.

Based on the research we offer in the area of e-logistics to share aggregated information about free production capacity by day. It is calculated by the formula:

$$\text{FPCD} = \text{TPC} - \text{COO}$$

FPCD – Free production capacity by day

TPC – Total production capacity by days

COO – Capacity occupied by orders

In addition, business processes and software systems in organizations in terms of participation, motivation and engagement of their users are improved (Stoyanova, 2018). Under the current conditions, when firms compete, betting on their supply chains, the task of managing managers is to strive for a complete synchronization in the work of customers and suppliers. Integration of business processes in the chain is a key factor in the development of successful supply chains (Milusheva, 2013). The proposed method is also important for the company's marketing because it helps the management of customer relationships in order to keep current customers and to gain new ones with low-cost and on time deliveries (Sulova, 2018). Under these conditions, one of the main possible factors for competitiveness not only for companies but also for their supply chains is the relationship with the company's suppliers (Milusheva, 2016).

CONCLUSION

By applying the proposed innovative approach, additional information is provided in the supply chain, which allows better logistic and supply chain participation. Traditionally, the supply chain participants issue invoices. The issued invoices generate transactions both with the supplier of the product/service/and the customer. In order to fulfil off-line data transmission, most often the invoices are structured as an electronic document in an approved standard and then they are sent to the client (electronically).

REFERENCES

1. KUYUMMDZHIEV, I., BANKOV, B. (2017) Migration of unstructured data into a relational scheme: problems, solutions, algorithm. Economics in a changing world – national, regional and global dimensions, Varna: Science and Economics, 1, 2017, 551-556.
2. MILUSHEVA, P. (2013) Relationships with Suppliers, Edition of the Conference of the University of Forestry Sofia, issue 1, p.49-53.

3. MILUSHEVA, P. (2016) Aspects in the Relationships of Companies with Suppliers, Economics and Computer Science, Volume 2, Issue1, pp. 6-9.
4. STOYANOVA, M. (2015) An Architecture Of A Gamification System. Economics and computer science, Varna: Knowledge and business, 1(2), 18-33. Available at: <http://eknigibg.net/spisanie-br2-2015.pdf>
5. STOYANOVA, M. (2015) Gamification process in information systems, Special Issue Proceedings of ETAEc 2015 Conference, Scientific Bulletin – Economic Sciences, 14, pp. 174–180.
6. STOYANOVA, M. (2018) Application of the Gamification Concept in Project Management Software Systems. Varna: Monographic Library “Knowledge and Business”. Available at: http://eknigibg.net/monogr/2018.10_Monogr_gamification.pdf
7. SULOVA, S. (2016) An Approach for Automatic Analysis of Online Store Product and Services Reviews. Izvestia Journal of Varna University of Economics, 60(4), pp. 447-490.
8. SULOVA, S. (2018) Integration of structured and unstructured data in the analysis of e-commerce customers. SGEM 2018: 18 international multidisciplinary scientific geoconference, Albena: STEF93 Technology Ltd., 18, pp. 499-506.

COMPUTATIONAL INTELLIGENCE IN SMART EDUCATION AND LEARNING

Abdel-Badeeh M. Salem¹

¹ *Ain Shams University/Faculty of Computer and Information sciences,
Cairo, Egypt, abmsalem@yahoo.com, absalem@cis.asu.edu.eg*

Abstract

Computational intelligence (CI) techniques are intelligent paradigms based on artificial intelligence (AI) concepts, approaches and theories. CI deals with the development of intelligent software that improves automatically through experience. Researchers have been used the CI algorithms in developing a new generation of smart tutoring and learning systems. This paper discusses the CI approaches for developing the digital e-Learning and intelligent tutoring systems. In addition, the paper presents some examples of the developed systems by the author and his colleagues during the last ten years.

Keywords: *Artificial intelligence, Computational intelligence, Machine learning, intelligent tutoring systems, Smart learning systems.*

INTRODUCTION

CI techniques aim to enable computers to learn from data and make improvements without any dependence on commands in a program. This learning could eventually help computers in building smart models such as those used in the prediction tasks. On the other side, artificial intelligence (AI) is an interdisciplinary science and based on many disciplines such as: computer science, mathematics, philosophy, psychology, life sciences, linguistics, library sciences and engineering. The goal of AI is to develop intelligent software models of the human behaviour, i.e. the abilities of thinking, hearing, talking, walking and also feeling. The AI field covers many research areas, e.g. knowledge engineering, action and perception, reasoning methodologies, cognitive computing and modeling, connectionist models, constraint satisfaction, machine learning, natural language processing, and AI planning and scheduling. The main AI technologies include: vision systems, robotics, intelligent games, expert systems, natural language (NL) understanding, intelligent multimedia systems, data mining and knowledge discovery systems and intelligent tutoring and learning systems.

On the other side, development of intelligent tutoring and learning systems are the very important industry in the area of AI in education. The main three components of efficient and robust intelligent tutoring and learning systems in any domain are the “knowledge base”, “inference engine” and NL interface. Concerning the knowledge base there are many knowledge representation and management techniques, e.g.; lists, trees, semantic networks, frames, scripts, production rules, cases, and ontologies. The key to the success of such systems is the selection of the appropriate technique that best fits the domain knowledge and the problem to be solved (Greer, 1995; Mazza and Milani, 2005). That choice depends on the experience of the knowledge engineer. Regarding the inference engine, there are many methodologies and approaches of reasoning e.g.; automated reasoning, case-based reasoning, common sense reasoning, fuzzy reasoning, geometric reasoning, non-monotonic reasoning, model-based reasoning, probabilistic reasoning, causal reasoning, qualitative reasoning, spatial reasoning and temporal reasoning. In fact, these methodologies receive increasing attention within the AI in education and learning communities (Salem, 2011; Salem, 2008; Salem, 2007; Salem and Parusheva, 2018b).

1. INTELLIGENT TUTORING SYSTEMS (ITS) AND SMART LEARNING SYSTEMS (SLS)

ITS is a knowledge-based software that act as an intelligent tutor used in real teaching. ITS is also used in learning, and training situations (Greer, 1995). From the technical point of view, ITS is composed of the following software components: (a) expert model, (b) student mode, (c) instructional module, (d) interface and (e) knowledge acquisition module. On the other side, SLS are AI-based systems that imitates the human mind. The main characteristics of these systems are the ability of inference, reasoning, perception, learning, and knowledge-based systems. To a limited degree, AI permits SLS to accept knowledge from human input, then use that knowledge through simulated thought and reasoning processes to solve problems. Many types of ITS and SLS are in existence today and are applied to different domains and tasks, e.g., geology, biological sciences, medical sciences, health care, commerce, and education (Salem, 2011; Mazza and Milani, 2005). From the development and

computing aspects, intelligent software authoring tools and shells (SATS) allow a course instructor to easily enter domain and other knowledge without requiring computer programming skills. THE SATS automatically generates an ITS/IeLS focusing on the specified knowledge (Greer, 1995).

2. COMPUTATIONAL INTELLIGENCE FOR ITS AND SLS

A. Case Based Reasoning (CBR) Technique

CBR is an analogical reasoning method provides both a methodology for problem solving and a cognitive model of people (Kolonder, 1993). CBR means reasoning from experiences or "old cases" in an effort to solve problems, critique solutions, and explain anomalous situations. People tend to be comfortable using CBR methodology for decision making, in dynamically changing situations and other situations where much is unknown and when solutions are not clear. From knowledge engineering point of view, the "case" is a list of features that lead to a particular outcome. From the CI perspective, CBR refers to a number of intelligent algorithms smart processes that can be used to record and index cases and then search them to identify the ones that might be useful in solving new cases when they are presented. In addition, there are techniques that can be used to modify earlier cases to better match new cases and other techniques to synthesize new cases when they are needed. CBR has already been applied in several different applications in many domains, e.g. medicine, industry, law, banking (Abdrabou and Salem, 2010; Kolonder, 1993).

B. Ontological Engineering (OE) Approach

The term "ontology" is inherited from philosophy, in which it is a branch of metaphysics concerned with the nature of being. The main objective of using ontologies is to share knowledge between computers or computers and human. Most of the usages of ontologies in the field of computer science are related to knowledge-based systems and intelligent systems. These types of ontologies include a small number of concepts and their main objective is to facilitate reasoning. During the last decade, increasing attention has been focused on ontologies (Tankelevciene and

Damasevicius, 2009). At present, there are applications of ontologies with commercial, industrial, medical, business and research focus (Salem, 2010; Salem and Alfonse, 2007; Su and Ilebrekke, 2002).

C. Data Mining and Knowledge Discovery in Databases Approach

Data mining methodology aims to extract useful knowledge and discover some hidden patterns from huge number of databases which statistical approaches cannot discover. It is a multidisciplinary field of research includes: AI, CI, machine learning, data science, and statistics, Data mining techniques aim at providing intelligent computational methods for accumulating, changing and updating knowledge in intelligent systems, and in particular learning mechanisms that will help us to induce knowledge from information or data.

From the intelligent soft computing and data science, knowledge discovery in databases (KDD) involves the following three main processes; (a) using the database along with any required selection, preprocessing, sub-sampling, and transformations of it, (b) applying data mining methods (algorithms) to enumerate patterns from it, and (c) evaluating the products of data mining to identify the subset of the enumerated patterns deemed knowledge. Data mining is supported by a host that captures the character of data in several different ways, e.g. clustering, classification, link analysis, sequence analysis, regression models, summarization, text mining, sequential pattern mining, association rules mining. In addition, there are a lot of intelligent techniques to perform these tasks e.g. bio-inspired algorithms, decision trees, k-means. For more technical details and aspects, we refer to the books (Cios, Pedrycz and Swiniarski, 1998; Witten and Frank, 2005).

3. BENEFITS OF CI TECHNIQUES TO ELEARNING/TUTORING SYSTEMS

This section discusses the benefits of the previous techniques in e-learning and tutoring systems.

A. Benefits of CBR Approach to eLearning/tutoring systems

There are several benefits where students/learners should be able to perform better using CBR methodology, e.g.:

- Students/learners will be able to recognize more situations and the solutions that go with these cases include failure cases, students will be able to benefit from the failures of others.

- Students/learners will have access to obscure cases that they otherwise would not be able to make use of. These obscure cases can help with any of the tasks previously listed.

- During a training period CBR system provides the student with a model of the way decision making ought to be done.

- * For tasks where there is much to remember, CBR systems can augment the memories of even educators.

- * Both educators and students tend to focus on too few possibilities when reasoning analogically or to focus on the wrong cases.

B. Benefits of Ontologies to Intelligent Educational Systems

Ontologies usage in educational systems may be approached from various points of view: (a) as a common vocabulary for multi-agent system, (b) as a chain between heterogeneous educational systems, (c) ontologies for pedagogical resources sharing or for sharing data and (d) ontologies used to mediate the search of the learning materials on the internet (Cakula and Salem, 2011). An intelligent learning system based on a multi-agent approach consists in a set of intelligent agents, which have to communicate. They collaborate through messages. Software agents can understand and interpret the messages due to a common ontology or the interoperability of the private ontologies.

C. Benefits of data mining methods to e-learning

This section presents the benefits of some of the data mining methods in e-learning domain. Further details can be found in (Salem, 2008):

1) Information Visualization (IV) in Smart Learning

The IV in e-learning can be used in the following educational tasks; admitted questions, complementary assignments, exam scores, etc. Moreover, intelligent visualization tools (e.g. GISMO CourseVis) enable instructors to manipulate the graphical representations generated, which allow them to gain an understanding of their learners and become aware of what is happening in distance classes.

2) Clustering and classification in Smart Learning

Clustering and classification approaches have been used in smart learning for the following tasks:

- Finding clusters of students with similar learning characteristics and to promote group-based collaborative learning as well as to provide incremental learner diagnosis.
- Discovering potential student groups with similar characteristics to a specific pedagogical strategy.
- Grouping learners/students as hint-driven or failure-driven and finding students' common misconceptions.
- Identifying learners/students with little motivation and finding remedial actions in order to lower drop-out rates.

4. SOME EXAMPLES OF SMART LEARNING AND TUTORING SYSTEMS

This section presents some of our intelligent learning/tutoring systems developed by the author and his colleagues during the last ten years

A. Expert Systems for Heart Diseases Diagnosis

In this application (Salem and Hodhod, 2002) we have developed two versions of expert systems for heart diseases diagnosis. The first one uses the rule-based reasoning while the second one uses case-based reasoning. The system's knowledge base of the first version is composed of 24 facts and 65 rules for 24 heart diseases. The system is implemented in Visual Prolog and has been tested for 13 real experiments (patients). The experimental results have shown 76.9% accuracy in estimating the right conclusion. In the CBR version, the knowledge is represented in the form of frames and built the case memory for 4 heart diseases namely; mitral stenosis, left-sided heart failure, left-sided heart failure, stable angina pectoris and essential hypertension. The system has trained set of 42 cases for Egyptian cardiac patients and has been tested by another 13 different cases. Each case contains 33 significant attributes resettled from the statistical analysis performed to 110 cases. The system has been tested for 13 real cases. The systems are able to give an appropriate diagnosis for the presented symptoms,

signs and investigations done to a cardiac patient with the corresponding certainty factor. It aims to serve as doctor diagnostic assistant and support the education for the undergraduate and postgraduate young physicians.

B. Web-Based Breast Cancer Ontology

The breast cancer ontology was encoded in OWL-DL format using the Protégé-OWL editing environment (Salem and Alfonse, 2007). The knowledge was collected from: MedicineNet, World Health Organization (Salem, 2010), breastcancer.org, ehealthMD and National Comprehensive Cancer Network. In this ontology we have two main super classes, namely, (a) MedicalThings which has sub classes Diseases, Medical_Interventions, Pathological_Category, References, and (b) People which has the sub classes; men and women. Our results indicate that, the breast cancer is described in terms of its symptoms, causes, stages, pathological category, diagnosis and treatment. In this context, we described causes, stages, and symptoms as references. While diagnosis and treatment are described as medical interventions. The main benefits from this ontology are to allow finding and locating information about breast cancer needed for interested users and domain experts.

C. Data Mining Approach for Slow Learners Students (SLS)

Recently, data mining approach have been improved in order to classify the students' records based on their educational activates. Slow learners students (SLS) suffers from low rate of understanding the educational materials; therefore, they need special education methods and strategies to educate them. In our research (Mohammad, Mahmud, El-Hobart, Rushdie, and Salem, 2013 a, b, c) we developed a smart classification model of English course e-Learning system for SLS. Our model is based on the decision trees technique to determine the most effective learning pattern for teaching English language for those students. The model was implemented on 300 students in the fourth grade at Kuwait, age 8 -10. We obtain promising results.

D. Ontology for e-business paradigms

Nowadays, e-business, or digital business, is the integrated execution of all business analytics processes of an enterprise by means of

smart computing and informatics. Recently we developed a web-based ontology for e-business paradigms (Salem and Parusheva, 2018a). In this work, five web-based ontologies were designed for the following: (a) e-business applications; (b) e-business participants; (c) e-business infrastructure; (d) e-business support areas, and (e) fields in e-business. The developed ontologies were implemented in ontology web-based language OWL2 using the Protégé smart tool version 5.0.0 editing environment.

E. Knowledge Discovery for Banking Risk Management

In the era of Internet of Things (IoT), decision makers and managers directed their efforts to use the knowledge discovery (KDD) and data mining (DM) process in order to enhance their decision-making policies, avoid unpredictable frauds and increase the revenue opportunities for their financial agencies. In our research labs (Sobhy, Abbas and Salem, 2015), we present a novel approach in banking data bases in order to spot the light on the most effective features, which affect the decision-making process. In the data preprocessing phase, the K-Nearest Neighbor algorithm has been implemented to measure the distance between objects' vectors. The feature selection process has been implemented using Microsoft SQL server analysis services and applied on real financial database obtained from PKDD'99. The results show that the features are reduced from 47 attributes into 17 significant attributes.

CONCLUSION

Computational intelligence (CI) techniques give tutoring and e-learning systems added computing capability, allowing them to exhibit more intelligent behavior. CI and Knowledge engineering paradigms offer robust intelligent techniques for accumulating, changing, updating, managing and representing knowledge. These techniques enable users with learning mechanisms that help to induce knowledge from raw data. Furthermore, ontological engineering offers a promising way to develop efficient tutoring and learning systems capable to facilitate knowledge sharing, refine, search, and reuse. On the other side, the development of intelligent tutoring and e-Learning systems is a very difficult and complex process that raises a lot of technological and research challenges that must be addressed in an interdis-

ciplinary way. Moreover, the convergence of artificial intelligence, data science machine learning, educational technology and web science is enabling the creation of a new generation of web-based intelligent e-learning and tutoring systems. The web based of such systems can enhance the online education/ learning/training processes.

REFERENCES

1. ABDRABOU, E. A. M. and SALEM, A.-B. M. (2008) Case-Based Reasoning Tools from Shells to Object-Oriented Frameworks. Advanced Studies in Software and Knowledge Engineering- Supplement to the International. *Journal INFORMATION TECHNOLOGIES & KNOWLEDGE*, Ed. Krassimir Markov, Krassimira Ivanova, Ilia Mitov Sofia: Institute of Information Theories and Applications FOI ITHEA, pp. 37-44.
2. ALFONSE, M., AREF, M. M. and SALEM, A.-B. M. (2012) Ontology-Based Knowledge Representation for Liver Cancer. Proceedings of the International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business, Luxembourg, G. D. of Luxembourg, April 18-20, 2012, pp. 821-825.
3. CAKULA, S. and SALEM, A.-B. M., (2011) ONTOLOGY-BASED COLLABORATIVE MODEL FOR E-Learning. Proc. of the Annual International Conference on "Virtual and Augmented Reality in Education" (VARE 2011) (combined with EEA and Norwegien Financial Instruments project practical conference "VR/AR Applications in Training"), Vidzeme University of Applied Sciences, Valmiera, Latvia, 18 March 2011, pp. 98-105.
4. CIOS, K. J., PEDRYCZ, W. and SWINIARSKI, R. W. (1998) Data Mining Methods for Knowledge Discovery. Kluwer.
5. GREER, J. (1995) Association for Advancement of Computing in Education (AACE). Proceedings of AI-ED 95, World Conference on Artificial Intelligence in Education.
6. KOLONDER, J. (1993) Case-Based Reasoning, Morgan Kaufmann.
7. MAZZA, R. and MILANI, C. (2005) Exploring usage analysis in learning systems: Gaining insights from visualizations. Proceedings of Workshop on usage analysis in learning systems at 12th international conference on artificial intelligence in education, New York, USA, pp. 1-6.

8. SALEM, A.-B. M. (2007) The Role of Artificial Intelligence Technology in Education. Proceedings of 5th International Conference on Emerging e-learning Technologies and Applications, Information and Communication Technologies in Learning (ICETA2007), The High Tatras, Slovakia, pp. 1-9.
9. SALEM, A.-B. M. (2008) Data Mining Technology in e-Learning. Proceedings of 6th International Conference on Emerging e-learning Technologies and Applications, Information and Communication Technologies in Learning (ICETA2008), Stara Lesna, The High Tatras, Slovakia.
10. SALEM, A.-B. M. (2010) Ontological Engineering in e-Learning. Proceedings of 8th International Conference on Emerging e-learning Technologies and Applications, Information and Communication Technologies in Learning (ICETA2010), Stara Lesna, The High Tatras, Slovakia.
11. SALEM, A.-B. M. (2011) Intellectual E-Learning Systems. Proc. of the Annual International Conference on "Virtual and Augmented Reality in Education" (VARE 2011) (combined with EEA and Norwegian Financial Instruments project practical conference "VR/AR Applications in Training"), Vidzeme University of Applied Sciences, Valmiera, Latvia, pp. 16-23.
12. Salem, A.-B. M. and ALFONSE, M. (2009) Ontological Engineering Approach for Breast Cancer Knowledge Management. Proc. of Med-e-Tel, the International eHealth, Telemedicine and Health ICT for Education, Networking and Business. Luxembourg. March 30 – April 5, 2009, pp. 320-324.
13. SALEM, A.-B. M. and HODHOD, R. A. (2002) Developing a Hybrid Expert System Prototype for Diagnosis of Heart Diseases. Proceedings of IFIP 17th World Computer Congress, Stream 8: IIP 2002 Intelligent Information Processing, Montreal, Canada, August 20-30, Poster Presentation.
14. SALEM, A.-B. M. and PARUSHEVA, S. (2018a). DEVELOPING A WEB-BASED ONTOLOGY FOR E-BUSINESS. International Journal of Electronic Commerce Studies, Vol. 9, No. 2, pp .119-132.

15. SALEM, A.-B. M. and PARUSHEVA, S. (2018b) Exploiting the knowledge engineering paradigms for designing smart learning systems. *Eastern European Journal of Enterprise Technologies*, Vol. 2, Issue: 33635, pp. 38-44.
16. SALEM, A.-B. M. and VOSKOGLOU, M. G. (2013) Applications of the CBR Methodology to Medicine. *Egyptian Computer Science Journal*, Vol. 37, No.7, pp. 68-77.
17. SOBHY, M., ABBAS, S. and SALEM, A.-B. M. (2015) Knowledge Discovery for Banking Risk Management., *IPAC '15*, November 23-25, Batna, Algeria ACM.
18. TANKELEVCIENE, L. and DAMASEVICIUS, R. (2009) Characteristics for Domain Ontologies for Web Based Learning and their Application for Quality Evaluation. *Informatics in Education*, Vol. 8, No 1, pp. 131-152.
19. TANTAWI, M., REVETT, K., TOLBA, M. F. and SALEM, A.- B. M. (2012) Fiducial Feature Reduction Analysis for Electrocardiogram (ECG) Based Biometric Recognition, *International Journal of Intelligent Information Systems*, Springer, pp.1-23.
20. Thakaa Mohammad, Abeer M.Mahmoud, EL-horbart, E., Roushdy, M., and Salem, A.(2013b), "An e-learning Classification Model of Teaching English for Slow Learners", *ICICIS'13 Conference and Workshop on Intelligent Technologies for E-learning ITEL 2013*, Cairo, Egypt, Dec 14-16.
21. Thakaa Mohammad, Abeer M. Mahmoud, EL-horbart, E., Roushdy, M., and Salem, A.(2013c), "Towards Applying Data Mining Techniques for Intelligent E-learning"1st *WSEAS International Conference on Computer Supported Education (COSUE'13)*, Vouliagmeni, Athens, Greece, May 14-16.
22. WITTEN, I. H. and FRANK E. (2005) *Data Mining – Practical Machine Learning Tools and Techniques*. 2nd ed, Elsevier.

PROPOSAL OF THE MULTI-ASPECT ANALYSIS OF ICT'S NEEDS IN MANAGEMENT

Marian Niedźwiedziński¹

¹ Collage of Computer Science & Skills in Łódź

Abstract

The Multi-aspect Analysis of Reserves in Administration (**MARIA**) is a proposal of assessing main factors determining effectiveness in a business organisation (BO) to show how important among them is information. In other words - the aim of MARIA is to assess most probably response of the organisation (in the sense of its effectiveness) on possible ICT investments. Without such an assessment, one can easily magnify the role of ICT in an organisation and in a consequence to magnify effectiveness of possible projects and initiatives within in ICT area. An approach, which will be presented, treats implementation of ICT as a risky business and therefore, it requires an adequate prior analysis, which may even take precautions against useless investments.

Keywords: *administration, effectiveness, Information & Communication Technology.*

INTRODUCTION

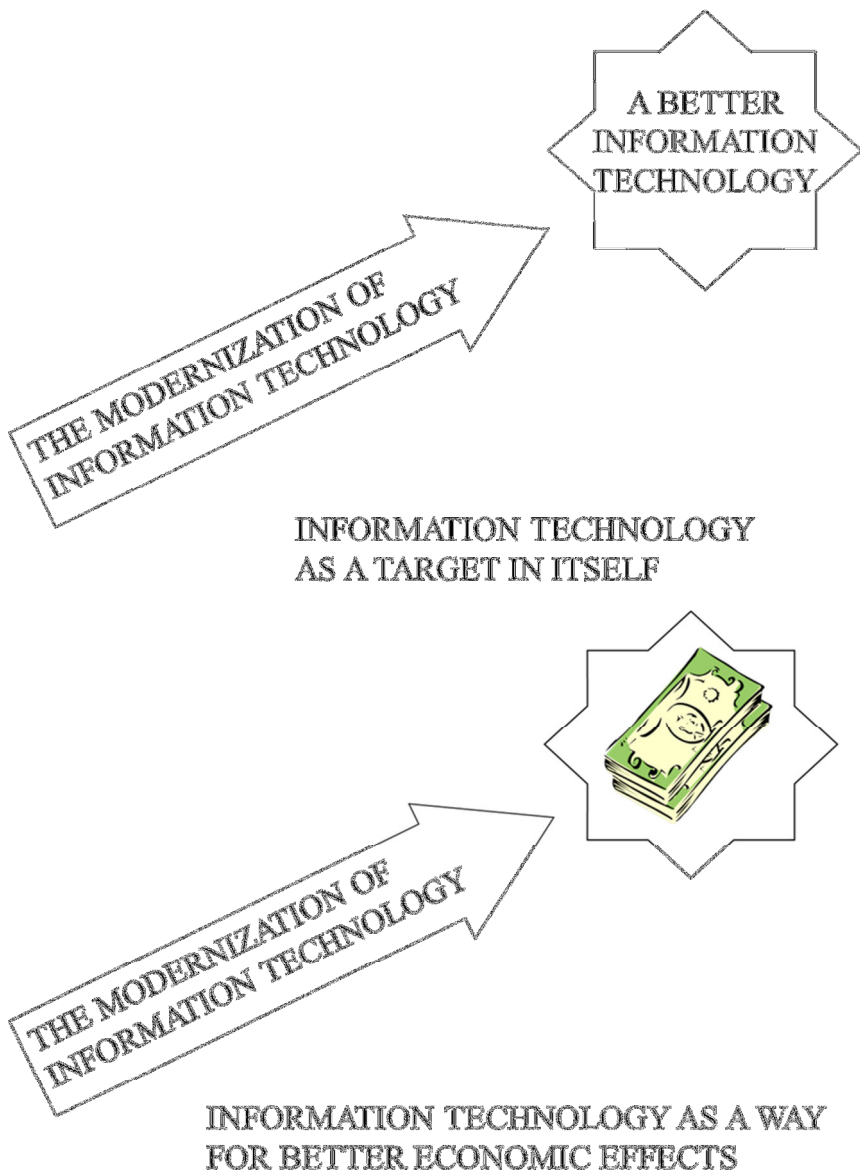
The most popular opinion says, that ICT effectiveness in BO is up to a degree, in which ICT provides for users' needs. That is why, the most popular way of doing a prior analysis in ICT projects is just asking future users (especially decision makers) about their information needs.

I do not think however, that it is good starting point for building effective ICT applications. I think so because – in my opinion there is no dead certainty, that a delivery of additional information (which is planned as a result of a given ICT investment) means better work of the organisation and in a consequence a higher effectiveness of the given BO. In other words - there is no dead certainty that in the given BO exist reserves, which can be exploited by better informed decision makers for achieving better BO's effectiveness.

This opinion seems to be credible taking into consideration a significant number of failures, which still can be observed in the area of ICT investments.

Among numerous methods for ICT effectiveness evaluation, I have not met anyone basing on analysis of reserves existing in a given organisation. In my opinion however whose internal reserves have a key importance. This opinion is conformable to a very widely known judgement, that a main source of effectiveness' growth, being a result of ICT investment, is an elimination (or a limitation) of organisational reserves, which exist in a given BO. Therefore I have proposed and elaborated an adequate method in the paper. The essence of the Multi-aspect Analysis of Reserves in Administration (MARIA) is defining how favourable (or not favourable) are the conditions for effective implementations of ICT in a specific business organisation, basing on its reserves' analysis. Another interpretation of the proposed approach is checking an organisation's sensibility of ICT implementation in terms of its effectiveness. The method treats ICT investments as a specific kind of business and therefore it requires an adequate prior analysis and takes precautions against useless investments.

So, consequently – a modernisation of information system should not be treated as a target in itself, but as a way of earning extra money from existing business - see fig.1.



**Figure 1. Two ways of thinking
on “The modernisation of information technology”**

1. RESERVES IN A BUSINESS ORGANISATION

The following schema (fig. 2) illustrates the most important determinants of business organisation's (BO's) effectiveness and basic relations between them.

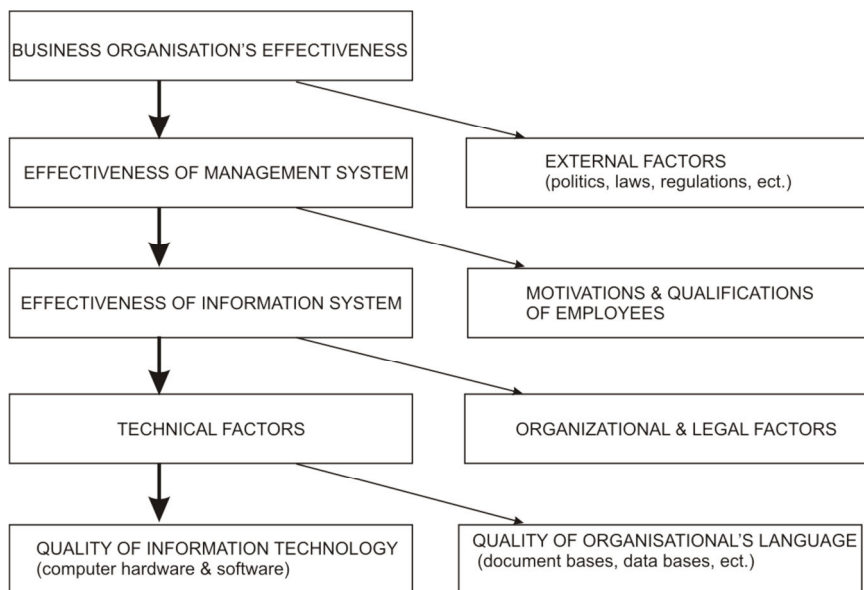


Figure 2. The main factors determining effectiveness in BOs' (thick, vertical arrows mark a specific “critical path” of ICT investments' effectiveness)

The above figure shows, that on the highest level of consideration, BOs' effectiveness is determined mainly by two groups of factors:

- effectiveness of management system (it means - internal factors controlled within a specific BO), and
- external factors (it means - politics, laws, regulations, economic situation and other outside circumstances).

External factors, as being out of control from a specific BO point of view, are beyond my interest in a further discussion. I will concentrate on an effectiveness of a management system.

The next level of elements in figure 2 proves that the effectiveness of management system within a specific BO is determined by:

- effectiveness of information system and
- employees' motivations and qualifications (human aspect).

The level of motivation (reward power) determines the effectiveness of a given BO, because no effects can be accomplished - unless employees are encouraged to work on them. In other words - reward power determines effectiveness of BO in the sense, that there are no effects if there is no system-inclining tasks' performers to obtain them.

At ECIS'07 (European Conference on Information Systems) was presented an interesting opinion that ICT applications' value depends on a quality of three elements: hardware, software and "human-ware". From their importance and visualization points of views – the iceberg analogy can be applied (see fig. 3 below). Similarly, as in the case of iceberg, in the case of ICT applications - easily visible factors (hardware and software) have not a dominating role. The most important factor is "under water's surface".

An importance of a human aspect within ICT value – an iceberg analogy

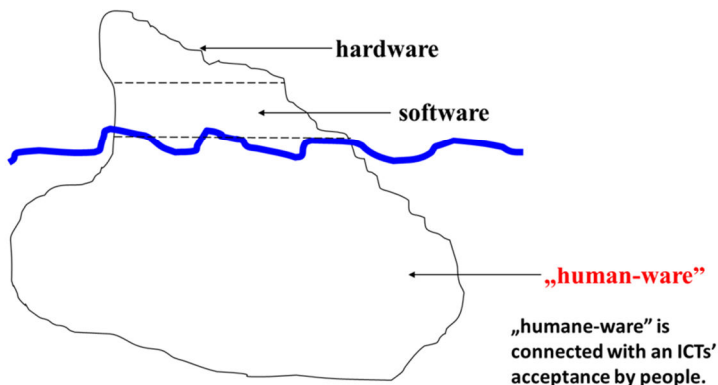


Figure 3. Iceberg analogy

The "human-ware" means users' readiness to transfer advantages of ICT into real, valuable effects for investing organizations. Such a readi-

ness can have many degrees. Up to the degree of such a readiness – an adequate degree of an ICT application's effectiveness can be achieved. In some extreme situations ICT applications can have even negative effects on BO. It can happen, when the organizations had spent money and put other efforts for applications, which were rejected by potential users. It is necessary to stress here, that no benefits can be achieved just by an ICT's application itself. Always is needed a positive approach of its potential users, who can make the applications workable and effective.

The second mentioned factor (qualifications) is well known. Every businessman understand clearly that skills of employees (as well as lack of skills) determine the effectiveness of a given BO. As qualifications' aspect however are not the main field of my interest in the paper, I will not discuss them further and concentrate on the effectiveness of information system.

According to the fig. 2, the determinants of information system's effectiveness (in a given BO) are as follows:

- technical issues, as well as
- organisational and legal issues.

Organisational and legal issues seem to be very important. The capacity of information channels may be significantly reduced by an excessive centralisation in organisation's management, absence of self-contained tasks' system and poor horizontal relations. The more levels in a hierarchy of management system - the bigger probability of deforming (including some intentional deformation), delaying or reducing amount of information in its channels. Too long or too complicated information channels are overloaded and therefore not effective.

The very important barrier of information system's effectiveness is lack of access to important business data arising from existing legal regulations. For instance – terms of delivery offered by competitors in public procurement processes. Although those terms are very wanted data – they are not attainable because of adequate legal rules.

As organisational and legal issues are not in my direct field of interest in the paper, I will concentrate on determinants, which fall into technical factors:

- quality of information technology and
- quality of an organisational „language”.

Quality of an organisational „language” is connected with such elements in BOs as: normative bases, documentation bases, databases, access to public registers etc. All kinds of possible problems with quality or accessibility to those sources of data may dramatically reduce the effectiveness of information system in a given BO.

Defining the quality of information technology two elements must be taken into consideration: quality of computer hardware and quality of computer software. Too poor or improper technical equipment of information system significantly reduces its capacity.

2. CONDITIONS OF EFFECTIVE ICT INVESTMENT

Considering all those aspects we can define a special set of conditions, on which investing in ICT can be really useful and effective for a given BO:

1. It must be possible to increase effectiveness of a given BO by improvements in its management system;
2. A progress in effectiveness of management system can be achieved first of all by improvements in a BO's information system;
3. An effectiveness of the information system depends first of all on technical factors;
4. The most important technical factor limiting capacity of the information system is poor or improper quality of the information technology used in a given BO;

As conditions mentioned above are in many BOs not fulfilled, in many of them investments in ICT are not effective or they are not effective enough. So, one may say, that vertical arrows in fig.2 show "a critical path" of the effectiveness in ICT investments.

Another way of presenting the ideas mentioned above is shown in fig.4.

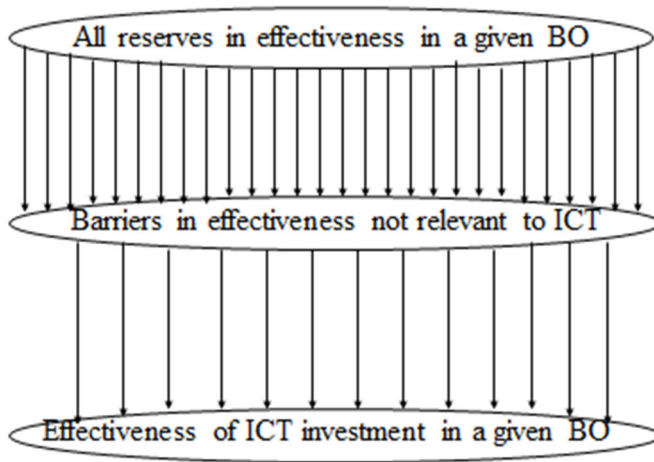


Figure 4. A process of defining an effectiveness of an ICT investment in a given BO

One can see in it all reserves in BO's effectiveness (represented by arrows), which are partly blocked by factors not relevant to ICT. In the end it appears, that ICT seems to be, in many cases, not as much important as one can expect.

The way of thinking presented above leads to two general conclusions:

1. Effectiveness of ICT investments in OBs is not automatic, it cannot be guaranteed in all situations. It depends on a specific constitution of various conditions in a given BO. This is why it is necessary to use an universal, objective and interdisciplinary analysis of those conditions, which can be specific to every specific BO. Such an analysis will enable:

- to have a clear understanding of the main determinants of organisation's effectiveness and ICT's role among them;
- to make a reliable evaluation of potential ICT investments' effectiveness,
- to create a reasonable strategy of investing in ICT.

Without such an approach one can easily magnify the ICT role in BO's improvement and as a consequence magnify the effectiveness of possible ICT's projects and initiatives.

2. There seems necessary to reorient a starting point for new information systems' development from considering information needs expressed by potential users - to analysis of reserves, which exist in a given BO.

I think so, because there is no equivalence between information needs declared by users and reserves in a given BO's effectiveness, which can be release due to delivery of additional information.

3. PRACTICAL ASPECTS OF THE PROPOSED METHOD'S IMPLEMENTATION

First of all, a special invention group¹ should be established in an organisation. The mission of the group should be representing users' point of view during a process of preparation and realisation of ICT investment.

The group should consist of employees, who have the following attributes:

- a work experience long enough to be familiar not only with formal but also with nonformal aspects of an organisation;
- an enthusiasm for innovations;
- an ability to see various details as combined elements of a complex system;
- an ability to separate important problems from not important ones.

Next, members of the invention group should estimate values of adequate reserves using a balance methodology as it is shown in the figure 5 below.

¹ The name comes from "inventics" - a new discipline of knowledge, which explores problems of invention, creativity, ways of looking for new, non-conventional solutions etc.

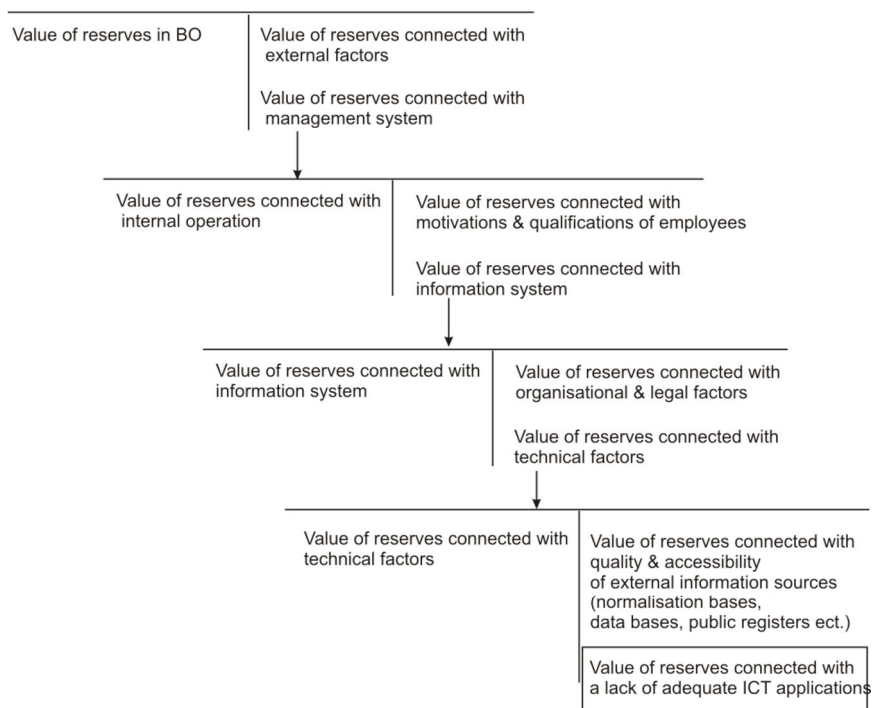


Figure 5. A balance estimation of reserves in succeeding levels of the proposed analysis

CONCLUSION

It seems, that it is necessary to use universal, objective and interdisciplinary approach in analysis of possible ICTs' investments in a given organisation:

- to make the reliable evaluation of an ICT application effectiveness,
- to make the adequate evaluation of ICT projects' variants profitability,
- to create the reasonable strategy of information system modernisation and in sum
- to treat a modernisation of information systems within an organisation as a business, but not as a target in itself and therefore to take precautions against needless investments.

Without such an approach, in my opinion, one can easily to magnify the role of ICT in Business Organisations' improvement and in a consequence to magnify effectiveness of possible projects and initiatives.

DIGITAL TRANSFORMATION IN BULGARIA

Krasimir Shishmanov¹

¹ *D. A. Tsenov Academy of Economics, Svishtov, Bulgaria,
k.shishmanov@uni-svishtov.bg*

Abstract

In today's dynamic and highly competitive environment, companies are aware that they need to transform their business in order to survive in the contemporary digital world. They need to employ a flexible approach and invest in developing the digital skills of their employees in all departments of the organizations so as to adapt to market changes and succeed in future.

Keywords: *digital transformation; digital skills; digital economy; digitization; e-commerce.*

Rapid development of e-services and e-commerce in recent years has led to significant changes in business and society and to the appearance of new products and services. A key factor in making an online deal is doing it over the Internet, which brings a number of advantages - overcoming time and spatial barriers, the convenience of viewing the goods which are offered, time-efficiency and easy access to services.

According to data provided by Ecommerce Europe, the number of consumers on the European market who do their shopping online is 296 million. More than 2,500,000 jobs are directly or indirectly linked to e-commerce. The number of set up online businesses exceeds 750,000. Over 4.2 billion orders were sent in one year. The share of the European Internet economy equaled 2.59% of European GDP in 2015. This percentage is expected to double by the year 2020.

In 2016, 20% of enterprises in the EU reported selling goods and products online to end consumers and other enterprises. The highest share of online business sales was recorded in Ireland (30%), followed by Denmark (29%), Germany and Sweden (28%). The highest volume of turnover from online business sales in EU member states was that in Ireland (35% of total turnover volume), followed by the Czech Republic (31%) and Belgium (29%).

In 2016, 13% of companies in the EU made online sales to consumers (B2C) and 12% made online sales to other companies (B2B). Five

percent of total turnover was generated from online sales, 2% from B2C sales and 3% from B2B sales. The highest number of companies with B2C sales was in Ireland (22%), while companies in Germany made the highest number of B2B sales (18%). The highest share of turnover generated from B2C sales was recorded in Ireland (12%), while the share of turnover generated from B2B online sales was the highest in Belgium (7%). (Online shoppers & e-purchases)

Exported data clearly indicate the need for digital transformation. It is triggered by the fact that the majority of consumers have ‘transported’ to the internet where they read the news, make purchases, play games, watch movies, etc. Hence, the question ‘How can companies organize their business so as to meet growing online needs?’

Digital economy provides numerous opportunities, yet it requires a deep understanding of its nature, the specifics and principles of its functioning, its advantages and disadvantages, as well as how to benefit from the opportunities it provides in order to promote its further development and produce a positive effect on the digital transformation of business.

In terms of the issues related to digital transformation, there are several major aspects that need to be taken into account in order to ensure the success of adopted measures:

- Transformation needs to be based on leading technologies. The choice depends on the goals of the organization, but there are other points to be considered as well, e.g. how strong community support for the project is and whether technical support will be provided. Another good idea is to find out how different technologies are accepted by the market;

- Transformation needs to happen transparently. Any process that radically changes the usual conditions entails certain risk and digital transformation is no exception. Since the process involves many departments, it often gives rise to discussions about the distribution of responsibilities. Transparent decision making is the most effective strategy when determining the role of each team member;

- Management is crucial to the implementation of the project. An autocratic management style cannot guarantee the success of the transformation. Rather, management should explain to employees the actions that have been planned and what is expected from the team. Manage-

ment can then count on the cooperation of employees. In addition, throughout the whole process of transformation, from the emergence of the idea to its full realization, including problems faced by individuals, employees need to be confident that they have the constant support of the management of the enterprise;

- It is necessary to document the process of digital transformation. Reporting systems, defects and progression, automation of assemblies and dispatches, automation testing platforms, monitoring of production systems and so on - these are the control functions, which are part of the companies;

- The success of any plan depends on the sequential implementation of the project at specified stages. It must be planned in a manner which makes it possible to assess economic effectiveness. To implement changes at an operational level, it is necessary to create a cross-functional team consisting of employees in different departments who will be responsible for specific aspects of the process.

We could therefore conclude that projects about the digital transformation of organizations are not limited to the introduction of ICT. The success of such projects is also determined by the changes that need to be systematically implemented in several aspects simultaneously. Based on the expectations about the results of digital transformation, we define these areas:

- Information and communication technologies as the main instruments of transformation;

- The results of business processes which depend directly on the development goals of an organization in accordance with the existing business strategy;

- All categories of participants in business processes (executors, managers, stakeholders, clients), i.e. all persons who contribute to the accomplishment of the process, or provide feedback for its evaluation and further improvement;

- The environment that creates the conditions for the best development and implementation of technology projects in order to improve business processes. It is established through the cooperation and interaction between the employees of the organisation.

The economic situation and the competitiveness of countries is largely determined by the progress of digital transformation. In order to identify the rank of each country, a special composite index is employed – the Digital Economy and Society Index (The Digital Economy and Society Index (DESI), 2018). It summarises relevant indicators of Europe's digital performance and tracks the evolution of EU member states in digital competitiveness.

According to index data (see fig. 1), Denmark, Sweden and the Netherlands were the countries with the best digital performance in the EU in 2018, followed by Luxemburg, Ireland, Great Britain, Belgium and Estonia, while Romania, Greece, Bulgaria and Italy scored the lowest ratings.

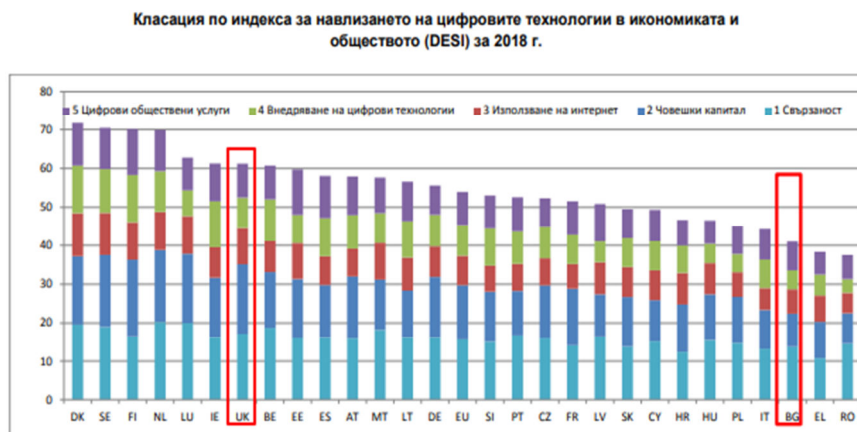


Figure 1. The DESI index ranking

Source: <https://ec.europa.eu/digital-single-market/en/desi>

In general, there has been an upward trend in the development of the DESI index in all EU member states, its growth being highest in Ireland and Spain and lowest in Denmark and Portugal. The DESI index is based on six dimensions that are employed to assess digital economy:

1) Connectivity

The dimension is used to measure the deployment of broadband infrastructure and its quality. Access to fast and ultrafast broadband-

enabled services is a necessary condition for competitiveness. In terms of connectivity, the highest score was that of the Netherlands, followed by Luxemburg and Denmark. The countries with the lowest scores on the dimension were Greece, Croatia and Italy.

2) Human Capital and Digital Skills

The dimension measures the skills required to take advantage of the opportunities offered by digital technology. The highest scores in terms of human capital were those of Finland, the Netherlands and Sweden, while the lowest ones were those of Romania, Bulgaria, Greece and Italy. Eighty-one percent of EU citizens use the Internet at least once a week, which is a 2% increase compared to the previous year.

3) Use of Internet Services by Citizens

This dimension accounts for various online activities, such as the consumption of online videos, music, games, making video calls, online shopping, banking, etc. According to the DESI index, the most active users of Internet services are those in Denmark, Sweden and the Netherlands, in contrast to users in Romania, Italy and Bulgaria – the countries that ranked at the bottom on this dimension.

4) Integration of Digital Technology by Businesses

The Integration of Digital Technology dimension measures the digitization of business and e-commerce. By adopting digital technologies, businesses can enhance their efficiency, reduce costs and better engage customers and business partners. In addition, the Internet offers access to wider markets and higher potential for growth. On this dimension, Denmark, Finland and Ireland scored highest, while Romania, Poland and Bulgaria scored lowest. European businesses are increasingly integrating digital technology. The share of businesses that implemented an electronic information sharing systems, for example, rose from 26% in 2013 to 34% in 2017; the share of companies that sent e-invoices grew from 10% in 2013 to 18% in 2016. The share of businesses that use social media to better engage their customers and partners increased from 15% in 2013 to 21% in 2017. There was a 3% increase in the percentage of small and medium-sized enterprises selling online in the period from

2013 (14%) to 2017 (17%). Less than half of those companies sold their products and services cross-border to another EU member-state, though.

5) Digital Public Services

This dimension measures the digitization of public services, focusing on eGovernment and eHealth. Modernisation and digitization of public services can lead to efficiency gains for the public administration, citizens and businesses. Finland, Estonia and Denmark had the highest scores on this dimension, in contrast to Greece, Hungary and Romania which were among the countries with the lowest scores.

6) Research and Development ICT

The Research and Development ICT presents analysis on the trends of the ICT sector and Research and Development provided by the European Commission, as well as external studies conducted at the request of the Commission. Over the first four years, the funding provided by the EU Research and Innovation Programme, Horizon 2020, amounted to EUR 5.7 billion.

Bulgaria ranked 26th out of the 28 EU member states in DESI 2018 (DESI 2018 Country Report Bulgaria). The country retained its ranking from the previous year despite some slight improvements to its score. Over the last year, the country made progress in connectivity and the availability of digital services.

Digital public services improved in particular, resulting in an increased number of e-government users. This is due to the adoption of a strategic framework which is actively employed at present, as well as the optimized ICT budget framework.

In 2017, a governance programme for the period 2017-2020 was adopted with measures and priorities related to the digitization of state administration. The programme includes the integration of a national scheme for electronic identification; developing basic infrastructure further; establishing links between important registers and ensuring operational registers to make the transition to automated exchange of data and electronic documents safe.

Digital public services for businesses show a significant improvement. As of 2018, tax declarations for legal persons are to be submitted

only in electronic form, while this is still optional for physical persons and is incentivized via a tax rebate. Bulgaria performs well in the area of open data – there are currently over 7,000 datasets from different national and regional administrations and agencies.

In terms of eHealth services, Bulgaria performs below the EU average. To improve performance, the Ministry of Health has initiated a project on the development and implementation of a National Health Information System (NHIS). The project includes developing health registers, implementing a pharmacotherapeutic expert system to flag electronically possible drug interactions, building an online platform of the NHIS, and introducing electronic health records, referrals and prescriptions. Overall, Bulgaria shows very good progress in the area of digital public services.

Overall, 50 % of Bulgarian companies consider market expansion and entering new markets their main priorities. E-commerce is considered to be essential for the digital transformation of companies by 57% of respondents, while 77% of respondents state that it is part of their policy for digitization of the business.

An interesting fact is that 74% of businessmen are confident that they can use a combination of the digital solutions and services necessary for their organizations. Entities also seek to ensure safety (61%), speed / responsiveness (60%), consistency across channels (48%), and personalization / relevance (47%). Of those that do not yet have a defined strategy (24%), many know it is important, but have not yet started the process of defining their strategy (49%) (Progress, 2016).

One of the major issues which organisations are facing is the design of their own digital transformation strategy. According to surveys, 76% of global respondents have defined their strategy. Most of them are not fully aware of the importance of a digital strategy accompanying all stages of production.

Two-thirds (67%) of the managers of Bulgarian companies confirm that digital transformation is part of their corporate strategy, while 59% are apprehensive that they are lagging behind in digital transformation, which renders them uncompetitive and could result in serious financial problems.

Overall, 70% of Bulgarian companies are planning to invest in digital metrics and data analysis in order to be more flexible in their web-site

management. On this dimension, Bulgaria scores higher than the average global score of 47%.

The factors which are globally viewed as critical for the digital transformation of business and are therefore considered an investment priority are presented in the table 1 below (Ibid).

Table 1

Key investment priorities

Establishing digital metrics and measurements	40%
Building applications that support the customer engagement model or other digital initiatives	40%
Overhauling customer-facing technology systems	41%
Using mobile to support operational improvement efforts, customer engagement initiatives/new business opportunities	42%
Process improvements to enable more agility with website/mobile/social platforms	45%
Building and managing web properties as part of an integrated approach	46%
Customer journey/digital touchpoint mapping	57%

The extremely low level of digital skills and their weak integration in economic activity are the main problems faced by Bulgaria. Low digital skills, the lack of ICT specialists and insufficient investment in digital infrastructure are the main factors behind the low digitization of the country.

The level of digital skills is one of the lowest in the EU. Despite the increase in the number of people with basic digital skills from 26% in 2017 to 29% in 2018, our country continues to rank among the lowest ranking countries in the EU.

More than half of Bulgarian companies state that they do not have the digital skills required to implement digital transformation (63%) or digital leadership (53%). In comparison, respondents on a global scale point out that high dependence on IT groups prevents the establishment of a specialized IT department for implementation of the strategy (58%), that they do not have digital skills to implement the strategy (58%), that there is no digital leadership for designing a strategy (56%), and there is no centralized strategy or management (61%) and those are key barriers to improving customer service through digital transformation.

Only 10% of Bulgarian enterprises are confident that they have the employees they need and that they have been appointed at the right position within the organization to enhance digital transformation. In contrast, 25% of all enterprises on a global scale are highly confident in their ability to implement and integrate all sources into a comprehensive digital business strategy, to manage the entire lifecycle of the application, to map user requirements and combine different initiatives into a meaningful and well-targeted digital transformation strategy.

In 2014, an e-skills strategy was elaborated to modernize the education system, improve access to quality education and increase the offer of IT training. As a result, the first IT courses were carried out in first and fifth grade in 2016/2017, to be continued in seventh grade in 2017/2018, and then gradually be extended in the following years.

In 2018, the number of Science, Technology, Engineering and Mathematics (STEM) graduates decreased slightly compared to 2017, but still remained a positive element compared to the other dimensions of the Human Capital index. The European Center for the Development of Vocational Training (Cedefop) found a mismatch between the increasing demand for highly-skilled engineering professionals and decreasing supply that was due to the declining number of STEM students and the frequent lack of job-specific skills in graduates. ICT specialists accounted for as little as 2.7% of total employment, which was below the EU average by 1 percentage point. Many of the companies operating in the IT sector in Bulgaria established corporate academies that offer digital skills in order to address the shortages of advanced digital skills in Bulgarian businesses. The country has also developed a number of strategies for broadband connectivity, digital technology skills, digital transformation of the business and eGovernance.

REFERENCES

1. ALTUKHOVA N.F. (2012) *Models of coherence as a tool for managing information systems* // Vestnik Universiteta. No. 5.. pp.71–77
2. BRIAN SOLIS, *The 2017 State of Digital Transformation*., URL: <https://marketing.prophet.com/acton/media/33865/altimeter--the-2017-state-of-digital-transformation?>

3. DAVID L. RODGERS (2017) *The Digital Transformation Playbook: Rethink Your Business for the Digital Age*. AITEKO, p. 344
4. Digital Economics // Strategy, No. 3 (28) 2017. pp. 86-87
5. Every fourth company has failed in digital transformation. // Chief information Officer, No. 9, November 2017. S. p.6
6. https://www.progress.com/docs/default-source/default-document-library/landing-pages/dach/ebook_digitaltransformation_final.pdf
7. <https://ec.europa.eu/digital-single-market/en/desi>
8. Progress (2016) Are businesses really digitally transforming or living in digital denial? https://www.progress.com/docs/default-source/default-document-library/landing-pages/dach/ebook_digitaltransformation_final.pdf

MODELING THE LIFE CYCLE OF E-LEARNING COURSE USING MOODLE CLOUD LMS

Tatyana Makarchuk¹, Valery Trofimov², Svetlana Demchenko³
¹ Saint-Petersburg state university of economics, Saint-Petersburg,
Russia, tmakarchuk@mail.ru
² Saint-Petersburg state university of economics, Saint-Petersburg,
Russia, twv@mail.ru
³ Saint-Petersburg state university of economics, Saint-Petersburg,
Russia, shh-svetlana@yandex.ru

Abstract

Many higher education institutions use Learning Management Systems (LMS) for e-learning. E-learning research shows the heterogeneity of approaches to creating an e-learning resource using LMS, which causes difficulties in assessing the quality of e-learning. There is a need to develop common e-learning course design technology using LMS for high school with considering trends in the development of digital society, such as mobile application, cloud computing, analytics tools and others. The article shows how you can use Scrum methodology for e-learning course development with Moodle Cloud LMS based on quality standard for e-learning ISO/IEC 19796-1 “Information technology for learning, education and training”, in particular, ISO/IEC 19796-1:2005 “Information technology. Training, education and preparation. Quality management, ensuring quality and metrics. Part 1. General approach” and ISO/IEC 19796-3:2009 “Information technology. Training, education and preparation. Quality management, ensuring quality and metrics. Part 3. Reference methods and metrics”. The authors describe the processes of creating e-learning course “Applied Informatics in Economics and Management” for masters with Moodle Cloud LMS.

Keywords: *e-Learning, cloud computing, Scrum-methodology.*

INTRODUCTION

With the rapid formation and development of the digital economy new requirements for employees of high-tech industries appears (Makarchuk, et al, 2017). These requirements determine the need for processing e-learning course in high school, taking into account the needs of users in obtaining high-quality and reliable information (Doychev, et al, 2016).

Many educational institutions with higher education programs use Learning Management Systems (LMS) for the organization of e-learning for increasing the availability of quality educational services (Stevens and Kitchenham, 2011).

Differences in the design of e-learning training courses are determined by aggregations of situational factors, such as IT infrastructure, didactic restrictions for training programs, a circle of interested persons. The heterogeneity of approaches to the organization of e-learning on the LMS platform causes difficulties in assessing the quality of e-learning. Improving the unified design methods of e-learning course with LMS will increase efficiency of quality management of e-learning in high school and reduce student barriers e-learning accessibility (Kumar and Owston, 2016.).

According to the concept of the third platform of the company IDC (<http://www.idc.com>), the world's leading provider of information and advisory services, and Information Technology development trends in 2013-2016 Gartner Inc. (<http://www.gartner.com/>), the basic requirements for the designing of e-learning's training course is determined by key strategies for the development of information technology in a digital society: mobile and cloud technologies, big data analysis methods and social networking technology. Special care when choosing technology for designing electronic courses with LMS need to pay to IT infrastructure and its effectiveness in supporting custom applications (Yélémou, et al, 2016). Using cloud infrastructure provides optimal affordable pricing package for educational organizations, in particular for coaches and students, for LMS launch (Veermanickam and Mohanapriya, 2016). Quality of e-learning course wins if it is accessible from mobile devices. Mobile learning has no limits in space and time (Kraut, 2013) and allows you to create special type of educational space conducive to communication and learning (Yépez-Reyes, V., 2018). Interface of on-line course in LMS in LMS should help e-learning participant to focus on learning, not on technology (Liliyana, et al, 2017). Analyzes student interactions with the learning management system during each learning activity will help predict student behavior so that provide the necessary support in a timely manner (Mothukuri, et al, 2017).

Project management at different stages of the e-learning course life cycle improves the quality of the e-learning course with the current trends: smart education, cloud computing, mobile apps, analytics, communication technologies.

1. MODEL OF THE LIFE CYCLE OF THE E-LEARNING COURSE

The main documents in the field of quality management of e-learning are quality standard for e-learning ISO/IEC 19796-1 “Information technology for learning, education and training”, in particular, ISO/IEC 19796-1 “Information technology for learning, education and training”, in particular, ISO/IEC 19796-1:2005 “Information technology. Training, education and preparation. Quality management, ensuring quality and metrics. Part 1. General approach” and ISO/IEC 19796-3:2009 “Information technology. Training, education and preparation. Quality management, ensuring quality and metrics. Part 3. Reference methods and metrics”.

The authors of the article consider the model of the life cycle of the e-learning course as a set of processes:

- Analysis of the need to design an e-learning course.
- Analysis of the structure of the e-learning course.
- Project development of the e-learning course.
- Development of the e-learning course.
- Implement the e-learning course.
- E-learning process.
- Evaluation of the e-learning course.

Throughout the entire life cycle of e-learning a flexible methodology Scrum Project Management has been applied for design, development, implementation and evaluation of e-learning course. Scrum methodology allows you to develop e-learning course of high quality in small groups of developers with minimal costs (Galván-Cruz, et al, 2017).

Moodle Cloud LMS (<https://moodlecloud.com>) was used as a Learning Management System, which is accessed by the model “Software as service” (SaaS) under the General Public License (GPU). Moodle Cloud LMS turned out to be the best alternative that combines the possibilities of open-source software available by subscription, accessi-

ble from student and teacher mobile devices As LMS alternatives were considered the best LMS platforms according to the Center learning technologies and improving productivity (Hart, 2016):

- Blackboard LMS;
- Docebo LMS;
- eFont LMS;
- Litmos LMS;
- Mirapolis;
- Moodle Cloud LNS;
- Torch LMS.

Building an e-learning life cycle model on the platform Moodle Cloud will help identify the strengths and weaknesses of this technology and the possibility of replicating the results of further work in the environment of e-learning.

2. DEVELOPING THE E-LEARNING COURSE WITH MOODLE CLOUD LMS

At St. Petersburg State University of Economics (UNECON) uses Moodle 3KL learning management system (de.unecon.ru) for implementation of e-learning. Most of the training courses offered at Department of Informatics, made in LMS Moodle and presented on the portal de.unecon.ru. Using the Moodle 3KL Learning Management System does not replace face-to-face full-time training and offers additional support of applied processes: management of educational content, document management, learning management.

At the first stage of the e-learning course life cycle model “Analysis of the need to design an e-learning course” were considered subprocesses: initiation, stakeholder identification, target setting and demand analysis, presented in Table 1 on the example of the organization of training master program “Applied Informatics in Economics and Management” UNECON.

Table 1

**Analysis of the need to design the e-learning course
“Applied Informatics in Economics and Management” UNECON**

№	Subprocess	Subprocess description
1	Initialization	The head of the master's program “Applied informatics in economics and management”
2	Stakeholders identification	Supervisors of Undergraduates Undergraduates External partners (employers, partner universities, etc.) Third-party students (applicants, student interns and etc.)
3	Goals Definition	Learning Management: - management of educational content - communications management - activity analysis - document management Promotion of the master's program
4	Demand analysis	Undergraduates – 50 % The head of the master's program – 20 % Supervisors of Undergraduates – 15% Third-party students –10 % External partners – 5 %

According to Table 1, the proposed e-learning course for as a goal, it does not directly involve economic benefits, but regarded as an effective tool for managing Master’s program and its popularization.

At the second stage of the e-learning course life cycle model «Analysis of the structure of the e-learning course», the requirements for an on-line course are studied according to the main educational program and other regulatory documents.

For the training course "Applied Informatics in Economics and Management" the course structure was created according to the following standards:

1. Federal State Educational Standard of Higher Education in Russian Federation Applied Informatics (Master's level);

2. Professional standard of the Ministry of Labor and Social Protection of the Russian Federation “Information Technology Manager”;

3. Chief Information Officer (C.I.O) Competency Requirements for Management information processes, systems and services.

In the structure of the course “Applied Informatics in Economics and Management” was proposed “By section” structure and the sections of the course were created: “News”, “Informational”, “Training modules”, “Scientific activity”, “Practices”, “Achievements of the program”.

The third stage of the e-learning course life cycle model “Project development of the e-learning course” is conceptual and can be implemented with using the flexible SCRUM methodology, the advantages of which are: iteration, minimization of risks, transparency of development, focus on end user, tight communication of project participants and their effective interaction (Faniran, et al, 2017.).

The project for the development of the training course “Applied Informatics in Economics and management” in LMS Moodle Cloud has been implemented in accordance with the main provisions of the SCRUM. According to this flexible methodology the following roles were approved: product owner, scrum master, development team, for each of which are defined relevant activities of roles: consulting, coordination, execution of works. As product owners were lecturers (professors and supervisors of undergraduates), as a scrum master was the head of the master's program. Development team was consisting of undergraduates “Applied Informatics in Economics and management”. Task list (sprints) was formed for a certain period, as a rule, constituting one to two weeks. In the framework of the project in question, the sprint was determined by weekly range, which is appropriate for the organization of the work of the teacher-student. Before each sprint for the future period determined the requirements, goals, objectives and the result to be obtained. Based on the analysis of the results for the period the team carried out either an iteration with the correction of defects, or execution next sprint.

At the stage “Project development of the e-learning course” special attention was paid to tool selection - learning management platform. Main criteria for choosing an LMS platform in a digital economy can be considered:

- support of mobile learning as one of the most popular forms the implementation of e-learning, available in time and space;
- access to the application on the SaaS model “application as a service”;
- the ability to view and edit content in various formats, including text, audio, video;
- supporting of various forms of social communication;
- data analytics in e-learning, incl. KPI Methodology Support (key performance indicators) and Performance Management.

At St. Petersburg State University of Economics an additional criterion of choice of the LMS platform is software with open source code.

To develop the training course “Applied Informatics in Economics and management” was chosen LMS Moodle Cloud, as the most popular among e-learning platforms available on the SaaS model and distributed under the General Public License (GPU). As of January 2017, according to GPU LMS Moodle Cloud license provides the ability to register 50 users and allocated 200 MB of free space, which can be increased using “personal clouds” (Makarchuk, 2017). On platform can accommodate an unlimited number of courses, there is supporting of mobile version.

At the stage of the e-learning course life cycle model “Development of the e-learning course”, the result is the finished product, in the project considered in the article - training course “Applied Informatics in Economics and Management”. At the development stage of the e-learning course, special attention was paid to user interface. When designing a digital educational environment, it is important to keep in mind that users from different cultures may have different expectations when interacting with the LMS interface. Limitation of functionality in Web platform design settings can complicate the interaction between LMS user and user interface (Bournel-Bosson and Kostov, 2018). In LMS Moodle Cloud user interface settings can have a different view for users with different roles on the site - guest, student, assistant, course developer and administrator. In this project the user interface was configured separately for users of each role with desktops or laptops and mobile users.

The subprocesses of the development stage of the e-learning course are presented in Table 2.

**Description of the process “Development of e-learning course”
on the example of a training course for undergraduates
“Applied Informatics in Economics and Management”**

№	Subprocess	Subprocess description
1	Content implementation	Master program lecturers (process owners): development of course content within a single structure
2	Project implementation	Lecturer – Project manager (SCRUM master) Undergraduates of the program "Applied Informatics" in economics and management", UNECON (performers)
3	Media implementation	LMS Moodle Cloud
4	Technical implementation	Undergraduates of the program "Applied Informatics" in economics and management", UNECON (performers) based on the tasks of weekly sprints in the environment LMS Moodle Cloud
5	Maintenance	It was carried out throughout this and subsequent stages by the development team

Intermediate results of E-Learning Quality Assessment Master's program “Applied Informatics in Economics and Management” (UNECON) on the model of the life cycle of e-learning showed an increase in the number 15% of students of the training course by adding guest access, time content editing reduced by 40% according to data in LMS Moodle 3K systems UNECON and Moodle Cloud LMS.

CONCLUSION

The proposed model of the life cycle of e-learning is based on International Standards ISO/IEC 19796-1 “Information technology for learning, education and training”, ISO/IEC 19796-1:2005 “Information technology. Training, education and preparation. Quality management, ensuring quality and metrics. Part 1. General approach” and ISO/IEC 19796-3:2009 “Information technology. Training, education and preparation. Quality management, ensuring quality and metrics. Part 3. Refer-

ence methods and metrics” on the example of e-learning of the master's program “Applied Informatics in Economics and Management”, UNECON with using a flexible SCRUM design methodology and learning management platform Moodle Cloud, available on the SaaS model and the GPU license, contributes to development of the theory in the field of research development of training courses in the system LMS. Intermediate results of building a training course on the model of the life cycle of e-learning has shown improving the quality of learning.

REFERENCES

1. BOURNEL-BOSSON, C.-Y. and KOSTOV, J. (2018) Inspiring self-confidence in French for academic purposes through the use of Moodle. *Recherche et Pratiques Pedagogiques en Langues de Specialit.* 37 (1).
2. DOYCHEV, E. et al (2016). *Agent-Based Support of a Virtual eLearning Space*. International Conference on Computational Collective Intelligence, ICCCI 2016: Halkidiki; Greece. pp.35-44.
3. FANIRAN, V.T. et al (2017). *Adopting Scrum as an Agile approach in distributed software development: A review of literature*. 1st International Conference on Next Generation Computing Applications, NextComp 2017: Mauritius; Mauritius. pp. 36-40
4. GALVÁN-CRUZ, S. et al (2017) *A Means-ends design of SCRUM+: An agile-disciplined balanced SCRUM enhanced with the ISO/IEC 29110 Standard*. 6th International Conference on Software Process Improvement: Zacatecas, Mexico. pp.13-23.
5. HART, J. (2016) *Modern Workplace Learning. Centre for Learning & Performants*. Technologies Publishers, USA.
6. KRAUT, R. et al (2013) *Mobile learning for teachers: Global themes*. UNESCO Publishers, Paris, France.
7. KUMAR, K. and OWSTON, R. (2016) Evaluating e-learning accessibility by automated and student-centered methods. *Journal of educational technology research and development*. 64(2). pp. 263-283.
8. LILYANA, F.N. et al (2017) *Techno MOOC development of the basis of remote lab access*. International Conference "Quality Management, Transport and Information Security, Information Technologies". St. Petersburg, Russian Federation. pp. 680-684.

9. MAKARCHUK, T. (2017) *Mobile learning on the basis of the cloud services*. International conference on e-learning, EL 2017. Lisbon, Portugal. pp. 175-178.
10. MAKARCHUK, T. A., MINAKOV, V. F., TROFIMOV, V.V. et al (2018) Quality management of e-learning in information technology management training. International conference on research paradigms transformation in social sciences: Irkutsk, Russia. pp. 742-748.
11. MOTHUKURI, U.K. et al (2017). *Improvisation of learning experience using learning analytics in eLearning*. 5th National Conference on E-Learning and E-Learning Technologies.
12. RUTH, M. and KASPAR, K. (2017). The E-Learning Setting Circle: First Steps Toward Theory Development in E-Learning Research. *Electronic Journal of e-Learning* (EJEL). 15(1). pp. 94 - 103.
13. STEVENS, D. and KITCHENHAM, A. (2011). *An analysis of mobile learning in education, business and medicine*. IGI Global Publishers, Hershey, USA.
14. VEERAMANICKAM, M.R.M. and MOHANAPRIYA, M. (2016) *Research paper on E-Learning application design features: Using cloud computing & software engineering approach*. International Conference on Information Communication and Embedded Systems, ICICES 2016: Chennai, Tamilnadu, India.
15. YÉLÉMOU, T. et al (2016). *Eliminate the delay backlog in the conduct of pedagogical activities by distance learning*. 8th International Conference on e-Infrastructure and e-Services for Developing Countries, AFRICOMM 2016: Ouagadougou, Burkina Faso. pp. 273-277.
16. YÉPEZ-REYES, V. (2018). *Mobile learning: Challenging the current educational model of communication studies*. Journal of Advances in Intelligent Systems and Computing. 721. pp. 1014-1021.

EXPLORING THE FUNCTIONALITY OF BULGARIA EGOVERNMENT SITE

Vanya Lazarova¹

*¹ University of National and World Economy/ITC, Sofia, Bulgaria,
vlazarova@unwe.bg*

Abstract

The object of the paper is the portal for electronic administrative services (eGov.bg), all the functions provided by it and study its functionality. The eGovernment site features offered must be examined in depth and all working and non-working functions highlighted. One of the most important tasks at the end is to derive and show the ratio of working and non-working functions and to show which of them have some activity and which have not. This information would be presented in tabular form and diagrams.

Keywords: *functionality, Bulgaria eGovernment, site.*

INTRODUCTION

In the last decade, the global economy has become noticeably digital, and the information and communication technology sector is steadily entering all areas of social and economic life. The expectations of citizens and businesses towards public institutions are increasing, related to providing more public control over their activities, improving the quality of service provided and ensuring a higher standard of living. Achieving a balance between expectations and opportunities is a daunting task that requires mobilizing the efforts and resources of the administration and society as a whole.

EGovernment represents the use of information and communication technologies to present administrative services to citizens who need them.

The definition of EGovernment in Bulgaria is given in document E-gov Strategy: "EGovernment is the management in the electronic environment of regulatory interconnections, administrative processes and services and interaction with users using information, statistical and mathematical models and methods of data processing, information and

knowledge that provide much higher level of management effectiveness. E-Governance is a tool for comprehensive increase of the efficiency of the processes in the administration, as well as facilitating the processes in the interaction between the administration, employees, citizens, business, through the use of e-Services”.

The object of this paper is the portal for electronic administrative services (eGov.bg), all the functions provided by it and study its functionality.

The eGovernment site features offered must be examined in depth and all working and non-working functions highlighted. One of the most important tasks at the end is to derive and show the ratio of working and non-working functions and to show which of them have some activity and which have not. This information would be presented in tabular form and diagrams.

1. EXPLORING THE FUNCTIONALITY OF BULGARIA EGOVERNMENT SITE

The Bulgarian portal for electronic administrative services (egov.bg) offers a number of features, but not all of them work to their full capacity, and behind some of them there is no activity at all. Thus, Bulgaria finds itself with a functioning e-government, but not in its full capacity.

1. Business and Entrepreneurs Menu

The Business and Entrepreneurs menu includes submenus providing information in that field, such as banking and financial services, taxes, investment projects, the ability to apply for registrations, permits and other procedures.

Table 1

Business and Entrepreneurs Menu

Functions	YES	NO
Banking, financial and investment services		✗
Taxes and social security payments		✗
Energy		✗

Health and medical insurance		✗
Investment projects and concessions	✓	
Customs regimes and permits. Import, export, forwarding	✓	
Bankruptcy and liquidation		✗
Patents and copyrights	✓	
Production and marketing	✓	
Free professions and crafts. Self-sufficient	✓	
Construction and commissioning of sites	✓	
Tourism, hotel and restaurant business	✓	
Business certification and reference services	✓	
Total (%)	62%	

2. Civil Rights Menu

The Civil Rights menu contains information and e-services for civil rights certification, legal proceedings, and deprivation of rights.

Table 2

Civil Rights Menu

Functions	YES	NO
Restriction of civil rights		✗
Witnesses and parties to the case		✗
Bailiffs		✗
Certification of property and property rights	✓	
Certification and exercise of civil rights		✗
Accessibility, publicity and transparency	✓	
Total (%)	33%	

3. Civil Status Menu

The Civic Status menu contains information and e-services related to citizenship, address registrations and identity, birth, marriage and inheritance documents.

Table 3

Civil Status Menu

Functions	YES	NO
Bulgarian citizenship		✕ *
Pets		✕
Name, address registration, identity documents	✓	
Corrections, endorsements, certificates and references concerning civil status acts	✓*	
Birth, adoption, foster care, guardianship and guardianship	✓*	
Family - marriage, divorce	✓	
Death and Inheritance	✓*	
Total (%)	71%	

Functions highlighted in Table 3. with ✓*, provided e-services that are accessible to 4 municipalities, but we believe that the process could easily be speeded up in the next months and other municipalities will be covered by the e-service.

E-function highlighted with ✕*, are accessible only for Pesteria municipality, so it could not be marked as working.

4. Citizens' Taxes and Fees Menu

The Citizens' Taxes and Fees menu contains information and e-services related to taxes and fees, social and health insurance.

Table 4

Citizens' Taxes and Fees Menu

Functions	YES	NO
Tax Refund and Tax Relief	✓	
Tax and social security calculators		✕
Taxes and social security payments	✓	
Local taxes and fees	✓	
Total (%)	75%	

5. Health and Health Insurance Menu

The Health and Health Insurance menu includes information and e-services related to citizens' social security rights and health insurance.

Table 5

Health and Health Insurance Menu

Functions	YES	NO
Health insurance	✓	
Tax and social security calculators	✓	
Total (%)	100%	

6. Property, Housing and Utilities Menu

The Property, Housing and Utilities menu includes information and e-services related to taxes, fees, utilities, construction and conversion, real estate.

Table 6

Property, Housing and Utilities Menu

Functions	YES	NO
Improvement		✗
Taxes and fees	✓*	
State and municipal property	✓*	
Condominium		✗
Utilities - plumbing, heating, electricity, gas, telecommunications		✗
Renting and letting of real estate		✗
Acquisition, sale and mortgage of real estate		✗
Construction and reconstruction		✗
Total (%)	25%	

Functions highlighted in Table 6. with ✓*, provided e-services that are accessible to few municipalities, but we believe that the process could easily be speeded up in the next months and other municipalities will be covered by the e-service.

7. Education and Qualification Menu

The Education and Qualifications menu provides information and electronic services related to kindergartens, schools, universities, students and students, as well as vocational guidance and qualifications.

Table 7

Education and Qualification Menu

Functions	YES	NO
Kindergartens, nurseries and pre-school education		✗
Children, students and students with disabilities		✗
Colleges and Universities		✗
Vocational guidance and qualification, adult education	✓	
Schools		✗
Teachers and teachers		✗
Total (%)	17%	

8. Environment Menu

The Environment menu includes information and electronic services related to energy sources, landscaping, water and various programs.

Table 8

Environment Menu

Functions	YES	NO
Alternative energy sources		✗
Household and construction waste		✗
Water		✗
Protected Areas		✗
Pollution control		✗
Vegetation (Green system, Landscaping)	✓	
Total (%)	17%	

9. Agriculture and Forests Menu

The Agriculture and Forests menu contains information and electronic services related to programs, registrations and permits in this field; veterinary medicine, forests and forestry, animal husbandry, agriculture, tourism and others.

Table 9

Agriculture and Forests Menu

Functions	YES	NO
Veterinary medicine		✗
Forests and logging	✓	
Other (herbs)		✗
Other (bees)	✓	
Livestock breeding		✗
Agriculture		✗
Hunting and fishing		✗
Growing fish and other aquatic crops		✗
Entrepreneurship and Tourism	✓	
Agricultural machinery		✗
Total (%)	30%	

10. Social Activities Menu

The menu "Social activities" contains information and electronic services related to social assistance, applications and registration of job seekers, children at risk, employment and pension insurance, entry in a public register.

Table 10

Social Activities Menu

Functions	YES	NO
Unemployment and employment promotion		✗
Children at risk		✗

Labor protection		✕
Social assistance	✓	
Labor and pension insurance	✓	
Disadvantaged people	✓	
Total (%)	50%	

11. Transport and Vehicles Menu

The Transport and Vehicles menu contains information and electronic services related to this field, such as legal qualifications and professional qualifications in the field of transport, acquisition, registration and tax of vehicles, transport infrastructure and more.

Table 11

Transport and Vehicles Menu

Functions	YES	NO
Competence and professional qualifications in the field of transport	✓	
Carriage of passengers	✓	
Freight	✓	
Acquisition, registration and tax of vehicles	✓	
Transport infrastructure	✓	
Total (%)	100%	

12. Spatial Planning, Construction and Cadastre Menu

The "Spatial Planning, Construction and Cadastre" menu contains information and e-services related to cadastral activities, investment design and construction, technical infrastructure, surveying and many more.

Spatial Planning, Construction and Cadastre Menu

Functions	YES	NO
Property rights. Alienation, compensation	✓	
Investment design and construction	✓*	
Cadastral activities	✓	
Purpose of the territories, development, reconstruction	✓*	
Provision of data from GEOCARTFUND	✓	
Professionals - Geodesy, Cartography and Cadastre	✓	
Technical infrastructure	✓*	
Contingency plans - creation, modifications	✓*	
TOTAL (%):	100%	

Function marked with ✓* includes services available only to 3 or fewer municipalities. Since we are monitoring the site, in just a few months the development of this functionality is very fast. The problem is that e-service consist of fulfilling the online form, which, in most cases, are received in paper at the proper desk. To accelerate site efficiency (Михова, Веска, 2012) all paper documents must be replaced with electronic ones.

After a study of the functionality of the menus of Bulgaria Egovernment Site (eGov.bg) it was found that at the end of august 2019 year, the total percentage of working function is 53%, and non-working functions is 47%. Figure 1 represents a summary chart of the working functions and figure 2 represents the ratio of working and non-working functions in eGov.bg.

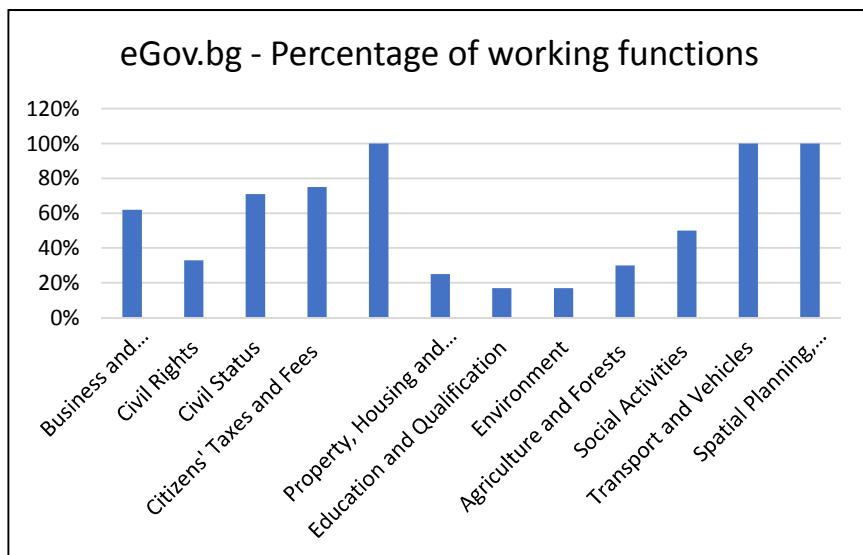


Figure 1. Working functions in "eGov.bg"

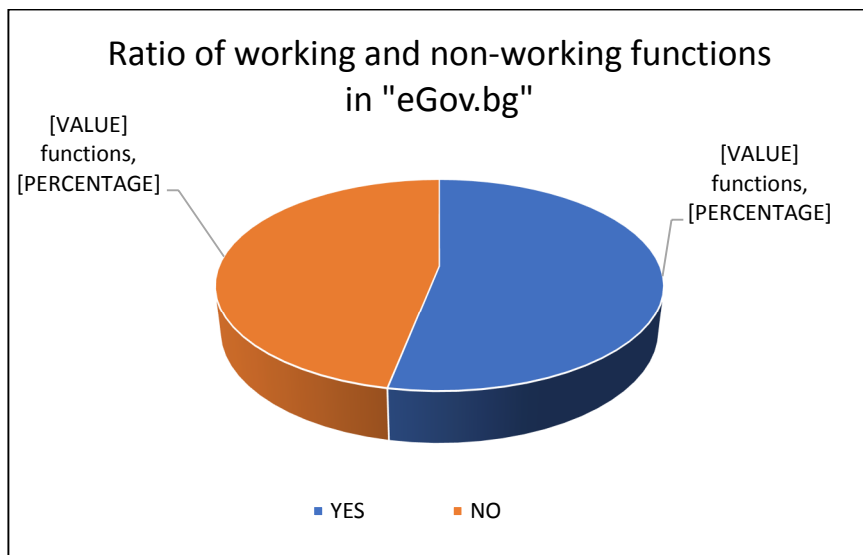


Figure 2. Ratio of working and non-working functions in "eGov.bg"

CONCLUSION

The study concludes that Bulgaria currently has an established e-government, which does not fully function and does not cover all the services offered by the administration. Many of the services provided by the eGov.bg are accessible only to a few municipalities. A significant percentage of the available e-services is a fulfilling of the online applications, which, in most cases, are received at the counter at the relevant administration. This impedes the complete transformation of paper-based services into electronic ones.

As the system always follows the business (Tsaneva, Monika, 2019) the future website development depends on the needs of business and citizens.

REFERENCES

1. Стратегия за развитие на електронното управление в Република България 2014-2020г., <http://www.strategy.bg/StrategicDocuments/View.aspx?Id=892> [Accessed 28/08/2019].
2. Стратегия за развитие на електронното управление в Република България 2019 – 2023. <https://www.strategy.bg/PublicConsultations/View.aspx?lang=bg-BG&Id=4223> [Accessed 28/08/2019]
3. МИХОВА, Веска. 2012. Мониторинг и анализ на бъдеща производителност на бизнес приложения. *Статистика, информационни технологии и комуникации: Юбилейна научна конференция*. - София: Издателски комплекс-УНСС, 2012, 496-503.
4. TSANEVA, Monika. 2019. A Practical Approach for Integrating Heterogeneous Systems. *Business Management*, issue2/2019, p.13.

BIG DATA AND INFORMATION SECURITY

Veselin Popov¹, Petya Emilova²

*¹ D. A. Tsenov Academy of Economics, Svishtov, Bulgaria,
v.popov@uni-svishtov.bg*

*² D. A. Tsenov Academy of Economics, Svishtov, Bulgaria,
petiaem@uni-svishtov.bg*

Abstract

Along with the many opportunities that Big Data provides, it brings many threats to society and businesses. Big Data Analytics in security refers to the ability to gather massive amounts of digital information to detect, analyse and visualize insights that enable the prediction and counteraction of cyberattacks. Along with security technologies, organizations are given the opportunity to identify activity schemes that represent network and other types of threats, which is of great practical importance.

The scope of paper is big data, its sources, features, protection, platforms and processing tools, analyses the abilities of big data for cyber security and define some challenges that refer to major Big Data security issues. The methodology of the study involves a research of the literature on the topic.

Keywords: *Big Data, Big Data Analytics, Cyber Security, Information Security.*

INTRODUCTION

Information security issues are among the leading contemporary business problems. Technologies such as Big Data are linked to information security – on the one hand because it needs serious protection and on the other hand because it itself provides new protecting capabilities in terms of analytics and security solutions to protect data and prevent future cyberattacks.

Big Data Analytics (BDA) refers to the strategy of analysing large volumes of data, or big data (TECHNOPEDIA, 2019). It has become an effective tool in many areas of modern information processing, including the area of information security. By providing tools for collecting and analysing large amounts of digital information generated from different sources or logged by different devices, BDA helps detect and define patterns and trends of irregular behaviour, search for methods to track cybercriminals, predict and stop potential cyber-attacks.

This paper focuses on the problems and challenges of protecting Big Data and on Big Data's ability to improve information security.

1. BIG DATA – SOURCES, FEATURES, PROTECTION, PLATFORMS AND PROCESSING TOOLS

In the specialized literature, the term "Big Data" refers to a huge amount of information that is stored, transmitted and processed in computer systems. Today, *the main sources* for generating Big Data are social medias, a huge variety of smart devices, videos, digital images, sensors, and sales transaction records (TECHNOPEDIA, 2019). In addition to their large volume, these data are also complex which is why traditional processing applications cannot handle them.

According to the National Institute of Standards and Technology (NIST, 2015), Big data is data where the volume, required speed for processing, or the manner in which the data is presented limit the possibilities of the data being analysed effectively by traditional relational approaches or the data require significant horizontal scaling to ensure efficient processing. This has created the need for a new generation of software applications and tools.

Big Data differs from traditional databases by four main characteristics related to: the *volume* of data; *velocity* of data generation and transmission; *variety* of data in the form of different types of structured and unstructured data; *complexity* related to the structure, behaviour, and permutation of data sets in the presence of critical factors.

One of the biggest challenges for Big Data platforms is the protection that must be provided at every stage of the platform's lifecycle, using a combination of *traditional* security tools, *new* tools and technologies, as well as *intelligent* security monitoring processes.

The Big Data life cycle includes four stages:

- a) organizing the data,
- b) processing,
- c) data storage,
- d) processing and production.

Organizing data involves the creation and / or access to the data. Big Data is a collection of different types of data that come from different sources. According to sources, Big Data is of three types: human-

generated data - text, photos, video; machine-generated data - computers operating documents, databases, multimedia, GPS, RFID, smart homes; and data generated by objects (various digital devices).

For example, user-generated data includes CRM data, data such as email messages, phones, SMS or social media posts, and more. There are many transaction data and data stored in different databases. There is also a huge volume of data generated from custom software, sensors and sensors. Data transmitted from sources to the platform must be protected.

Data storage. Different tools such as encryption, user authentication, and intrusion protection are used for protection. These security tools must operate in a distributed platform environment with many servers and branch nodes. Security tools must protect log files and analytics tools when they run on the platform.

Many organizations use cloud services to store their data. Cloud technologies also allow companies to use pre-built Big Data solutions or to quickly build or deploy powerful server complexes without the significant cost of acquiring hardware. In addition to the benefits, however, the new technologies pose challenges to security.

The analytical processing (analysis and data outcomes) of different types of data is the essence of Big Data. The results obtained from this processing are directed to applications, reports, dashboards and are goal for attacks. Therefore, the encryption of results, control of access and traffic are crucial.

The classic technologies used to protect Big Data are: encryption, user access control, intrusion detection and countering, physical security. Encryption protects data both during storage and transition. It should be noted that encryption must work with different types of data and with different analytical tools, relational and non-relational databases, special file systems, etc. User access control is a major network-level security tool. The control is important for the Big Data platform, it must be very precise and based on well-defined policies. Intrusion detection and countermeasure systems are also important to the Big Data platform as they contribute to the timely detection of intrusion attempts. Physical security is related to the security of the building premises of the organization's or cloud provider's data centre.

The Big Data Processing Software Market offers hundreds of software tools and integrated platforms. Some of them have a long history

while others have appeared recently. This market is dominated by large enterprise software vendors such as Oracle, IBM, SAS, SAP, Teradata, Microsoft, and more. Their Big data products are typically integrated into complex business management solutions.

Successful products are also available from smaller software companies specializing in Big data such as Tableaa, RapidMiner, Pentaho, Alteryx, Alpine and more.

Open source products like Apache Hadoop, MapReduce, GridGain, Storm and others are also available on the market. Some open source software solutions are often included in projects by commercial vendors.

The best software tools for Big Data processing, according to their purpose, functions performed and security level implemented, are (IMPORT.IO, 2018):

- for data storage and management - Hadoop, Cloudera, MongoDB, Talend;
- for data cleaning - OpenRefine, DataCleaner;
- to extract knowledge from data - RapidMiner, IBM SPSS Modeler, Oracle data mining, Teradata, FramedData, Kaggle;
- for data analysis - Qubole, BigML, Statwing;
- for data visualization - Tableau, Silk, CartoDB, Plot.ly, Datawrapper;
- for data integration - Blockspring, Pentaho;
- data languages - R, Python, RegEx, XPath;
- data Collection - Import.io.

2. BIG DATA ANALYTICS FOR CYBER SECURITY

Big data analysis is much more sophisticated than traditional database applications. Big data are large volumes, not aggregated, in different formats and their processing can hardly be performed in the memory of only one computer. Big data processing involves mechanical processes and algorithms. The Big Data analysis methods used are of two main types - responsive and predictive (CHAOVALITWONGSE, W., HUANG, S., 2015).

Reacting analysis aims to provide statistics on current and historical data and information about what happened and why it happened. It includes methods such as statistical modelling, trend reporting, visualization, association, and correlation analysis.

Predictive analysis focuses on the use of known data (training data) that includes input data properties (attributes) and response values (target models) to build a predictive model (solution) to make predictions for invisible data (test data). It uses methods such as vector machines, linear regression / classification, nonlinear regression (generalized linear model), decision tree, Bayes theory, neural networks, and more.

Big Data technologies used in security systems are able to detect threats in advance. For example, they can detect atypical web behaviour, predict an attack, and analyse the sources of an attack.

Big Data creates conditions for the efficient and effective application of some fraud detection techniques. In the specialized literature, they are divided into two main groups: *statistical techniques* and *techniques using artificial intelligence*. Table 1 provides examples of these techniques.

Table 1

Examples of data analytics techniques

Statistical technics
<ol style="list-style-type: none"> 1. Techniques used in pre-processing data to detect, validate, correct errors, fill in missing and inaccurate data. 2. Calculation of various statistical parameters such as averages values, performance indicators, probability distributions, etc. 3. Models and probability distributions of various business activities. 4. Processing user profiles. 5. Time series analysis of time-dependent data. 6. Clustering and classifying to discover possible patterns / schemas and dependencies / associations in the data groups. 7. Combining algorithms to detect anomalies in the behaviour of transactions or users
Artificial intelligence technics
<ol style="list-style-type: none"> 1. Data mining 2. Expert systems. 3. Automatic pattern recognition 4. Machine training techniques 5. Neural networks

Source: (BAJPAI, A., DAYANAND, and ARUSHI, A., 2018)

3. CHALLENGES

Big Data Working Group defines four aspects of Big data security which are: infrastructure security; data privacy; data management, integrity and reactive security (BIG DATA WORKING GROUP, 2013). Each area has many problems. For example, infrastructure security has the following issues:

- Single Layer Security - The companies need to integrate multi-layer defence within the company's defence strategy, which addresses both internal and external security threats.
- Data transfer across multiple devices, which requires additional levels of security and monitoring to ensure that data is not captured on the route from one device to another.
- A rapid development of Big Data technology and its supporting infrastructure (such as cloud services) that must be able to process data from an infinite number of points with speed, security and reliability. Infrastructure should include security measures to keep information in place at all stages of the process.

In addition to enhancing business intelligence, Big Data provides the opportunity to enhance cyber security. This, however, increases the existing problems and challenges that require attention and await solutions.

Major Big Data security issues are related to: Threats to data security; Privacy risks; Big Data credibility needs to be confirmed; Lack of privacy protection security technology for Big Data.

These Big Data security issues require many challenges which need to be tackled, the most important of which are:

a) Adopting protection as a top priority for Big Data platforms, which will focus the attention of managers and developers in this direction.

b) Introducing reactive and proactive protection.

c) The physical protection of devices and servers that contain sensitive information must be managed with particular care and isolated from other devices.

d) Application security is as important as device security. Data protection solutions are crucial in this regard. These solutions are at the

heart of Big Data platforms, they contribute to the discovery of behaviors and development trends, and on this basis offer business strategies. This importance of Data mining solutions requires that they be protected against both external threats and internal individuals who abuse their privilege to access sensitive information.

e) High level of access control to be provided with encrypted authentication and to determine who can see what data.

f) Use of real-time security tools. These tools generate a wealth of information. The problem here is to ignore the unimportant alerts so that employees are being targeted for real violations.

g) Data storage management. For Big Data, the architecture is typically to store data at multiple levels, depending on their importance to the business and the cost of storing it.

h) Performing a detailed audit that can help determine when missed attacks can occur at this time, what the consequences are, what to change in the current system.

i) Use of distributed systems. For faster analysis, most Big Data platforms actually distribute the huge amount of data processing work across many systems.

CONCLUSION

Along with the many opportunities Big Data provides to consumers, it creates conditions for multiple threats to society, business and individuals. These threats must be countered. At the same time, the BDA has emerged as an effective tool in many areas of modern information processing, including in the field of information security. This report addresses the problems and challenges of protecting Big Data in these two aspects.

REFERENCES

1. ARADAU, C., BLANKE, T. (2015) The (Big) Data-security assemblage: Knowledge and critique. *Big Data & Society*. 1-12, bds.sagepub.com.
2. BAJPAI, A., DAYANAND, and ARUSHI, A. (2018) Big Data Analytics in Cyber Security. *International Journal of Computer Sciences and Engineering*. 6 (7). pp. 731- 734.

3. BIG DATA ANALYTICS TECHNIQUES. Some Effective Techniques of fraud detection analytics. EDUCBA, [Online] Available from: <https://www.educba.com/fraud-detection-analytics/> [Accessed 2/08/2019].
4. CHAOVALITWONGSE, W., HUANG, S. (2015) Computational Optimization and Statistical Methods for Big Data Analytics: Applications in Neuroimaging. [Online] Available from: http://faculty.washington.edu/artchao/INFORMS_Tutorials_2015-Web.pdf [Accessed 12/08/2019].
5. FAN, Y. (2016) Big Data Security and Privacy Protection. *Electronic Technology and Software Engineering* (1): 227
6. IMPORT.IO (2018). All the best big data tools and how to use them. [Online] Available from: <https://www.import.io/post/all-the-best-big-data-tools-and-how-to-use-them/> [Accessed 21/08/2019].
7. NIST. (2015) Big Data Interoperability Framework. NIST [Online] Available from: <https://www.nist.gov/el/cyber-physical-systems/big-data-pwg>. [Accessed 2/08/2019].
8. TECHOPEDIA. Big Data Analytics. [Online] Available from: <https://www.techopedia.com/definition/28659/big-data-analytics>. [Accessed 2/08/2019].
9. ZENG, G. (2015) Big Data and Information security. *International Journal of Computational Engendering Research*. 5(6). pp. 17-21.
10. BIG DATA WORKING GROUP (2013). Expanded top Ten Big Data Security and Privacy Challenges. [Online] Available from: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf. [Accessed 2/08/2019].

APPROACHES FOR ENHANCING DIGITALIZATION AND DIGITAL TRANSFORMATION IN SUPPLY CHAIN MANAGEMENT

Luben Boyanov¹

*¹ University of National and World Economy/Department of
Information Technologies and Communications, Sofia, Bulgaria,
lboyanov@unwe.bg*

Abstract

The paper reviews the importance of digitalization and digital transformation in supply chain management. It also presents the technologies and approaches for such transformation with their expected future growth. Business processes in supply chain management, which can be digitalized are identified and the priorities of such transformation are presented. Good practices are also presented and the effects of digital transformation are given. Other important factors in the process are also presented.

Keywords: *digitalization, digital transformation, supply chain management.*

INTRODUCTION

The world has entered the Digital Age (known also as Computer Age, Information Age, etc.) following the Digital (Information) revolution, which marked the next major human transformation after the Agricultural and the Industrial Revolution. They all converted drastically human society and the entire world. Just like the cultivation of grain and invention of machines and manufacture changed history, digital computers and communication technologies reformed almost every aspect of people's life in the last decade. People are going through drastic modernisation in the way they work and live. The invention, development and advance of high-tech electronics – computers and their associated devices, their miniaturisation and wide spread, in parallel with the development of Internet and mobile technologies, shaped a completely different world, existing only in science fiction 40 years ago. Everything is getting digital, everything transforms in a digital form. And while we all know well about the success of digital companies like Google, Facebook, Am-

azon, Netflix, etc., the challenge of transforming the rest, the more traditional companies has emerged as the most important task for them. This paper reviews and examines the modern technologies and approaches for such transformation in the field of Supply Chain Management (SCM), but most of them can be applied in any other human activities and field.

1. NEW AND EMERGING DIGITAL TECHNOLOGIES

There are a lot of new and promising digital technologies that have emerged together with the fusion of the micronization and nanoization of products and devices and their connection to personal, local, metropolitan, global, mobile, private and public networks. All this is forming new paradigms and approaches in every and each direction for man. We shall only list the most important of them, as they fit in the activities and goals of SCM.

Internet of Things (IoT) is the buzzword for the last two decades and one of the priority areas for almost every IT company and business in the world. Its applications are everywhere. Google, Amazon, Microsoft, Intel, IBM, Oracle, Cisco, AT&T, Huawei, etc. have already implemented and developed platforms for IoT, which can be used for any business and application (Computerworld, 2019). IoT “makes” anything “smart”- smart TV, smart watch, smart industry, smart transport, smart logistics, smart energy, smart cities, smart homes, etc. etc. IoT is in fact a set of processes, objects and all things of our surrounding, which are digitally connected, allowing people to monitor and collect data on those objects/processes and to add feedback on them. Capture, collect, transmit and manage data from and to any “thing” forms the IoT. Applying IoT in SCM requires integration of data and their collection from different types of production, transport, logistics, customers, suppliers, etc. Despite the advantages of providing better efficiency, restructuring the entire chain in a consistent and proper way is no easy task. Very few operations have experience in such data collection and management – the airspace fleet and some of the military production.

A direct consequence but not actually related to IoT in origin is **Big Data**. This is an area of investigating and developing methods and approaches for fetching, processing and analysing of data, that are too big and complex to be processed in the traditional way. Some of this data can

be a source of unknown dependencies and relations, or to point out unknown until now problems and errors, or to lead to discovery of new materials, drugs or inventions. The challenges in Big Data are data collection, storage, analysis, search, sharing, transfer, visualization, queries, update, security and variety of sources. Big data is associated with three concepts - volume, variety and velocity, to which some authors add two more - veracity and value, giving the five V (Cano, 2014). Supply chain management has some of the typical Big Data areas of application – industry, transport and logistics, real time analytics, customer analytics, etc.

Artificial Intelligence (AI) is sometimes called machine intelligence, as it is in fact a program, executed on digital machine. Its main feature is the presence of human like intellect with the ability to learn, in a way similar to people. Intelligent objects can watch and analyse the environment (or data) and act in such a way that a predefined objective can be achieved. AI mimics the cognitive functions of humans and their interaction with the environment. Learning and problem solving are major functions of AI. This technology has huge potential in the area of SCM – it can reshape business processes and ecosystems, creating new areas of production and new types of clients. The decision making prospective of AI will allow planning in the entire chain of the supply management, where by now, all decisions on different scenarios and big data are taken by human. The rise of efficiency will be huge.

Two other important technologies with serious chances to become widespread are **Virtual Reality** (VR) and **Augmented Reality** (AR). Virtual reality is an interactive computer technology that creates simulated (virtual) reality to its users, where they participate and act. It comprises of audio and video environment, in which users actively participate – typically they are connected to the VR via sensors, and a possible feedback can be asserted. VR can be close to some realistic environment but can be also a product of humans' imagination. Augmented reality is also interactive computer technology, where the object from the real world are augmented with computer generated environment, where audio, video, smell or other senses can be provided to the human. The sensor information can be constructive (added on the natural/real environment) or destructive (to be absent from the real environment) resulting in a new, integral reality for the user. Augmented reality is also associated with the

terms “mixed reality” and “computer-mediated reality”, which often refer to similar and closely related concepts.

Cloud services are modern IT technologies, which unlike those listed above have already passed the age of innocence and became part of our everyday digital experience. They allow the use of computer resources like memory (for data and programs) and CPU time/usage (for execution of various programs) where the end user has no direct access to the resources – as if they come from an unspecified/unshaped cloud. Cloud services are part of modern IT life and both companies and private users are their clients. Cloud systems are big data centres with powerful computer systems (and programs running on them), offering computer resources as servers, storage, programs and networking via the pay-as-you-go model. Clouds can be owned by a certain company or organization and provided limited access to it (private clouds), or can be accessed to many organizations and individuals (public clouds). Amazon Web Services (AWS) is regarded the biggest and most popular cloud, but Microsoft Azure and Google Cloud Platform have also their notable place in the cloud community.

The concept of resource sharing for getting unity and coherence, avoiding creation and maintenance of IT infrastructure is very convenient for a variety of users, and in this respect the area of SCM with its variety and diversity of services and function is a proper candidate and user of the technology. Clouds can support the digital transformation in SCM, synchronising and unifying the resources and users within the entire supply chain.

Blockchain technology is another modern buzzword, that was born of the bitcoin phenomenon. It consists of ever-growing list records, called blocks, that are linked to each other (as a chain) by cryptography. Each block contains hash from the previous block, the time of its creation and transaction data. According to that concept and the approach used, the blockchain is resilient to change of the data in it. As defined in Nakamoto, 2008, it presents a distributed ledger technology, where transactions between parties can be recorded efficiently in a verifiable, permanent and transparent way.

A number of attempts exist by industrial organizations to introduce blockchain in SCM. The alliance Blockchain in Transport Alliance

(BiTA) works for the creation and development of SCM open standards (BiTA, 2017). IMB and Walmart use blockchain systems for control and monitoring of SCM – the nodes are administrated by Walmart and reside in IMB’s cloud (Corkery and Popper, 2018). Some distributed SCM can use smart contracts, which also have important functions in verification and enforcing the execution of a contract. Table 1 summarizes advantages and disadvantages with the expected future market share of those technologies.

Table 1

Advantages and disadvantages of modern digital transformation technologies and their expected market share according to
<https://www.statista.com/statistics> and
<https://www.globenewswire.com>

№	Technologies	Advantages	Disadvantages
1	<i>Internet of Things</i> Expected market size in 2025 – 1.6 trillion USD	Makes anything smart – can monitor, capture, collect, transmit data and manage/feedback things	Creates Big data, sometimes hard to control and manages, difficult integration, no standards
2	<i>Big Data</i> Expected market size in 2027 - 103 billion USD	Stores, analyzes, manages, processes data in very large volumes, in large varieties and fast speed	No standard and no universal and accepted tools for work with Big Data
3	<i>Artificial Intelligence</i> Expected market size in 2025 - 118.6 billion USD	Analyze big data, take decisions – strategic or in real time	No universally programs and approaches for different application areas
4	<i>Virtual and Augmented Reality</i> Expected market size in 2025 - 814.7 billion USD	Removes the necessity of testing and modeling real processes and objects	No widespread and unified tools for various applications
5	<i>Cloud services</i> Expected market size for public clouds in 2022 -331.2 billion USD	Use of unified and coherent computer resources without creation and maintenance of IT infrastructure	May have some issues to overcome for SCM applications

6	<i>Blockchain</i> Expected market size in 2023 - 23.3 billion USD	Provides reliable and resilient tracking of in- formation/data	Lack of experience in working with it
---	--	--	--

2. BUSINESS PROCESSES AND DIGITALIZATION

Digitalization and digital transformation present opportunities for improvement of the effectiveness and efficiency in SCM. It is important to list some of the main business processes in the chain. The Supply Chain Management consists of the following key business processes: Customer Relationship Management, Supplier Relationship Management, Customer Service Management, Demand Management, The Order Fulfilment, Manufacturing Flow Management, Product Development and Commercialisation, and Returns Management (Croxtton and Rogers, 2002).

The first business process in the list is the ***Customer Relationship Management*** (CRM). It controls the interaction with present and future clients and helps the automation and integration within the framework of an enterprise on activities like sales, marketing and support. CRM defines the framework for the development of the relationship with customers. It can become a valuable practice to identify individual clients of groups of clients, that have a similar quantitative value in time and their loyalty can be reaffirmed by offering personalized products and services. CRM can be further enhanced by teams, offering agreements on products and services (Product and Service Agreements - PSA), that can satisfy the needs of key players or areas, as well as other clients' needs. The teams will pursue improvement of processes and elimination of non-profit activities. Digital transformation will allow companies to collect, support and process big, client-oriented data bases, that will help enhancing the profit and turnover.

Next comes the ***Supplier Relationship Management*** (SRM). This process defines the way of interaction between the company and its suppliers. It is important the organization to establish close cooperation with some (or all) of its suppliers. This involves the formation of a solid grounds and signing agreements for products and services (PSA), so that

the company and its suppliers can have mutual benefit of the best trade practices.

Another key process is the ***Customer Service Management***. It is an important step towards the administration of agreements for products and services, providing information for orders to the clients, as well as timing for delivery and access to products. Data for this process comes from the production and logistics management.

The forth process is the ***Demand Management***. It allows the company to be active in the process of balancing the demand and supply. This process envisages prediction and synchronization of demand and supply in order to improve the choices and reduce the variations in demand. The process should include information and even intelligence on the client, historic sales and planned marketing, prediction efforts and demand impact.

Another important process is the ***Order Fulfilment***. It is more than just an order of goods or service. It includes all necessary activities for determining the client's wishes and designing a process, that allows the company to satisfy client's demands, minimizing the total value of deliveries at the same time. This is not only logistic function but also inter-sector activity in line with the operation of the main suppliers and the clients. The objective is to achieve a seamless process from the supplier to the organization and her various segments and clients.

Probably one of the earliest processes in the big picture is the ***Manufacturing Flow Management***. It includes all necessary activities for the route of goods on the production line and managing the manufacturing flexibility in line with the supply chain management. The choices in manufacturing reflects the opportunities for creation of a wide variety of products in due time at lowest possible cost. To achieve such levels of flexibility, planning and execution should be carried out off premises and must encompass the entire supply chain. The management of the manufacturing flow has to be integrated or synchronized with the process of human resources management.

The seventh process is the ***Product Development and Commercialisation***. It concerns the structure and approaches for the product development and its realization on the market. The teams responsible for the

product design and the commercialization process should coordinate them with those, responsible for Customer Service Management to identify the needs of the client – both present and future. Also they must identify the choice of materials and supplier in coordination with the Supplier Relationship Management process. At the same time, proper production technology should be developed, integrated with the best flow in the supply chain management for goods and processes for a particular product-market arrangement.

Last but not least comes the *Returns Management* process. It is linked to the Customer Service Management process, where the actions are on return of goods, reverse logistics, control and limitation of access to goods or avoiding such access. All this is carried out in the framework of the company and its key players in the supply chain management. The correct application of the process allows the company management to control efficiently those flows and at the same time to identify the options for reducing unwanted goods and reuse tools like containers. The effective management of return goods is an important link between marketing and logistics and presents options for high level of competitiveness.

As one can see, the listed business processes can be easily selected for digitalisation and digital transformation for a company or organization, and their effect can be visible in a relatively short time span. Most of them like the Returns Management process, Order Fulfilment, Customer Service Management, Supplier Relationship Management, and Customer Relationship Management do not require significant efforts, investment and time but can bring significant benefits for the organizations.

Changes and transformations in SCM are to be made by digital domains, using flows of information and data, which might be too big for people to identify, work and analyse, or to find the optimal sequences or flow of activities and processes and/or find errors or erroneous practices. Automated digital processes and use of AI will accelerate decision taking and will improve the services for clients, reducing costs. SCM will hold the opinion and requirement of the clients in the centre of the process, even during the manufacturing process.

3. GOOD PRACTICES IN SCM DIGITAL TRANSFORMATION

There are already a number of good examples of how digital transformation can be beneficial and accelerating the work for different organizations.

One good illustration is the acceleration of the deliveries in the transport/logistics area. At the end of 2017, Amazon launched its application for trucking, called Relay, designed to make trips to Amazon warehouses faster and more effective (Cunnane, 2017). Drivers enter cargo information into the app before they arrive to the destination point. Once enter cargo information in the application, they obtain a QR code, which is used at the entry gate. In such way a pre-checking in is performed, accelerating the process. There appears a better visibility in regard to the current location of the deliveries and the warehouse can better prepare for the arrivals. The entire process of accepting, checking and processing the deliveries is optimized and human errors can be avoided. Amongst the biggest challenges in the logistic/transport sector, where hundreds of millions of dollars are invested, are the trust factor and scalability.

Artificial Intelligence (AI), described above, is one of the most promising tools for digital transformation and a facilitator for the use of big (or unstructured) data. Sometimes companies develop their own expertise and skills in the AI area. The Plated company is an example for such approach. It uses AI and machine learning for prediction of various options and recipes for healthy food (Banker, 2018). The first activity of the Chief Data Science Officer (CDSO) at Plated was to build a data science team, consisting of data engineers, systems engineers, analysts, and modellers. Before the arrival of the CDSO, the company offered twelve recipes per week, with six new ones being swapped in each week. The CDSO raised the issue of the number of recipes Plated should offer. It is well known fact that more offers can lead to increase in customer acquisition and retention, but too many choices could drive up supply chain costs and make it difficult for customers to select easily what to eat. So with the help of AI, the cuisine tags were closely examined, describing what was in the meals, the types of food customers react positively to, the molecular nutrition, and the corresponding recipes. The

result was recommendation model per customer! And as a sequence of such forecasting model – a better inventory management was achieved.

A number of other examples exist and their quantity is increasing but the mentioned above clearly show the advantages of the digital transformation and digitalization in SCM. The best existing practices in the world show that digital transformation is not a term of the future, or some exotic experiment but a necessity for better effectiveness and efficiency in SCM. The listed (and other emerging) digital technologies are applicable to SCM and will support the optimisation of all business processes in the field.

CONCLUSION

In this paper, we have enlisted some of the most important factors, responsible for enhancing digitalization and digital transformation in supply chain management. Out of the paper, but bearing in mind are also the methodology, architecture and strategies to be considered and they are non-less important than the technologies and business processes themselves. An important conclusion is that the inevitable improvement of effectiveness and modernization in the SCM comes from the digitalization and digital transformation. The forecast for raise in the use and market share in the next few years of the applicable digital technologies clearly shows the new way. The links amongst the technologies, strategies, functionalities and methodologies have also to be synchronized and integrated.

A very important point is the approach of appointing Digital officer, or creating Digital transformation department/structure, responsible for proper and efficient digital policies. There is no sole or dominant solution for such transformation but any efforts will lead to a positive result. The educational efforts for acquainting the new graduates in SCM with digital transformations practices are also a must! The existing good and workable models, can be applied for any organization with similar activities. The pace of introduction of the digital technologies will be crucial for the success of an organization. The most important point is to make the first steps in the process and to develop a strategy and digital thinking as an integral approach for the development of the organization. It will payback and raise its competitiveness.

ACKNOWLEDGEMENT

This study was partially supported by the Digitalization of Economy in a Big Data Environment Grant BG05M2OP001-1.002-0002-C02, funded by Operational Programme “Science and Education for Smart Growth” 2014-2020.

REFERENCES

1. COMPUTERWORLD (2019), The Most Powerful IoT Companies In The World, *By Computerworld UK Staff, Computerworld*, [Online], 15 February 2019, Available from: <https://www.computerworld.com/article/3412287/the-most-powerful-internet-of-things--iot--companies-to-watch.html> [Accessed 10/08/2019].
2. CANO J. (2014), The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity, *Rethink Maintenance*, [Online], 11 March 2014, Available from: <https://www.xsnet.com/blog/bid/205405/the-v-s-of-big-data-velocity-volume-value-variety-and-veracity>, [Accessed 10/08/2019].
3. NAKAMOTO S. (2008), Bitcoin: A Peer-to-Peer Electronic Cash system, [Online], October 2008, Available from: <https://bitcoin.org/bitcoin.pdf>, [Accessed 11/08/2019].
4. BiTA (2017), Blockchain in Transport Alliance DRIVING STANDARDS AND ENABLING TECHNOLOGY ADOPTION, 2017, Available from: <https://www.bitastudio/>, [Accessed 11/08/2019].
5. CORKERY M. and POPPER N. (2018), From Farm to Blockchain: Walmart Tracks Its Lettuce, *The New York Times*, [Online], 24 September 2018, Available from: <https://www.nytimes.com/2018/09/24/business/walmart-blockchain-lettuce.html>, [Accessed 11/08/2019].
6. CROXTON K. and ROGERS D., (2002), The Demand Management Process, in *The International Journal of Logistics Management* 13(2), pp. 51-66, Available from: https://www.researchgate.net/publication/235304082_The_Demand_Management_Process [Accessed 12/08/2019].
7. CUNNANE C. (2017), Amazon Launches Trucking App, *Logistics Viewpoints*, [Online], November 2017, Available from:

- <https://logisticsviewpoints.com/2017/11/29/amazon-launches-trucking-app/>, [Accessed 13/08/2019].
8. BANKER S. (2018) Data Science And The Meal-Kit Subscription Business Model, *Forbes*, [Online], January 2018, Available from: <https://www.forbes.com/sites/stevebanker/2018/01/02/data-science-and-the-meal-kit-subscription-business-model/#39f70eb600bf>, [Accessed 13/08/2019].

CHALLENGES OF GDPR COMPLIANCE IN CONSUMER FINANCING COMPANIES

Monika Tsaneva¹

*¹ University of National and World Economy/Department of
Information Technologies and Communications, Sofia, Bulgaria,
mtzaneva@unwe.bg*

Abstract

From the point of view of information systems, EU General Data Protection Regulation (GDPR) is traditionally associated with the imposition of strict procedures and restrictions on the storage, processing and transmission of personal data. The aim of this paper is to propose a comprehensive approach to implementing GDPR requirements in information systems and applications operating in the consumer financing area. The research is based on an analysis of the business process and the typical information infrastructure of a credit institution on one hand, and on specifics of GDPR compliance in this sector, on another. As a result of this development, basic guidelines are proposed for how, while implementing GDPR's requirements, business can be expanded by creating the fundamentals for introducing cutting-edge information technologies, upgrading existing applications, developing new integration solutions, and developing B2B platforms. The main conclusion made from this research is that when carefully planned and implemented with the right technological solutions, GDPR compliance in the consumer financing companies can open up new business and technological opportunities, thereby ensuring further optimization of company operation and eventually enhancing customer satisfaction.

Keywords: *GDPR, integration solutions, consumer financing applications.*

INTRODUCTION

Nowadays everyone is familiar with the main rules and regulations that comprise GDPR. Consumers associate them with the binding consent to the processing and storage of their personal data, IT professionals with the technological challenges to their implementation, and business - most often with restrictions that must be respected. This material addresses the major challenges and opportunities that GDPR compliance presents to Business Information Systems (BIS) operating in consumer credit companies.

1. TYPICAL INFORMATION INFRASTRUCTURE OF A CONSUMER FINANCING COMPANY

Whenever some significant new features of data storage and data processing must be implemented, the starting point must be the general architecture of the information infrastructure. This is because all the applications and system modules run as interconnected components of a system that serves the needs of the main business process of the company Tsaneva (2019, p. 8), consisting of three main stages – Apply, Contract and Perform (see Figure 1).

1. **Apply:** Both applicants and retailers use some front-end (web or mobile) system to apply or to register the application for a credit. On this stage the applicant signs a Consent for personal data storage and processing. An individual's consent for each specific type of financial operation must be requested. (Remoortel, 2016), so usually, as a part of this consent, the applicant authorizes the credit institution to perform an automated scoring of his application, and to share his personal data with other institutions like national registers, third party credit scoring companies and insurers. After its registration, the credit risk of each loan application must be scored. Usually, this assessment is performed automatically by a standalone system. The most significant specific of Credit risk scoring systems is that usually they work in compliance with the rules and restrictions established for accessing data in cloud systems (Lazarova, V. 2016) and do not store and visualise personal data, they just receive this data, perform the evaluations and return the score without jeopardizing GDPR requirements in any way. Another specific is that in the common scenario, the scoring involves some data requested and received via API from some National Registers (Central Credit Register - CCR, National Social Security Institute – NSSI, Directorate General Civil Registration and Administrative Services - GRAO etc.) and this data is sent back to the operational system as an additional information to the evaluated credit risk score. If the loan application is rejected by the scoring system, the customers personal data may be stored for 5 years (if the consent is not withdrawn earlier) and then must be deleted.

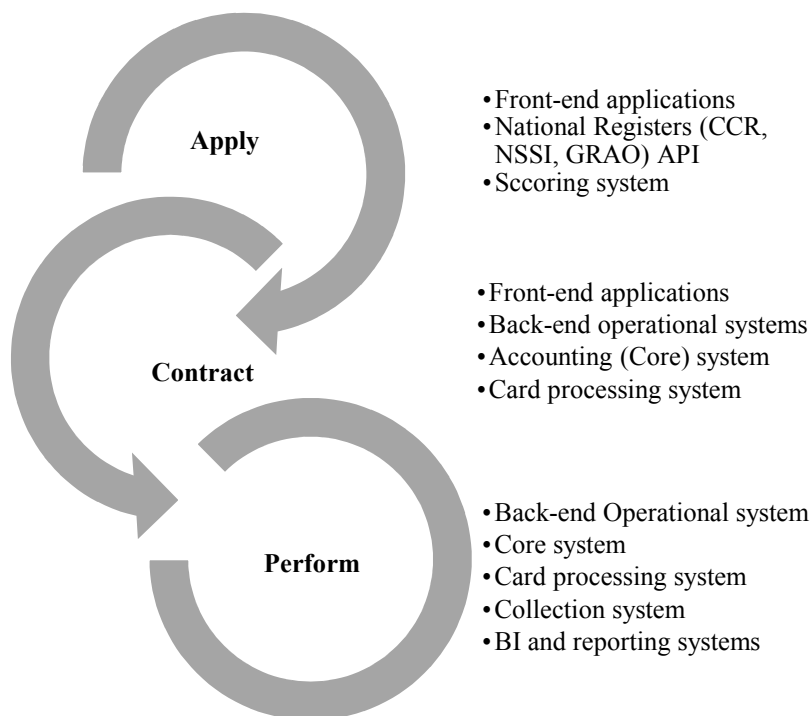


Figure 1. Typical information infrastructure of a Consumer Financing company

2. **Contract**: If the loan application is approved, a contract between the customer and the credit institution is signed. On this stage all the application/contract data is propagated from the operational system to the interconnected systems like the core (accounting) system, card processor (if a credit card has been contracted).

3. **Perform** (the contract): This stage includes all the administrative tasks that ensure regular loan repayment, quality communication with customers, reporting to company management and national regulators (Bulgarian National Bank -BNB, Central Credit Register - CCR etc.) On this stage additional data about bank or cash loan repayment transactions are collected on regular basis (usually monthly), all the communications (phone calls, e-mails, chats) with clients are recorded. On this stage

many indicators of contract performance are estimated. In the most common scenario, these indicators together with customers data are:

a) Exported in fixed record text files to national regulators on a weekly or monthly basis;

b) Loaded daily in the company's data warehouse. The company warehouse can be another relational database, or a Non-relational data store (Radoev, 2017, p. 153) to be presented to the management as dashboards providing informed data decision making abilities, so that business can continue to advance with the digital age (Mihova, Stefanov and Marzovanova, 2016, p 518);

c) Exported daily to Collection system or to another standalone company systems in files with predefined format.

All these activities are performed till the contract is fully repaid or terminated on some other legal basis.

Basically, all the above systems and modules have been developed during a long period of time and using completely incompatible information technologies, so the information infrastructure of a Consumer Financing company is heterogenous and ensuring the interoperability of its components in all the aspects including GDPR compliance is a complicated task

2. SPECIFICS OF GDPR COMPLIENCE IN CONSUMER CREDIT COMPANIES.

“GDPR another obstacle in an already complex system” and “Today's regulatory implementation, tomorrow's opportunity” – that's how the expectations of some managers (Kemp, 2017, p. 1) in financial business can be summarized.

MetaCompliance (2018, p.4) states the following GDPR principles:

- ✓ Lawfulness, fairness and transparency;
- ✓ Purpose limitation;
- ✓ Data minimisation;
- ✓ Accuracy;
- ✓ Storage limitation;
- ✓ Integrity and confidentiality.

Implementation of these principles in practice means (EY, 2018, p.6):

✓ Profiling is prohibited except when it is founded on a legitimate basis like explicit consent, performance of a contract and compliance with a legal obligation;

✓ Data protection impact assessments should take place prior to the processing of financial data and must serve to estimate the risks of processing data and to define mitigating measures;

✓ Services should adhere to data protection by design and by default principles which means that require service providers must think about the impact their services will have on data protection before delivering them;

✓ Data subjects always have the right of access to their information that is being processed and they have the right to receive a copy;

✓ Data subjects have the right of erasure which requires providers to take these rights into account when designing services, so that personal data can be deleted if requested and permissible.

A proper data subject consent gives the consumer credit companies legitimate basis to perform profiling in order to evaluate the credit risk of each loan applicant, so the prohibition of profiling is actually not an issue, but an opportunity for further automation of one of the most time and resource consuming activities which is credit risk assessment. This opportunity is extended by developing advanced integration modules for data retrieval from national registers via API provided by CCR, NNSI and other Bulgarian national institutions.

The following points need to be considered during the design process of GDPR compliance of information systems and their integration solutions:

1. Properly determine the scope of personal data (see Figure 2) that needs protection and must be ready to be delivered to the data subject in a machine readable format or be deleted ("forgotten").

Based on the business process, the scope of personal data that must be returned to the data subject or must be "forgotten" upon request is:

✓ Personal data registered as a part of the application for a loan like names, addresses, e-mails, phones and other contact data, education, workplace, incomes, debts and so on;

✓ Data received from national registers – from NNSI retrieved during application assessment, from CCR retrieved both during application

assessment and imported on regular basis (monthly) during loan repayment;

✓ Terms and conditions of the loan that have been contracted like purchased goods, loan amount or credit card limits, credit card numbers, bank accounts numbers, insurances, loan duration, interest percentages and so on;

✓ Data collected during contract performance about regular payments (bank or cash transactions) made by the client, customer calls, correspondence regarding collection process and so on;

✓ All the data that is transferred to the company data warehouse and can be used to identify a customer including names, EGN, identification card, addresses, phones, e- mails, workplace etc.

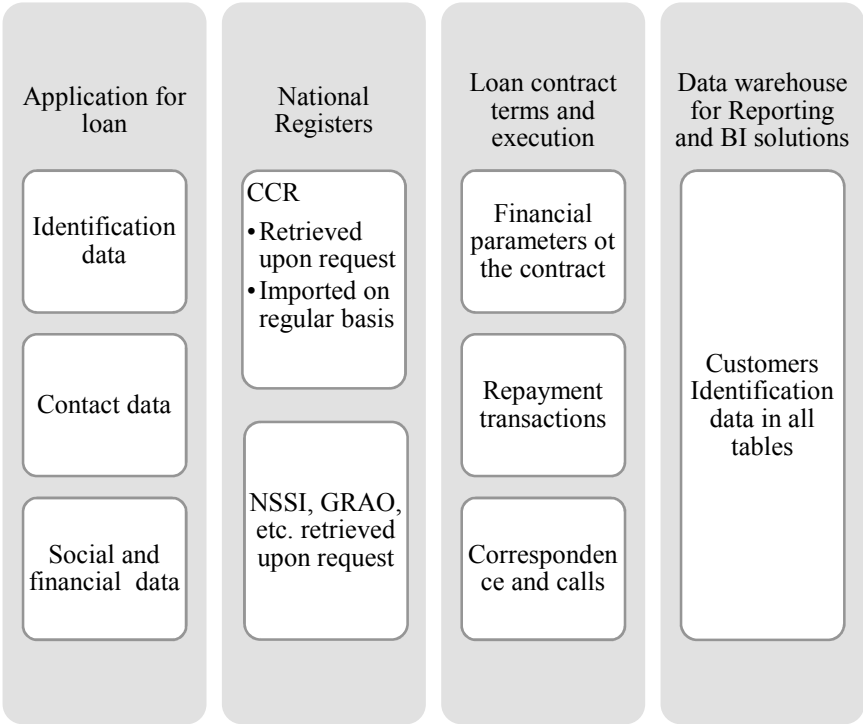


Figure 2. Scope of personal data that must be returned or “forgotten”

In addition to this, when a customer or employee needs to be forgotten, all the backups of the databases that are not currently destroyed must also be included in the scope of deletion. All drives/folders that contain obsolete data exported by legal obligations for national institutions must also be checked and cleared.

2. Check and, if necessary, redesign the user access control system, in order to guarantee, that every user of each system will have access to read and/or modify only to this data that are relevant to his job description. Some hard choices must be made because such changes usually require a large amount of work for user interface components redevelopment especially for legacy desktop systems. When new functionality is designed, data visualization must be organised considering data usage defended by Data Protection Officer (DPO) for different user roles. An alternative approach like using data encryption options embedded in modern databases must also be considered. Generally, database embedded data encryption is more suitable for new systems development. When applied to existing systems, this approach will require both data migration and redevelopment of application's data access tier, which can become a very big issue for large size databases that need a large amount of time to be migrated.

3. Choose and develop a proper single-entry module for data subjects (see Figure 3) who want to withdraw their consent, change its individual permissions, receive a copy of their data or request to be "forgotten". Opposed to traditional vision, this module is not a simple form (or a web page) containing a set of checkboxes, an "I accept" and a "Reject" button. Such a "Consent processing" module is an integration solution that includes substantial business logic and eventually results in a sequence of method invocations in all the interconnected modules and systems that store a particular set of personal data.

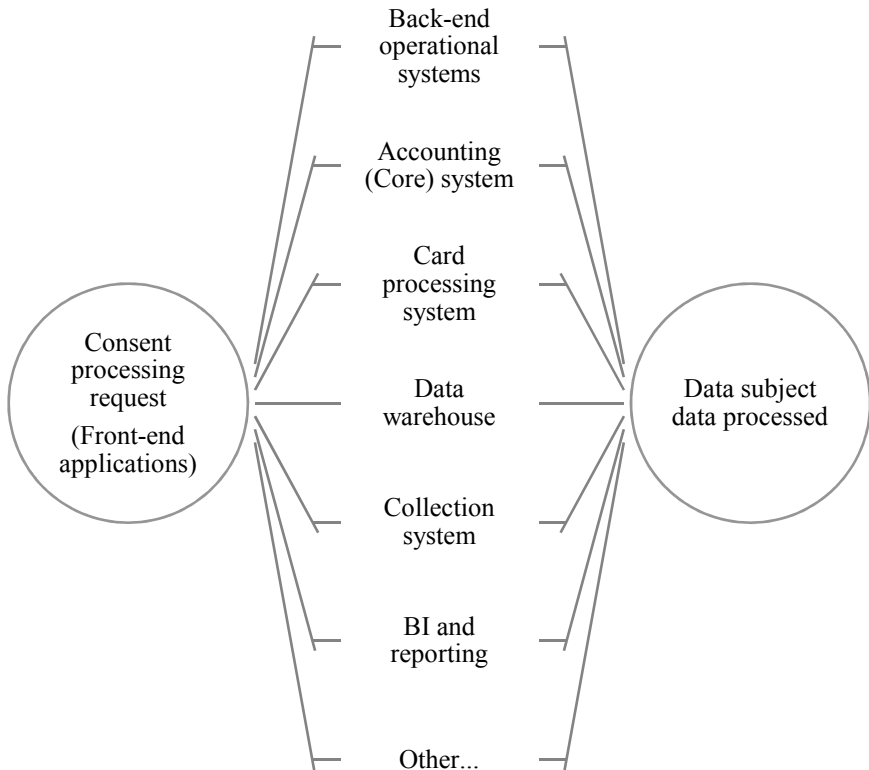


Figure 3. Single-entry “Consent processing” module for data subjects

4. Consider and implement the “anonymize” instead of “forget” personal data functionality. The main reason for such a consideration is the process of score card development for loan application assessment. Usually the content of a score card is based on contract performance indicators estimated for different customers segments depending on contract point of sale, purchased goods, customers age, education, social and financial status etc. So the simple deletion of the customer and all his related data (as required by relations in the data base) is not a suitable option. The more flexible approach that can be implemented is to anonymize all the sensitive data that can be used to identify a particular per-

son. A simple option is to replace all these actual values with hash codes, thus guaranteeing on one hand consistency of data and database relations and on the other - the impossibility of unauthorized sensitive data recovery. Anonymization must be performed simultaneously over all the data bases and data warehouses in the company thus ensuring both GDPR compliance and reporting systems operability.

3. BUSINESS AND TECHNOLOGICAL OPPORTUNITIES GIVEN BY GDPR COMPLIANCE IN CONSUMER CREDIT COMPANIES.

GDPR compliance is a time consuming and costly process because it involves significant redevelopment of data base structures, API methods and user interface components in a complex heterogeneous environment. Implementing well thought-out, technology-relevant and business-friendly solutions can achieve serious competitive and strategic benefits for the company (see Figure 4). The GDPR is not the only regulation which targets financial services industry, but consumers will also benefit from Payment Services Directive (PSDII), which requires financial services companies to open their data and payments infrastructure to promote increased competition. “PSDII, actually provides an opportunity for the financial services industry. This directive should allow innovative banks to improve their customer experience, increasing value to corporate customers” (Kemp, 2017, p. 1) because the treatment of data under PSDII must also comply to GDPR requirements.

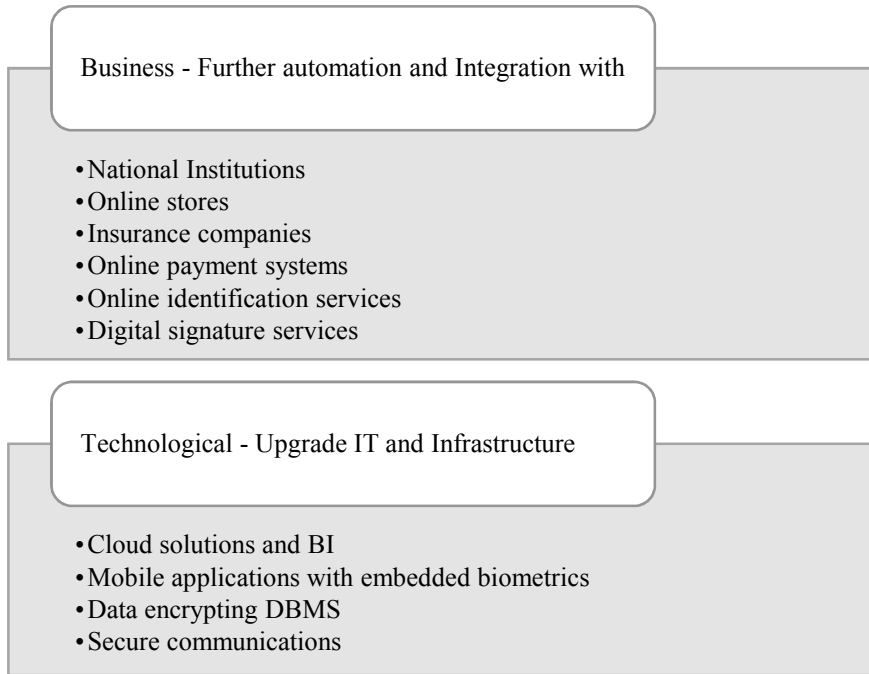


Figure 4. Opportunities given by GDPR compliance

Many new business opportunities are opened by data portability - the right to transfer consumers personal data in a structured, commonly used and machine-readable format. That way further automation of business activities and data entry reduction are achieved. A variety of additional services may be offered via Integration via API with third party systems for common business solutions.

✓ Integration with some national institutions like CCR, BNB, NSSI using the API they provide;

✓ The commonly used interoperability between digital (online) stores and credit institutions is greatly facilitated;

✓ Thanks to the GDPR-regulated data exchange, financial institutions and insurers offer common products and their overall business process is fully automated using specialized API methods. Additional options may be sought in the direction of automating periodic calculations

and reporting between them, for which the use of API is not the most appropriate option due to the large volumes of data exchanged;

- ✓ Common business solutions can also be implemented with external online payment systems (like ePay, EasyPay etc.), although most credit institutions encourage customers to use the implemented internal modules and applications;

- ✓ A very modern and promising area is the integration of services for electronic identification and electronic signature of documents (credit agreement, insurance policies, even the consent for processing personal data).

Investment in bringing information systems and GDPR compliance can increase their effectiveness if combined with upgrading the entire IT infrastructure by:

- ✓ Deploying cloud technologies, that ensure complete and much more secure control over user access to company services and data;

- ✓ Development of mobile applications using embedded biometric data for secure consumer identification;

- ✓ Migrating to the latest DBMSs that have built-in data encryption capabilities for whole tables or individual table columns;

- ✓ Construction of advanced network equipment with encrypted communication channels and high level of intrusion protection.

CONCLUSION

Implementation of GDPR compliance solutions in consumer financing companies requires hard work and significant amount of time and finances. But when carefully planned and implemented with the right technological solutions, this process can open new business opportunities and create prerequisites for technological upgrading of the information infrastructure, thereby ensuring further optimization of business processes eventually enhancing customer satisfaction. Financial sector companies passed the first easy and cheap steps – they have legal basis to store and process the personal data of their customers and have implemented the minimum GDPR requirements by designating DPOs and demanding data subjects to sign an individual consent for each purpose. To reap the potential benefits of GDPR compliance, the firm's business strategy must be updated, and information technology and infrastructure evolution strategy must be aligned with it,

then development and investment priorities must be set, and finally a detailed plan for their realization must be created and implemented.

REFERENCES

1. EY, Navigating the PSD2 and GDPR challenges faced by banks (2018), [Online], Available from [https://www.ey.com/Publication/vwLUAssets/ey-navigating-the-psd2-and-gdpr-challenges-faced-by-banks/\\$FILE/ey-navigating-the-psd2-and-gdpr-challenges-faced-by-banks.pdf](https://www.ey.com/Publication/vwLUAssets/ey-navigating-the-psd2-and-gdpr-challenges-faced-by-banks/$FILE/ey-navigating-the-psd2-and-gdpr-challenges-faced-by-banks.pdf) [Accessed 17/08/2019].
2. KEMP, J. (2017) GDPR: A challenge for the financial services industry, [Online], Available from: <https://www.euractiv.com/section/economy-jobs/opinion/gdpr-a-challenge-for-the-financial-services-industry/> [Accessed 15/08/2019].
3. LAZAROVA, V. (2016). Managing User Access to Cloud Services by Company Administrators. 2016 TEM Journal, 5(3), p.289-293
4. MetaCompliance (2018) GDPR Best Practices Implementation Guide, [Online], Available from https://www.infosecurityeurope.com/_novadocuments/355669?v=636289786574700000 , [Accessed 17/08/2019].
5. MIHOVA, V., STEFANOV, G., MARZOVANOVA, M. (2016). 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. [Online] pp. 514-519. Available from: <http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2016.pdf>, [Accessed 17/07/2019].
6. RADOEV M. (2017), A Comparison Between Characteristics of NoSQL Databases and Traditional Databases, Computer Science and Information Technology, Vol. 5(5), 2017, San Jose, CA, USA, [Online], pp. 149 – 153, Available from: <http://www.hrpub.org/download/20171130/CSIT1-13510351.pdf> [Accessed 16/07/2019].
7. REMOORTEL, FR.,(2016), Financial institutions and the General Data Protection Regulation, Financier Worldwide Magazine, November 2016, AVAILABLE from: <https://www.financierworldwide.com/financial-institutions-and-the-general-data-protection-regulation#.XS2i4OgzaHt> [Accessed 16/07/2019].

8. TSANEVA, M. (2019) A practical approach for integrating heterogeneous systems. Business Management 2/2019 [Online] pp. 5-15. Available from: <http://bm.uni-svishtov.bg/title.asp?title=1383>, [Accessed 17/08/2019].

MAIN FACTORS INFLUENCING DIGITIZATION IN CONSTRUCTION COMPANIES

Todorka Atanasova¹

¹ University of Economics, Varna, Bulgaria, t_atanasova@ue-varna.bg

Abstract

Digitization is an inevitable process for companies' prosperity. In the construction industry, its pace of entry is low. Its acceleration depends on many factors, as well as on the level of development of information and communication technologies. The report examines the major national factors and the European digitization policies affecting the digitization of construction companies in Bulgaria. Information technologies, applied in this area, have been researched and presented.

Keywords: digitization, construction companies, information technologies.

INTRODUCTION

Digitization began with the fourth technological revolution and is an irreversible process. It leads to new models of business, commercial and social relations, offering new goods and services and more, and in general, to a higher level of productivity and standard of living. In different parts of the world, digitization is proceeding at a different pace. "Globally, Europe, the Middle East and Africa are well behind the leaders in the rankings - Asia and Oceania. One fifth of the companies in Asia and Oceania are already reaping the benefits of the digital revolution. North and South America ranked second with 11% of digitized companies" (Capital, 2018). In construction, the rates are slower than in other business areas. The area is quite conservative, full of responsibilities and uncertainty. It is expected that digitization in construction companies will significantly reduce the cost of construction of the sites and the time for their completion, as well as improve the quality of work.

The purpose of this report is to outline the main factors influencing digitization in construction companies and information technologies that are specific to this field. The report was developed under project BG05M2OP001-1.002-0002-C02 "Digitizing the Economy in a Big Data Environment (DIGD)".

1. CONDITION OF DIGITIZATION IN CONSTRUCTION COMPANIES IN BULGARIA

In Gartner's IT dictionary, digitization is defined as "the process of changing from analog to digital," as well as "the use of digital technology to change the business model and provide new revenue and value opportunities." (It-Glossary). Digitization is seen not only as a process, but also as an environment integrating digital resources, services and professionals with the necessary knowledge and skills at the technology level of that environment. Digitization should be considered in terms of assets, their use (business processes, transactions, interactions) and the workforce (digitization of labor, deepening importance of digital capital, etc.). According to the progress made in the process of digitization, three levels of maturity in an organization can be determined (Fletcher, G., 2014)- Fig. 1.

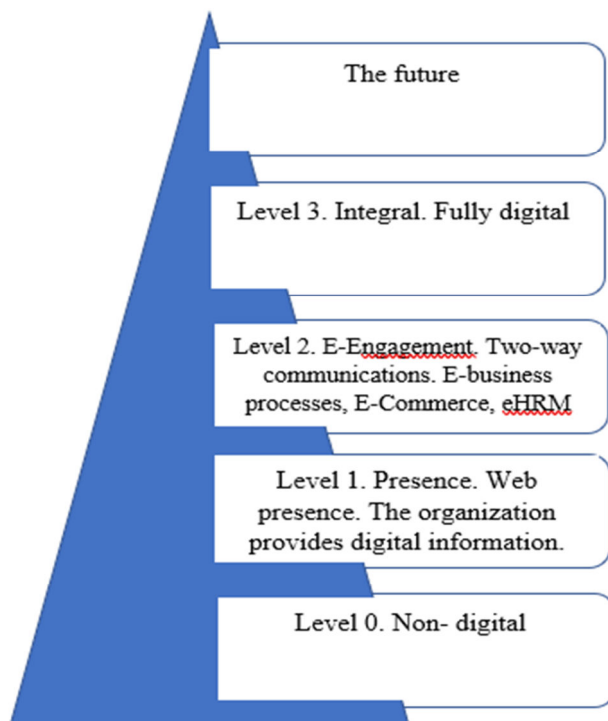


Figure 1. Levels of digital maturity of the organization

If we look at the levels of digital maturity of the company, the achieved level in most construction companies in Bulgaria is Level 1 and partly Level 2. Parts of the construction process, such as design stages and customer relationships, are digitized. But in the implementation of the construction site activities, digitization is far behind. Project planning is often done on paper. Contracts do not include incentives for risk-sharing and innovation; performance management is inadequate and supply chain practices are still unprepared.

Technical challenges specific to the construction sector play a negative role in the slow pace of digitization. Developing solutions for construction sites in different sectors that are geographically dispersed is not an easy task. Due to the lack of digitization, the exchange of information is slow and not universal.

In addition, project owners and contractors often use different IT platforms that do not synchronize with one another. As a result, a single source is not maintained to provide integrated, real-time project overview, cost and schedule.

There are also many problems in the sector, such as inadequate schedules and costs, poor quality and design errors, as well as inaccurate and incomplete plans and lack of cooperation.

2. FACTORS AFFECTING DIGITIZATION IN CONSTRUCTION

Factors affecting digitization in construction can be considered as national and European.

Research by some authors (Billon, Lera-Lopez, Marco, 2010) shows that for poorly digitized countries, like ours, infrastructure is a major national factor in the process under consideration. It gives access to the digital economy and stimulates its development. Infrastructure includes "all necessary facilities and facilities that indirectly serve the economy and life (roads, sewage, energy facilities, etc.)" (Bulgarian Interpretative Dictionary).

Of the various types of infrastructure for the digitization in the construction industry, production and communication are essential.

Digitization of production infrastructure in construction means the use of digital machines, tools and technologies, interactive work processes.

Applying digital sensors to machines, cameras, and monitoring devices, connecting them to a common platform, enables the data to be online. This is a fundamental necessity in the process of digitization, which also becomes the engine for its success. The platform may use a cloud service to reduce the cost to the company of purchasing equipment, providing access to data and protecting it. The availability of up-to-date data on-line loads workflows, provides work continuity, and provides a basis for analysis and decision-making. Digital models of reality - simulations, augmented reality, 3D printing provide on-line information on the projects being developed.

The use of digital tools is part of the digitization of the work environment in the company. They have a number of advantages:

- the accuracy of the digital electronic instrument is very high. Their readings are in digital form, which reduces the error.
- the digital instrument consists of sensitive elements that are easily responsive to ambient temperature and humidity.
- use less energy.
- digital output is obtained from a tool that acts as an input to memory devices such as disks, recorders, etc.

Creating and implementing smart products and technologies leads to a new organization of business processes and a work environment. Companies form new workflow schemes, for example, with web-based coordination of activities through linked schedules, transportation, material delivery, document sharing and more. Digitization cannot be accomplished without a virtual picture of physical processes.

When making a change in workflows, it is necessary to monitor its impact on other components of the business.

An essential factor for successful digitization is the management of software and data through data-based services. Advanced analysis capabilities are being built; and developing customer-based products. A ProductXchange software platform has been created in Bulgaria to categorize and digitize product information in construction. The software product was developed by the Bulgarian company ICB by coBuilder. With the help of ProductXchange, construction industry companies can more effectively control the use of construction materials and monitor their compliance with European and national regulations.

Communication infrastructure is influenced by new forms of machine-to-machine, machine-product, machine-to-human, environment-to-machine communications, etc. They are a factor which is necessary for the digital transformation of the business and its automation, incl. use of self-organizing systems.

Internet usage is related to the communication infrastructure. These include: electronic data collection and analysis, electronic data access (mobile Internet access and corporate networks), connectivity and activity synchronization. The network needs wireless and risk-free high-speed connections and clear rules for access and use. Even short-term loss of connectivity can lead to a delay or blockage of the manufacturing process, leading to high costs.

Internet access for small and medium-sized construction companies is an important factor in their digitization. It will also provide easy access to digital technologies. Any measure that facilitates the creation of a secure online transaction environment would have a positive impact in the process under consideration. High cybersecurity will be a driving factor.

The competitiveness of information and communication providers and the degree of competition between different operators play a well-defined role in providing new services and setting the pace for new platforms and applications.

Another extremely important national factor influencing the digitization of the construction industry is the human resources - the managers as well as the employees in the company. They are at the heart of the transformation. In particular, this refers to the thinking and qualification of the management and staff of the company.

Management must have a vision of the intended digitization state it is aiming for. An important factor for digital transformation in a construction company is the presence of a strategy for digitization, defining the concept and design of products and services. Leader, resources and necessary investments need to be identified. The critical question to ask is: not adapting to the digital world can be professional suicide for the company. It is important to create a sense of urgency.

Management decisions should be based on data and their analysis.

Significant influence is the people working in the company, their age and their level of qualification. IT support for software and technology is required, requiring an appropriate level of knowledge and experience.

Digitization has been identified as a key priority in the work program of the European Construction Industry Federation (FIEC), and its importance is reflected in its work program of 2016. This is the reason FIEC to work together with other major European associations working in the construction industry and prepare a European manifesto for the digitization of the construction industry (FIEC, 2018). On the Bulgarian side, the European Manifesto is supported by the Union of Civil Engineers in Bulgaria (SEIB).

The European factors influencing digitization include the digitization policies of European industry, which are reflected in the initiatives:

- COM (2015) 0192 Digital Single Market Strategy
- COM (2016) 180 Digitization of European Industry. This initiative is a key element of the Digital Single Market Strategy. A set of measures is being introduced to build on and complement the various national digitization initiatives of all industrial sectors and related services.
- COM (2016) 0176 ICT standartization priorities for the digital single market. The aim is to guarantee a new approach to standards in the areas of: 5G networks, the Internet of Things, cloud services, cybersecurity and data processing.
- COM (2016) 0178 European Cloud Service Initiative - Building a competitive economy in Europe based on data and knowledge.
- COM (2017) 0572 Increasing the efficiency of public procurement in Europe and for the benefit of Europe.
- SWD (2017) 0157 Digital Development: integrating digital technologies and services into EU development policy.

European digital platforms have been set up to support the development of the construction industry. They are operating systems that integrate different technologies, applications and services. The construction aims to meet the challenges of using digital tools to support the digital development of the sector, as well as Building Information Modeling (BIM), building passports, and more. It is the process of

creating and managing data for a building over its entire life cycle. BIM is at the center of digital transformation in construction and is its equivalent. It integrates all the information about a building - architectural, structural, technological and financial, tracking all interconnections and dependencies. It allows you to view the building as one complete object.

Funding for digitization projects is an important factor and driver for their success. To do this, companies must apply for state or European funding, which means they have a vision and a plan in advance of how the digital transformation process will proceed.

An important factor is the European level of standardization and interoperability of information technologies.

3. INFORMATION TECHNOLOGY FOR DIGITIZATION IN CONSTRUCTION

Information technologies are available to support most aspects of a construction project.

According to some researchers (Budiac,D.,2018) in North America (mostly in USA and Canada), trends in information technology with wide application in construction are shown in Figure 2.

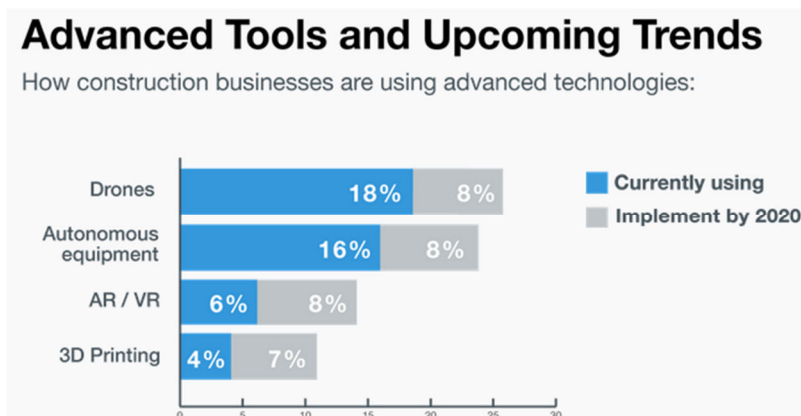


Figure 2. Ongoing use of information technologies and trends in their application in construction

Drones are used for aerial surveys, photography, laser scanning and thermal imaging.

Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality technologies allow a computer-generated 3D model to create a preliminary image of the object that will be built and navigated in it.

Network and communication technologies: Local Area Network (LAN), Wide

Area Network (WAN), Electronic Data Interchange (EDI), Internet, Intranet,

VRML, wireless, groupware, etc

Network and communication technologies: Local Area Network (LAN), Wide

Area Network (WAN), Electronic Data Interchange (EDI), Internet, Intranet,

VRML, wireless, groupware, etc

Cloud computing has a significant place in construction. The constant shift of workers and the creation of new jobs require access to data at any time and place, as well as access to the necessary software. The study cited above (Budiad,D.,2018) shows that the construction business is open to deploying cloud-based software.

Open to cloud software	2017	2016	2015
Construction	87%	84%	78%
All other industries	82%	79%	77%

Figure 3. Construction business is open to cloud hosted software

Network and communication technologies are presented in: Internet of things, Local Networks, Electronic Data Interchange, wireless, mobile connections, etc.

Internet of things (IoT) technology allows construction machines, equipment and materials to "communicate" with a central platform that takes into account various parameters. Sensor and wireless technologies

allow instruments to become "intelligent" when connected to one another. Provides monitoring and repair of equipment through wear sensors; inventory management and automation of material orders, security (prevention of workplace accidents). The innovative leap of this technology over automated factories and control software is that different elements can be identified in the respective systems automatically without contact.

Artificial intelligence (AI) and machine learning are very useful digital technologies for construction companies. The high volume of data can be used to predict risks and increase competitiveness. AI in the construction industry is focused mainly on finding models in large datasets. Also, algorithms used by AI report errors occurring on the construction site, based on the recognition of different buildings by their shape, size and location.

CONCLUSION

Construction is one of the industries with the most intensive use of information, since the main processes require a large amount of data. Due to a number of peculiarities of the activity, as well as the impact of the factors considered in the report, digitization in construction has been delayed. The implementation of information technologies, together with the improvement of the skills for working with them, will allow to accelerate the digitization in the construction industry.

REFERENCES

1. BILLON, MARCO, LERA-LOPEZ. Differences in digitalization levels: A multivariate analysis studying the global digital divide. Review of World Economics 146(1) April 2010
2. BUDIAC, D. Construction Technology Trends- 2018 Report <https://softwareconnect.com/construction/technology-trends-2018-report/> (Accessed 08.2019)
3. Bulgarian Interpretative Dictionary
4. Europe is lagging behind with digital transformation, 2018, https://www.capital.bg/politika_i_ikonomika/sviat/2018/06/07/319010_0_evropa_izostava_s_digitalnata_transformaciia/ (Accessed 08.2019) (in Bulgarian).

5. FLETCHER, G. Digital Business Maturity Model: The future of #Digital Business. <http://blogs.salford.ac.uk/business-school/digital-business-maturity-model/>, 2014 (Accessed 08.2019)
6. <http://productxchange.com>
7. INFORMATION TECHNOLOGY IN THE CONSTRUCTION INDUSTRY, <https://www.clearpathit.com/information-technology-in-the-construction-industry> (Accessed 08.2019)
8. IT Glossary. <https://www.gartner.com/it-glossary/digitalization/> (Accessed 08.2019)
9. STEFANOVA, M. Digitalization of the construction industry. European initiatives and standards for building information modeling, 2018 <http://www.bef.uceb.eu/wp-content/uploads/2018/05/08-MARIA-STEFAANOVA-Industria.pdf>, (Accessed 08.2019) (in Bulgarian).

DIGITALIZATION AND DIGITAL TRANSFORMATION IN CONSTRUCTION – BENEFITS AND CHALLENGES

Silvia Parusheva¹

*¹ University of Economics – Varna, Department of Informatics,
Varna, Bulgaria, parusheva@ue-varna.bg*

Abstract

Digitalization is essential for the development of the construction sector with the opportunities it offers to change and optimize the construction business. Digitalization affects every stage and process, the whole value chain. As for a number of other industries, as well as in the construction sector, it provides a number of advantages and benefits, including: increased productivity; increased speed of construction and saving time in the implementation of construction projects, significantly higher quality of accompanying construction documentation, etc. At the same time, there are some challenges related to the lag of digital transformation in construction, compared to other sectors, the presence of some specific technical challenges, different degree of application of new IT in smaller construction companies, which are often in the role of subcontractors and others. Despite the challenges, digital transformation has no alternative in view of the future upward development of the construction sector.

Keywords: *construction, digitalization, digital transformation, benefits, challenges, IoT, robots, 3D printing, RFID.*

INTRODUCTION

The construction sector is of key importance for the development of the world economy. It accounts for 6% of global gross domestic product (GDP) (World Economic Forum, 2017). Construction is associated with almost all industries. For other industries to function, they usually rely on buildings or assets that have to be built. In addition, it should be noted that constructed assets, with all their diversity (residential buildings, roads, bridges, schools, hospitals, etc.), have a direct impact on the quality of life of people. Furthermore, the prospects of intensive development of the construction sector in the future related to rapid urbanization and investment based on it plus the construction of smart cities and all related infrastructure and the introduction of the industrial Internet of Things concept, etc., should also be taken into account. As an example of the

expected prospects can be pointed an assessment of growth of urban population with 200,000 people a day. In this connection, it should be borne in mind that these residents need housing and related transportation, communal and other infrastructure and this construction is related to the construction sector (World Economic Forum, 2017). It is these global trends that pose significant challenges to the sector and, accordingly, the issues of digitalization and digital transformation are on the agenda, which will most effectively help address these challenges related to the expectations of rapid construction.

This report aims to highlight the key benefits that digitalization can provide to the construction sector and on the basis of which information and communication technologies and approaches it is possible to realize these benefits. On the other hand, it is also important to identify the challenges that construction companies must face in order to be able to take the most effective advantage of digitalization opportunities.

1. IMPORTANCE OF DIGITALIZATION FOR THE CONSTRUCTION SECTOR AND ITS BENEFITS

Digitization is essential for the development of the construction sector with the opportunities it offers to change and optimize the construction business. Digitization affects every stage of it, its processes, the whole value chain. As for many other industries, as well as in the construction sector, it has many **advantages**. The following may be indicated (Aghimien et al., 2018):

- increased productivity and efficiency;
- increased speed of construction activities;
- saving time in the implementation of construction projects and improved opportunities for adhering to construction schedules;
- significantly higher quality and safety of construction, incl. and the quality of the accompanying construction documentation;
- improved design of buildings, etc.

Among these strengths, some authors point out as leading one the opportunities that digitalization offers to **increase the productivity of firms in the construction sector** (Berger, 2016). Research shows that despite the lagging behind the construction industry reaping the benefits of digitalization, it is also growing, though not as large as other indus-

tries. To illustrate this fact, Berger gives an example of construction productivity in Germany, which has grown by 4.1% over the last ten years, while in the whole German economy it has increased by 11% over the same period. The digitalization in construction contributes to reduced construction costs and to have more successful projects. Considering that construction is the world's largest consumer of raw materials, digitalization helps make more efficient use of scarce materials globally.

2. TECHNOLOGICAL SOLUTIONS AND APPROACHES CARRYING BENEFITS FOR THE CONSTRUCTION SECTOR

In the field of construction, certain **technological solutions and approaches** in the field of information and communication technologies and digitalization can be mentioned, which influence the business practices of companies in the construction sector in a very favourable way and give them a number of advantages (Berger, 2016). These include:

- use of **digital platforms for the supply of raw materials**;

As a rule, raw materials represent a substantial part of the total cost in the construction sector, and digital platforms contribute to reducing these costs. According to statistics, electronic supply of raw materials helps *save, for example, 5% of the value of catalog-based purchases and 10% of value when auctioned.*

- implementation of **digital tools for smart logistics and maintenance of construction sites** - software for managing supply activities, exploiting the capabilities of the Internet of Things and radio frequency identification technology;

In construction, the following specifics are observed - construction workers use only about 30% of their working time to carry out their main activity. The remaining 70% is used for supplementary activities such as transporting materials, cleaning and rearranging the construction site, searching for materials and equipment, etc. This ratio in the use of working time can be optimized with the help of **digital tools**, with which the **benefits** of digitalization are widespread. Such digital assets may be: the use of *supply management software*, which results in materials being shipped to the construction site at the exact time when they are needed and work on their storage and rearrangement can be minimized.

Another possibility of intelligent logistics is the *intelligent, connected construction machines* as part of the work environment in the context of the Internet of Things. These machines can have appropriate *sensors*, which can optimize the work of construction workers and their auxiliary equipment (hoists and vehicles, etc.) - less waste of time for coordination and efficiency by synchronizing the base sensor signals, which can be seen as an undoubted **advantage** as a result of digital transformation.

Radio Frequency Identification (RFID) technology is also an important digital tool. Building materials, equipment and products, equipped with this technology can be identified and tracked by electromagnetic fields. They can be recorded and scanned, thus optimizing their location on the construction site.

- use of **drones, robots, 3D technology and 3D printing.**

The use of **drones** in the construction industry can potentially be useful at various stages of construction projects - in the preliminary planning, detailed exploration and mapping of the construction site, monitoring of the construction process, post-construction inspections, sales and marketing (Liu et al., 2014; Anwar et al., 2018). Drones can have a variety of equipment, such as a high-resolution camera, incl. for 3D capture and video streaming, RFID reader, GPS device, Wi-Fi communication and more. The data obtained from the sensors of these unmanned aerial vehicles can be processed and analyzed to provide important business information. Drones can be applied at every stage of the construction value chain - from pre-construction throughout the construction phase by providing information from the site of designers, construction contractors, construction supervision to the final stages of the construction project, incl. when preparing impact assessment reports. Intelligent monitoring of the construction process can significantly reduce the effort and cost of monitoring and reporting construction procedures, which are particularly large when implemented in the traditional way, especially in large-scale construction sites. **The benefits of using them** in construction work are contributing to increasing efficiency and productivity at all these stages by helping to monitor: progress of construction site activities, compliance of contractors' reports with actual activities performed or with regulatory requirements.

Robots can also bring significant benefits to the efficiency and productivity of construction work. They are used for mapping the construction site, for laying bricks, for preparing the necessary materials, etc. (Jayaraj & Divakar, 2018). Robots can replace humans in construction activities, such as working with chemical dyes, at high altitudes, which pose a threat to normal working conditions. Specific construction practices are cited in the publications, according to which activities requiring several weeks of normal construction activity are carried out for 48 hours with the help of a robot using 3D construction plans, according to which it arranges, processes and lays every brick (Berger, 2016). The **benefits** of implementing this technology are cost savings and improved performance.

3D construction printing is another innovative method that offers **advantages** in terms of digitizing the construction sector. 3D construction printing is a new and innovative way of construction and can be used to produce construction components of various sizes with high precision - from micro to macro components, creating layer by layer (Ledi-ga & Kruger, 2017). At the beginning of the process are computer-aided design and digital scanning models. Specialized software "cuts" the components into thin sections, each of which is then printed. When properly designed and used, this method has different advantages over traditional construction methods, as it creates opportunities to reduce project time and cost.

3. CHALLENGES TO DIGITALIZATION IN THE CONSTRUCTION SECTOR

The implementation of digitalization and digital transformation in the construction sector can be associated with some the challenges that need to be taken into account. In its study, the McKinsey Research Company (McKinsey, 2015) found a **lag in the construction sector** in terms of its degree of **digitalization** based on its 2015 McKinsey Global Institute industry digitization index among 22 compared sectors. Construction ranks second to last (after only the Agriculture and Hunting sector.) The construction sector has not yet embraced the new digital technologies that need initial investment, despite the fact that the long-term benefits are considerably high. Expenditure on research and devel-

opment in construction are far behind those in other industries: they represent less than 1% of revenue - a benchmark in automotive and aerospace sectors are between 3.5% to 4.5%. For its part, IT costs are also insufficient and account for less than 1% of construction revenues. McKinsey identifies **technical challenges** specific to the construction sector as the cause of the slow pace of digitalization. The **varying degrees of application of new IT in smaller construction companies**, which often play the role of subcontractors, may also be challenging.

Researchers and research firms (Berger, 2016) see several key points directly related to the ability to unleash the potential of digitalization - digital data, digital access, automation and connectivity. These key points can have an impact and become a link in the value chain in the construction sector: in logistics, in the supply of raw materials and production and in production, in marketing and sales, and in the after-sales marketing. Firms in the construction sector face the challenge of deciding which approach to target and how to best implement it. In the studies, Berger points out that companies in the construction industry that are more focused on technological development and carefully consider how to apply technological innovations in the value chain are more likely to outperform competitive players, because with the introduction of digital methods, will become more productive and efficient.

Other researchers and research companies (Solis et al., 2014; Capgemini Consulting, 2013; Ezeokoli, 2016) argue that firms need to face the multiple challenges of digital transformation to create a complete digital consumer experience, meets their expectations with every possible interaction. The major **challenges** include:

- aligning the right resources and the proper management of the right teams to focus on digital services;
- the perception by companies of the thinking of a prime mover in the pursuit of digital innovation, or so-called “digital-first”;
- building "digital enthusiasm" and motivation among mid- and operational-level management;
- requiring appropriate minimum digital knowledge and skills of employees;
- integrating all digital initiatives and upgrading to the next technological steps;

- search for the most suitable partners for the implementation of digital initiatives;
- improving digital services to meet new requirements and the need for adaptation, etc.

CONCLUSION AND FUTURE WORK

This study identifies the main benefits that digitalization brings to the construction sector, related to increased productivity and efficiency, increased speed of construction activities, shortening the deadlines for implementation of construction projects and adhering to construction schedules, the contribution of IT solutions for improved design of buildings, etc. In addition, the main and the most up-to-date technological solutions and approaches that implement the digitalization in the construction sector are outlined.

On the other hand, the findings indicate the major challenges associated with the digitalization of the construction industry, that include some lag in the construction sector in terms of its degree of digitalization, some technical challenges specific to the construction sector, identifying the right resources needed for digitalization, as well as adequately managing the right teams to accomplish digital transformation. Companies in the construction sector are aware of the potential of emerging technologies such as Building Information Modeling (BIM), 3D printing, robots, wireless sensing, but they are not fully aware of exactly how to implement them in their organizations' and digitalization's strategy. There are difficulties in identifying the right resources needed for digitalization, as well as adequate management of the right digital transformation teams.

The scope of future work is to identify key success factors for the construction sector to maximize the potential benefits of digitalization with special with special attention to companies operating in the Bulgarian construction sector, which is currently again in a stage of particularly intense development, supported by the growth of the Bulgarian economy.

ACKNOWLEDGEMENTS

The research is supported by the Project BG05M2OP001-1.002-0002-C02 Digitization of the economy in an environment of Big data

(DEBD) with basic organization and beneficiary University of National and World Economy and the partnership of University of Economics – Varna.

REFERENCES

1. AGHIMIEN, D., AIGBAVBOA, C., OKE, A., KOLOKO, N. (2018) Digitalisation in Construction Industry: Construction Professionals Perspective, Streamlining Information Transfer between Construction and Structural Engineering, Shiau, J., Vimonsatit, V., Yazdani, S., and Singh, A. (Eds.), ISEC Press.
2. ANWAR, N., IZHAR, M. A., NAJAM, F. A. (2018) Construction Monitoring and Reporting using Drones and Unmanned Aerial Vehicles (UAVs). *Proc. of The Tenth International Conference on Construction in the 21st Century (CITC-10)*, July 2nd-4th, 2018.
3. BERGER, R. (2016). Digitization in the Construction Industry, *Think Act*, 1-16.
4. CAPGEMINI CONSULTING (2013) Accelerating Digital Transformation, Digital Transformation Review, Available from: <https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2017/08/nummer-4-digital-transformation-review.pdf>, [Accessed 03/05/2019].
5. EZEOKOLI, F. O., OKOLIE, K.C., OKOYE, P. U., BELONWU, C. C. (2016) Digital Transformation in the Nigeria Construction Industry: The Professionals' View. *World Journal of Computer Application and Technology*, 4(3), 23-30.
6. JAYARAJ, A., DIVAKAR, H. N. (2018) Robotics in Construction Industry, *IOP Conference Series: Materials Science and Engineering*, 376.
7. LEDIGA, P., KRUGER, D. (2017) Optimizing Concrete Mix Design for Application in 3D Printing Technology for the Construction Industry, *Solid State Phenomena*, 263, 24-29.
8. LIU, P., CHEN, A. Y., HUANG, Y. N., HAN, J. Y., LAI, J. S., KANG, S. C. (2014) A Review of Rotorcraft Unmanned Aerial Vehicle (UAV) Developments and Applications in Civil Engineering. *Smart Structures and Systems*, 13(6), 1065–1094.

9. MCKINSEY (2016) Imagining construction's digital future. Available from: <https://www.mckinsey.com/~/media/McKinsey/Industries/Capital%20Projects%20and%20Infrastructure/Our%20Insights/Imagining%20constructions%20digital%20future/Imagining-constructions-digital-future.ashx>, [Accessed 02/05/2019].
10. SOLIS, B., LI, C., SZYMANSKI, J. (2014) Digital Transformation: Why and How Companies are Investing in New Business Model to Lead Digital Customer Experience, Altimeter Group, Available from: <https://www.socialmediatoday.com/content/report-why-and-how-businesses-are-investing-digital-transformation-and-new-customer>, [Accessed 07/05/2019].
11. WORLD ECONOMIC FORUM (2017) Shaping the Future of Construction: A Breakthrough in Mindset and Technology. Geneva. Available from: http://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction_Inspiring_Innovators_redefine_the_industry_2017.pdf, [Accessed 2/08/2019].

THE USAGE OF DATA LAKE FOR BUSINESS INTELLIGENCE DATA ANALYSIS

Snezhana Sulova¹

*¹ University of Economics – Varna/ Department of Informatics,
Varna, Bulgaria, ssulova@ue-varna.bg*

Abstract

Data analysis is now becoming increasingly more important for business. The accumulation of large amounts of different types of data in organizations is a prerequisite for seeking new ways of storing, processing and analyzing them. The following paper presents the nature of the data lake concept and examines its capabilities to organize all the data, both those generated by the organization and those extracted from Internet sources. Storing large amounts of data, regardless of its type, structure, or format, allows for the integrated use of structured and unstructured data and the application of a variety of techniques for intelligent business analysis.

Keywords: *data lake, big data, data analysis, BI, data warehouse.*

INTRODUCTION

Today's dynamic society is increasingly more tied to internet technologies and intelligent systems. The advancement in information technologies and the development of artificial intelligence are leading to the creation of new ways of working and remotely accessing devices. More and more physical devices have built-in electronic elements, software that allow them to be connected to the Internet and receive, collect and exchange data. Although organizations currently work mainly with information systems in which the data is organized and managed by database management systems (DBMS), according to research, unstructured data makes up about 80% of all the information resources in them (Grimes, 2008). A research on the topic of big data indicates that unstructured data tends to grow exponentially in number and that they represent 95% of new data, a large part of which is not processed or used (Minelli, et al., 2013, p. 11).

It is well known that in order for data to be converted into information and business knowledge, it must be transformed and organized

with a specific purpose. The analysis process is becoming increasingly more complex and involves not only the generation of reports using SQL queries and the calculation of statistical dependencies, but also data mining technologies. Another tendency has been noted which involves looking for and adding new additional sources of data for processing, as well as improving the approach and technologies used to store and retrieve this data. Business intelligence (BI) is a top priority for the organizations in most industries (Richards, et al., 2019). In regard of that, this report aims to clarify the essence of the data lake (DL) data storage concept and to demonstrate its capabilities for supporting smart business analyses.

1. THE ESSENCE AND CONCEPT OF DATA LAKE

James Dixon, CTO of Pentaho introduces a new big data storage concept called data lake (Dan, 2011). The main idea is to store different types of data in a relatively cheap way and then to apply ETL functions (extraction, transformation, loading) to them. “A data lake is a central location in which to store all your data, regardless of its source or format” (Laplante and Sharma, 2016, p.2). IBM researchers say the concept is designed to provide storage of virtually inexhaustible materials in the form of raw data that analysts have easy access to (IBM Corporation, 2016).

DL can be defined as a data storage strategy that provides flexibility for organizations when working with data. It allows the same data to be structured and processed differently, and this is essential for processing unstructured data where there are no well-defined algorithms for data extraction, processing and analysis, and different approaches are usually used.

The DL concept also allows data to be stored from external Internet sources, such as social networks, devices using the Internet of things (IoT) concept, and other unstructured corporate data (Fig. 1).

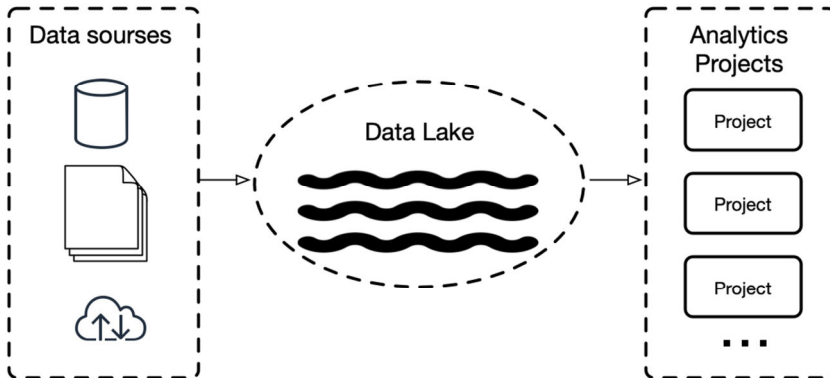


Figure 1. A common view of data lake

The following main advantages of the data lake concept can be pointed out:

- data from different types – structured, semi-structured and un-structured, can be stored;
- the types of data than can be extracted are an infinite amount;
- data can be stored in its raw form, which allows its conversion when it is needed;
- various tools can be used for extracting and processing the data;
- all data in the organization can be stored in one place, which allows a unified view of the data.

There are also several challenges associated with data lake, which is why there is still debate in literature whether this concept is necessary and useful for business. Some of the concerns connected to it include the following:

- a bad quality of data because they are received without supervision and control;
- the data process needs to be started from scratch with every data analysis;
- the performance of data operations is usually not guaranteed;
- weaknesses regarding the security and control over the access to the data;

- a danger of the data being turned into a “swamp” because they are stored without being sorted and organized into topics, categories and without maintaining metadata for them.

It should be noted that the use of DL cannot replace the traditional data warehouse (DW), which aims to integrate large-scale enterprise data into a united storage. According to Bill Inmon, DW “is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management’s decision-making process.” (Inmon, 2002, p. 31). Usually DWs are subject oriented. It is necessary to select and extract the data from the company databases and then transform, reformat and store it (Curtis and Cobham, 2008, p. 247). Therefore, creating a data warehouse takes time and resources in order to model and prepare the data. The knowledge and skills of the specialists are also important for its successful creation (Marinova, 2016). The benefits of data warehouses are numerous: they save time for users, improve decision-making processes, and help achieve strategic business goals.

Organizations looking to upgrade their analytics platforms could use both the DL and the DW storage concepts. This working method will allow them to explore how traditional analytical architectures work with new storage methods that include both relational databases and NoSQL databases. According to research, for every type of DBMS, there is an effective strategy for archiving and restoring databases, which is of crucial importance (Kuyumdzhev, 2019).

2. DATA LAKE AND BUSINESS INTELLIGENCE DATA ANALYSIS

Business intelligence can be viewed as a generic term for a set of approaches that serve to analyze the activities and functioning of an organization and support the decision-making process (Curtis and Cobham, 2008, p. 228). The goal of BI is to enable the processing of this large volume of data easily, to support the search of new opportunities for development and to build effective knowledge-based business strategies.

Business intelligence systems are constantly evolving, new functionalities are being added to them (Todoranova, 2013), they are evolving from single applications to large-scale business Intelligent Ecosystems (Kisimov and Stefanova, 2010). Their main components are: DW; ETL tools, OLAP techniques and data mining tools.

Trends in the development of BI systems indicate that the application of the data mining technology will expand and will be applied more and more to unstructured data (cio.bg, 2018). The World Wide Web has become one of the richest sources of data. Companies have started using data mining technologies more to extract knowledge from Internet sources or the so-called Web Mining to increase the precision of their business analytics. There are many documents, data, audio and video files on the web that can be used to extract new and useful business knowledge through appropriate processing. The knowledge is generated not only by the content of the web pages themselves, but also by their unique features, the structure of the web sites and the information connected to accessing them.

BI methods are usually applied when analysing structured datasets from a specific type of business, e.g. banking or credit institutions (Vasilev, Stoyanova, Stancheva, 2017, 2018). Time series data are quite popular for storing data in different types of businesses. In these cases, BI methods are applied to data stored in relational databases. But the real view of business needs more information on opinions, moods. In this case data from data lakes are useful. This fact shows the need to create, test and apply BI methods for analyzing data stored in data lakes.

The data extracted from the Internet are in most cases unstructured, and performing an automatic analysis, generating summaries, classifications, trending and anomalies requires that they be pre-processed and given a certain structure. Processing unstructured data is not an easy task (Bankov, 2018). It is very often necessary to use different approaches that require different sections of the data. Business analytics also require rebuilding of business rules and the need to use unconverted or so-called raw data. Therefore, we believe that creating and maintaining a data lake data storage in the case of large heterogeneous data is a good base for modern BIs that are focused on providing machine learning, NLP and AI to their customers.

Using DL together with DW offers a modern and optimal basis for data analysis, as shown in Figure 2. Data from enterprise databases that are collected as a result of multiple applications in enterprise information systems are the basis for the creation of a DW, which aims to store and track historical, archival information, consolidate large volumes of data

from various subsystems, analyzes and forecasts. Unstructured and semi-structured data from other external Internet sources, such as social networks and devices using the Internet of things (IoT) concept, server log files, and more are entered and stored in a DL. In this way, the necessary data will be provided for each analytical process. Integrated data will also be able to be used, which is a good basis for obtaining more detailed and in-depth analyzes and will help to make informed decisions.

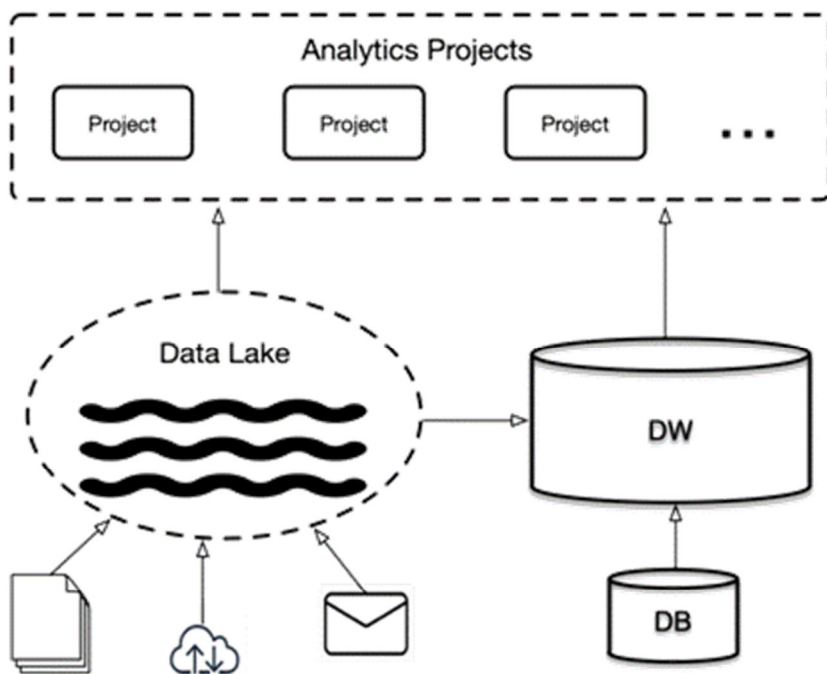


Figure 2. Use of DL and DW

The main advantages of the BI approach based on the combined use of DW and DL are the following:

- in addition to storing traditional structured data, DL allows the cheap storage of all types of data (including audio and video formats) coming from Internet applications, social networks and from various devices;

- various and optimal data analysis approaches can be used, including those that work with unstructured data – processing text, audio, video;

- enables real-time data extraction, rapid data analysis and the implementation of agile analytics schemes;

- allows working with large volumes of data, for example, which is a good basis for predicting possible future states and processes.

Although as mentioned above, there are a number of problems with the use of DL, it is considered that creating a single integrated data management framework, where they are managed with metadata that helps to find and connect information, can build a successful model for integrated data storage and management. Such a model is a good basis for conducting numerous analyzes and improving the BI strategies of companies. It would also allow the implementation of Agile BI, which is built on the idea of flexible analysis and adaptation to specific needs and is responsive to rapidly changing business conditions. A proper implementation of the DL concept would be in favor of adhering to one of the basic principles of Agile BI, which is to provide the right data at the right time for the correct analytical process.

It is proved that good organization of projects in the first phases of product development reduces errors and costs of subsequent stages (Nacheva, 2015).

We believe that the main conditions for the successful implementation of DL as a data source are the following

- clearly defining the need and purpose of using the DL;
- creating and following a data management strategy in DL;
- creating procedures for security and control over the access and use of the data;
- building DL as a new, additional source of data for analysis, rather than as a sole component of BI infrastructure.

Hadoop's distributed file system (High-availability distributed object-oriented platform) is currently considered the most popular DL build technology (Khine and Wang, 2018; Yordanova and Stefanova, 2019). Hadoop is a software framework managed by the Apache Software Foundation and designed to organize a distributed processing of big data

when using the MapReduce programming model. It is used to build large-scale projects on Yahoo!, Facebook, Oracle, in IBM's Watson supercomputer, and in Azure Cloud.

CONCLUSION

Growing volumes of heterogeneous data are a prerequisite for finding ways to extract, process and store them. Combining data across systems represents a challenge for many organizations when making management decisions. Using the DL concept can bring benefits for organizations that want to avoid the costly and cumbersome process of pre-processing storage data in a data warehouse. DL is a good approach for storing big data, but it should be noted that an incorrect design and use of it carries risks related to data quality, security and control of their access and usage. Well-trained professionals are needed to properly anticipate and plan the data lake so that it can help organizations successfully manage structured and unstructured data.

REFERENCES

1. BANKOV, B. (2018) An Approach for Clustering Social Media Text Messages, Retrieved from Continuous Data Streams. *International Scientific Journal "Science. Business. Society"*, 3(1), pp. 6-9.
2. CIO.BG (2019) *11 тенденции в развитието на платформите за бизнес анализи*. [Online] Available from: https://cio.bg/softuer/2008/05/12/3450655_11_tendencii_v_razvitiето_na_platformite_za_biznes/ [Accessed 31/07/2019].
3. CURTIS, G. and COBHAM, D. (2008) *Business Information Systems: analysis, design and practice*, sixth edition, Canada: Pearson Education.
4. DAN, W. (2011) *Big Data Requires a Big, New Architecture* [Online] Available from: <https://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/#37aa44d31157> [Accessed 22/07/2019].
5. GRIMES, S. (2008) *Unstructured data and the 80 percent rule*. [Online] Available from: <https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/> [Accessed 22/07/2019].

6. IBM Corporation (2016), *IBM Industry Model support for a data lake architecture*. [Online] Available from: <https://www.ibm.com/downloads/cas/DNKPJ80Q>, [Accessed 22/07/2019].
7. INMON, B. (2002) *Building the Data Warehouse*. 3rd Ed. Toronto: John Wiley & Sons, Inc.
8. KHINE, P. and WANG, Z. (2018) Data lake: a new ideology in big data era, *ITM Web of Conferences 17*, WCSN 2017, [Online] Available from: https://www.itm-conferences.org/articles/itmconf/abs/2018/02/itmconf_wcsn2018_03025/itmconf_wcsn2018_03025.html [Accessed 20/07/2019].
9. KISIMOV, V. and STEFANOVA, K. (2010) Design Principles and Methods for Distributed Business Intelligence Ecosystem, *Yearbook of UNWE, 1*, pp. 215-250.
10. KUYUMDZHIEV, I. (2019) Comparing Backup and Restore Efficiency in Mysql, MS SQL Server and MongoDB. *19 International Multidisciplinary Scientific Geoconference SGEM 2019*, Albena, Bulgaria, 19(2.1), pp. 167-174.
11. LAPLANTE, A. and SHARMA, B. (2016) *Architecting Data Lakes. Data Management Architectures for Advanced Business Use Cases*. O'Reilly Media, Inc.
12. MARINOVA, O. (2016) Business intelligence and data warehouse programs in higher education institutions: current status and recommendations for improvement. *Economics and computer science: Electronic journal*, 5, pp. 17-25.
13. MINELLI, M., CHAMBERS, M. and DHIRAJ, A. (2013) *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*. John Wiley & Sons, Inc.
14. NACHEVA, R. (2015) Principles of User Interface Design: Important Rules that Every Designer Should Follow. *Izvestia, Journal of the Union of Scientists – Varna. Economic Sciences*, pp. 140–149.
15. RICHARDS, G., YEOH, W., CHONG, A. Y. L. & POPOVIC, A. (2019) Business Intelligence Effectiveness and Corporate Performance Management: An Empirical Analysis. *Journal of Computer Information Systems*. 59(2). pp. 188-196.

16. TODORANOVA, L. Trends in the Development of Business Intelligence Systems. *Izvestia – Journal of University of Economics – Varna*, 1, pp 98-106.
17. YORDANOVA, S. and STEFANOVA, K. (2019) Big Data Challenges – Definition, Characteristics and Technologies, *The Scientific Papers of UNWE*, 1, pp. 13-31.
18. VASILEV, J., STOYANOVA, M., STANCHEVA, E. (2017) Business intelligence data analysis of a loan dataset. In *2nd conference on innovative teaching methods (ITM 2017)*, Varna: Science and economics, pp. 17-23.
19. VASILEV, J., STOYANOVA, M., STANCHEVA, E. (2018) Application of business intelligence methods for analyzing a loan dataset. *Informatyka Ekonomiczna*, 1, 47, pp. 97-106.

WEB SECURITY TECHNOLOGIES USED IN BANKS OF ESTONIA, LATVIA AND LITHUANIA

Pavel Petrov¹

*¹ University of Economics – Varna/Department of Informatics,
Bulgaria, petrov@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 this to be done a valid certificate issued by authority body should be used. In the scope of the study in the summer of 2019 we examined the web sites of banks licensed in Estonia, Latvia and Lithuania. The survey excludes the foreign bank branches, because we try to outline the "good practices" used by domestic administrators of banking websites.

Keywords: *Estonian banks; Latvian banks; Lithuanian banks; HTTPS; SSL/TLS certificates.*

INTRODUCTION

In this comparative study we choose banks which are regulated by local central banks and belong to three neighbour European countries, located in Northern Europe on the eastern coast of the Baltic Sea - the so called "Baltic states" - Estonia, Latvia, and Lithuania. These countries have a lot of similarities in demographics, economics and politics characteristics - they are members of the European Union and NATO. In their financial system they are using Euro as currency and are members of the Eurozone.

In Table 1 are summarized some overall data for countries which have general meaning in the context of the current study. It should be noted that there are some variations. For example the Gross Domestic Product per capita indicator of Estonia is the largest, but the Gini Index indicator is the lowest. This could be happen because of different metrics and the time lag in collected data. In general the three countries are very similar.

Before the comparison, we raise the hypothesis that in such very similar countries, with similar level of living standard, it should be ex-

pected that the web technologies used in local banks’ web sites should be also similar. In these institutions usually there is no problem with funding, and there are opportunities to use expensive software.

Table 1

Large-scale overall comparison between the Baltic States

Feature	Estonia	Latvia	Lithuania
Population [millions]	1.3	1.9	2.8
Area [km ²]	45 339	64 589	65 300
GDP (nominal) per capita (2018) [€]	19 500	15 300	16 100
Gini Index (2015)	32.7	34.2	37.4
HDI (2018)	0.871 (Very High)	0.847 (Very High)	0.858 (Very High)

Sources: Wikipedia 2019; Eurostat, 2019; World Bank DRG, 2019; UN HDRO, 2019.

1. METHODOLOGY AND EMPIRICAL RESULTS

In our study home pages of 9 Estonian, 14 Latvian and 4 Lithuanian banks were inspected in August 2019. The main method used in the survey includes analysis of the responses given by the web servers. An up to date browser Google Chrome Version 76.0.3809.100 (Official Build) (64-bit), working under typical desktop PC with Windows 10 Professional Edition x64, was used as a web client with "Developer tools" module activated. 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", but using real web browser is more straightforward. The methodology of the study is based partially on methodology used in previous studies (Petrov, 2018/19) on web technologies used in banks.

The lists of banks authorized to operate in Estonia, Latvia and Lithuania (Table 2) were taken from the websites of Estonia Finan­santsinspektsioon, Latvia Financial and Capital Market Commission and Bank of Lithuania (see reference list at the end). In this study websites of foreign bank branches and representative offices of foreign banks operat-

ing in the local financial markets are excluded. We surveyed only domestic ones, which operate under regulation of the domestic authority body. So the websites of those banks that operate on an EU branch or on an EU cross-border basis are not included.

Table 2

**HTTPS protocol usage in public web sites of banks
in the Baltic States**

№	Bank Name	Bank domain	HTTPS
Estonia			
1	AS Inbank	www.inbank.ee	yes
2	AS LHV Pank	www.lhv.ee	yes
3	AS Luminor Bank	www.luminor.ee	yes
4	AS SEB Pank	www.seb.ee	yes
5	AS TBB pank	www.tbb.ee	yes
6	Bigbank AS	www.bigbank.ee	yes
7	Coop Pank aktiaselts	www.cooppank.ee	yes
8	Holm Bank AS	www.holmbank.ee	yes
9	Swedbank AS	www.swedbank.ee	yes
Latvia			
1	AS Baltic International Bank	www.bib.lv, www.bib.eu	yes
2	AS Citadele banka	www.citadele.lv	yes
3	AS LPB Bank	www.lpb.lv	NO
4	AS Reģionālā investīciju banka	www.ribbank.com	NO
5	AS Rietumu Banka	www.rietumu.lv	yes
6	AS Meridian Trade Bank	www.mtbank.eu	yes
7	AS PrivatBank	www.privatbank.lv	yes
8	AS BlueOrange Bank	www.blueorangebank.com	yes
9	AS Expobank	www.expobank.eu	yes
10	AS PNB Banka	www.pnbbanka.eu	yes
11	AS SEB banka	www.seb.lv	yes
12	Rigensis Bank AS	www.rigensisbank.com	yes, but redirects to HTTP

13	Signet Bank AS	www.signetbank.com	yes
14	Swedbank AS	www.swedbank.lv	yes
Lithuania			
1	Akcinė bendrovė Šiaulių bankas	www.sb.lt	yes
2	AB SEB bankas	www.seb.lt	yes
3	Swedbank AB	www.swedbank.lt	yes
4	UAB Medicinos bankas	www.medbank.lt	yes

The summarized results of the studied home web pages are presented in the next table (Table 3) based on the following key indicators: presence of automatic redirection to HTTPS, certificate type, the name of certification body and validity period of the SSL certificate.

Table 3

**Main features in usage of the HTTPS in public web sites
of banks in the Baltic States**

Bank №	Automatic redirection to HTTPS	Certificate type	Certification authority body	Validity
Estonia				
1	yes	DV	Sectigo RSA Domain Validation Secure Server CA	2 y. 2 m.
2	yes	EV	DigiCert SHA2 Extended Validation Server CA	2 y. 3 m.
3	yes	DV	Let's Encrypt Authority X3	3 m.
4	yes	EV	GlobalSign Extended Validation CA - SHA256 - G3	1 y. 1 m.
5	NO	DV	DigiCert SHA2 Secure Server CA	2 y. 1 m.
6	yes	EV	DigiCert SHA2 Extended Validation Server CA	2 y. 2 m.
7	yes	DV	Amazon	1 y. 1 m.
8	yes	EV	Sectigo RSA Extended Validation Secure Server CA	1 y.
9	yes	EV	DigiCert SHA2 Extended Validation Server CA	1 y.

Latvia				
1	yes	EV	DigiCert SHA2 Extended Validation Server CA	2 y. 1 m.
2	yes	EV	Thawte EV RSA CA 2018	2 y.
5	yes	DV	DigiCert SHA2 Secure Server CA	2 y. 2 m.
6	yes	EV	Thawte EV RSA CA 2018	2 y. 1 m.
7	yes	DV	Go Daddy Secure Certificate Authority - G2	1 y. 1 m.
8	yes	DV	DigiCert SHA2 Secure Server CA	2 y.
9	yes	DV	Let's Encrypt Authority X3	3 m.
10	yes	DV	Thawte RSA CA 2018	1 y.
11	yes	DV	GlobalSign Organization Validation CA - SHA256 - G2	2 y. 1 m.
12	NO	DV	DigiCert SHA2 Secure Server CA	1 y. 3 m.
13	yes	DV	Go Daddy Secure Certificate Authority - G2	2 y.
14	yes	EV	DigiCert SHA2 Extended Validation Server CA	1 y.
Lithuania				
1	yes	DV	Thawte TLS RSA CA G1	1 y. 7 m.
2	yes	DV	GlobalSign Organization Validation CA - SHA256 - G2	2 y. 1 m.
3	yes	EV	DigiCert SHA2 Extended Validation Server CA	1 y.
4	yes	DV	COMODO RSA Domain Validation Secure Server CA	3 y.

2. COMPUTATIONAL DETAILS AND DISCUSSION

Three Latvian banks web sites are not using HTTPS (№3, №4 and №12) - two are not using HTTPS at all (№3 and №4) and in one case (№12) HTTPS requests are redirected to use HTTP. The number of banks web sites not using HTTPS is not high, but this situation is quite strange, because the prices for a simple DV certificate starts at around

30€ per year and also there is a free alternative. Well reputable 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, with the main goal to issue free certificates. These certificates are currently valid for 3 months. The so-called "wildcard certificates" covering all subdomains of a domain was introduced in 2018. One Estonian (№3) and Latvian (№9) banks are using free certificates from Let's Encrypt Certificate Authority.

About the case of redirecting the HTTPS requests to use HTTP (Latvian bank №12) it will be better either to support HTTPS according to the good practices or not to use HTTPS at all, because these problems could weaken the confidence of customers in the bank's capability to keep up to date its systems.

They are three types of certificates: Domain Validated (DV), Organization Validated (OV), and Extended Validated (EV) (Cooper, 2008; Saint-Andre, 2011). 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 that 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. Because there is no sure way to tell with confidence if a SSL certificate is Domain Validated or Organization Validated, in this research we didn't provide separation between them. 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 connection is protected. In general, the DV certificate is cheaper than EV certificate, but EV certificates are more prestigious.

From the data presented in Table 3 and aggregated in Table 4, it is clear that the majority - 15 banks are using simple DV certificates, and only 10 banks are using the more complicated for issuance EV. Only in Estonia the majorities (56%) of banks are using EV certificates while in Latvia and Lithuania the situation is the opposite - the majority (57% in Latvia and 75% in Lithuania) uses DV.

Table 4

**Type of the SSL certificates in public web sites of banks
in the Baltic States**

Certificate type	Estonia		Latvia		Lithuania	
	Count	%	Count	%	Count	%
No certificate	-	-	2	14	-	-
DV	4	44	8	57	3	75
EV	5	56	4	29	1	25
<i>TOTAL</i>	9	100	14	100	4	100

In two cases - one in Estonia, one case in Latvia the good practices are not followed and HTTP requests are not automatically redirect to use secure HTTPS connection.

There is a wide variety of preferences for a certification authority, but the most popular choices in Baltic States are:

- DigiCert - 10 banks;
- Thawte - 4 banks;
- GlobalSign - 3 banks;
- Sectigo/Comodo (Sectigo is formerly Comodo) - 3 banks;
- Go Daddy - 2 banks;
- Let's Encrypt (free of charge 3 months-long certificates.) - 2 banks;
- Amazon - 1 bank.

The data about certification authority which is presented in Table 3 are aggregated for convenience in Table 5 and represented on Figure 1.

Table 5

**The issuers of the SSL certificates used in public web sites
of banks in the Baltic States**

Certification authority body	Estonia		Latvia		Lithuania	
	Count	%	Count	%	Count	%
Amazon	1	11	-	-	-	-
DigiCert	4	45	5	42	1	25

GlobalSign	1	11	1	8	1	25
Go Daddy	-	-	2	17	-	-
Let's Encrypt	1	11	1	8	-	-
Sectigo/Comodo	2	22	-	-	1	25
Thawte	-	-	3	25	1	25
<i>TOTAL</i>	<i>9</i>	<i>100</i>	<i>12</i>	<i>100</i>	<i>4</i>	<i>100</i>

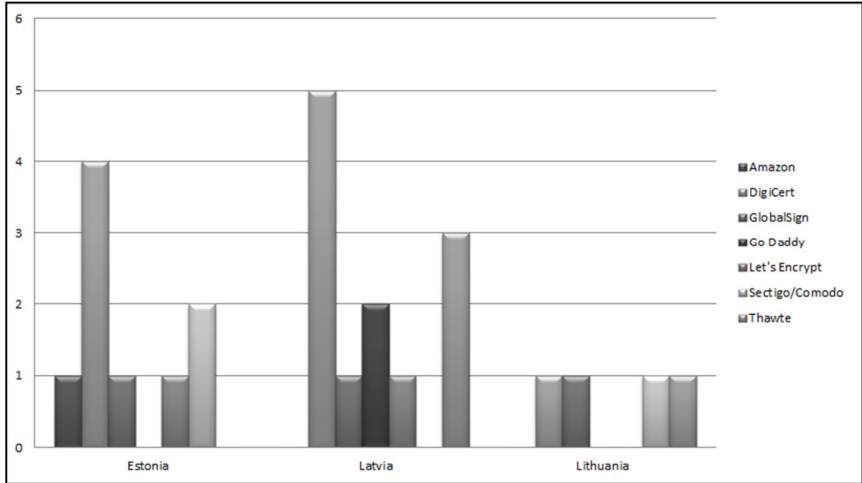


Figure 1. Certification authorities used by banks in the Baltic States

CONCLUSION

This research leads to the following conclusions. First, the banks sector in Lithuania, which is the largest Baltic state, is more consolidated than this in Estonia and Latvia. Second, as for the use of SSL certificates the most popular SSL certificate provider is DigiCert with share of 40% web sites. It is interesting that one Estonian and one Latvian bank are using free certificates from Let's Encrypt Authority. One bank in Estonia, two banks in Latvia are not redirecting automatically from unsecure HTTP to secure HTTPS connection. Three banks in Latvia are not using SSL at all or redirect secure HTTPS requests to unsecure HTTP connection. The last one we consider as a very bad practice.

The average validity of certificates is 1 year and 7 months with median - 1 year and 11 months.

The collected data are related to particular period - August 2019. The results of the study could have important practical impact for banks managers and IT specialist when evaluating options which technologies to implement in order to minimize the risk to the financial institution. Also the results reveal some good and bad practices used in the Baltic States banks. The research conducted on the use of the HTTPS protocol on the banks' public web sites covered the sites of all 9 Estonian, 14 Latvian and 4 Lithuanian banks licensed to operate on the respective country territory by the domestic National Banks or other government institution.

REFERENCES

1. BANK OF LITHUANIA (2019). Supervision of financial market participants. Banks authorised in the Republic of Lithuania. [Online] Available from: https://www.lb.lt/en/sfi-financial-market-participants?business_form=82&market=1 [Accessed 07/08/2019].
2. COOPER, D., et al. (2008) Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile. *RFC5280*. [Online] Available from: https://www.ietf.org/rfc/rfc_5280.txt [Accessed 07/08/2019].
3. ESTONIA FINANTSINSPEKTSIOON (2019). Supervised Entities. Licensed credit institutions in Estonia [Online] Available from: <https://www.fi.ee/en/banking-and-credit/banking-and-credit/credit-institutions/licensed-credit-institutions-estonia> [Accessed 07/08/2019].
4. EUROSTAT (2019), Gross domestic product at market prices. *Eurostat Database* [Online] Available from: <https://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tec00001> [Accessed 07/08/2019].
5. LATVIA FINANCIAL AND CAPITAL MARKET COMMISSION (2019). Market. Credit institutions. Banks. [Online] Available from: <https://www.fktk.lv/en/market/credit-institutions/banks/> [Accessed 07/08/2019].

6. PETROV, P., BUEVICH, A., DIMITROV, G., KOSTADINOVA, I. and DIMITROV, P. A Comparative Study on Web Security Technologies Used in Bulgarian and Serbian Banks. *19 International Multidisciplinary Scientific Geoconference SGEM 2019, Vol. 19, Issue. 2.1*, Sofia: STEF92 Technology Ltd., 2019, pp.3-10.
7. PETROV, P., DIMITROV, G. and 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*, Sofia: STEF92 Technology Ltd., 2019, pp.3-10.
8. SAINT-ANDRE, P. and HODGES, J. (2011) 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). *RFC6125*. [Online] Available from: <https://tools.ietf.org/rfc/rfc6125.txt> [Accessed 07/08/2019].
9. UN HUMAN DEVELOPMENT REPORT OFFICE (2019), Human Development Indices and Indicators. *2018 Statistical update*. [Online] Available from: <http://hdr.undp.org/en/2018-update> [Accessed 07/08/2019].
10. WIKIPEDIA (2019), Baltic states. *Wikipedia.org*. [Online] Available from: https://en.wikipedia.org/wiki/Baltic_states [Accessed 07/08/2019].
11. WORLD BANK DEVELOPMENT RESEARCH GROUP (2019), GINI index (World Bank estimate). *International Comparison Program database*. [Online] Available from: <https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD> [Accessed 07/08/2019].

ABC ANALYSIS OF INVENTORY IN MS EXCEL

Tanka Milkova¹

*¹ University of Economics – Varna,
Bulgaria, tankamilkova@ue-varna.bg*

Abstract

The technologies in IT industry take important part in all the areas of business and education in the contemporary world. It is hard to imagine for most of the activities to go smoothly without computer and software. The purpose of this paper is to present an idea for inventory ABC analysis. A sample dataset is used. Classification of inventory by empirical and graphical method in MS Excel is carried out. The proposed methodology can be used in the educational processes for students and also for real practical applications in inventory management in logistics.

Keywords: *ABC analysis, MS Excel.*

INTRODUCTION

Inventory management is a combination of complex processes and activities which are important elements in the logistical system of each organisation in the contemporary economy. Inventories are found in every logistics system as they represent the static state of material flow and they are major management objects in logistics. This determines their importance and is the main reason for the continuous theoretical and practical interest in the problems related to their management. There are different concepts and methods for inventory management.

However, the first step in inventory management is to classify their nomenclature and to group them and for each of these a set of management models and methods is recommended. The classical methods for classifying the inventories' nomenclature are the well-known ABC analysis method and XYZ analysis method (Dybskaya, V. V., etc., 2008, p. 594 - 609; Sterligova, A. N., 2012, p. 353 – 376; Sergeev, V. I., etc., 2005, p. 535 – 545.) and some of their more recent modifications which are known in the literature, such as FSN, VED, HML, SDE analysis methods, etc. (<<https://www.syntacticsinc.com/news-articles-cat/inventory-analysis-methods/>> [Accessed 23/08/2019]; <<https://scialert.net/fulltext>

[/?doi=ajava.2018.85.90](https://doi=ajava.2018.85.90)> [Accessed 23/08/2019]; <<https://mywestford.com/blog/fsn-analysis-in-inventory-management/>> [Accessed 23/08/2019]; <<https://www.unleashedsoftware.com/blog/fsn-analysis-can-useful-tool-inventory-management>> [Accessed 23/08/2019]).

The application of ABC analysis is a laborious process, especially in the presence of a volumetric inventory. With the help of information technology, this method can be applied without serious difficulty.

In this report the author aims to present an opportunity to perform an ABC analysis of inventories' nomenclature using MS Excel. The proposed method can be used both in the students teaching process and in the actual inventory management in logistics.

1. ESSENCE OF ABC ANALYSIS AND METHODS OF CLASSIFICATION OF INVENTORY NOMENCLATURE

The ABC method of analysis is based on the well-known in the economy Pareto rule or 80/20 rule. According to the 80/20 rule, 20% of the elements composing a phenomenon determine 80% of its occurrence. With regard to stock inventories, the 80/20 rule can be interpreted in the following ways: 20% of the maintained inventories are related to 80% of the total cost of purchasing all types of inventories; 20% of the maintained inventories are related to 80% of the total management costs of all types of inventories, etc.

The ABC method is a way of analysing and controlling the status of inventories, which is boiled down to the distribution of the nomenclature of all stock inventories into three groups A, B and C on the basis of some formal algorithm.

The first stage of the application of this method involves the choice of a classification criterion. The classification criterion should be chosen in accordance with the objectives of inventory management and may be: total value of inventories of each kind; total costs associated with each type of inventory; profitability of each type of inventory, etc.

The second stage of the implementation of the ABC method is connected with the ordering of the nomenclature of stock inventories in descending order according to the selected classification criterion. For that purpose, it is necessary to determine the importance of each item in the inventory nomenclature according to the chosen classification criterion.

For example, if a classification of the total value of inventories is made, the monetary total value of inventories should be determined from each item of the nomenclature corresponding to the average level of inventories of that type, etc.

In the third stage, the nomenclature of all inventories is divided into three groups A, B and C, and subsequently, for each of the groups of inventories appropriate management methods are set out, which are inspected in detail in the specialized literature. In Group A, inventories are of the highest significance and importance to the organization, group B refers to inventories of intermediate importance, and group C refers to inventories that are the least important to the organization.

In the literature (Lukinsky, V. S., 2007, p. 3 – 20; Lukinsky, V. S., 2003, p.26 – 38.) there are three basic methods for classification of the stock nomenclature - empirical method, differential method and graphical method. The following two will be used in this study:

- An empirical method. It is the easiest one to use and is expressed in the following. In Group A, the first 20% of the descending order of inventory is regarded. Group B accounts for the next 30% of the descending order of inventory nomenclature. Group C refers to the last 50% of the descending order of inventory.

- Graphical method. In this method, after the stock inventory nomenclature is arranged in descending order according to the chosen criterion, the relative indexes q_i (в %) are calculated - the relative share of the i -th type in the total value of all stock inventories according to the selected criterion ($i = 1, 2, \dots, N$, where N is the number of inventories in the nomenclature).

It is calculated by the formula

$$q_i = \frac{c_i}{\sum_{i=1}^N c_i} \cdot 100, \quad (1)$$

where c_i denotes the i -th inventory type according to the selected classification criterion.

The quantities q_i are summed up in increasing order and a summary part of the total value of the inventory in percentage $\sum q_i$ is obtained, according to the chosen criterion. In the graphical method, the obtained quantities are plotted on the ordinate axis of the coordinate system, and

the indexes $1, 2, \dots, N$ are plotted on the abscissa. The points with coordinates $(I; \sum q_i)$ of the coordinate system merge lightly with the cumulative curve $OO'D$ (т. $O(0;0)$, $D(N; \sum q_N)$, $\sum q_N = 100\%$, which is generally concave. Then a tangent l_1 is constructed to the curve $OO'D$, which is parallel to OD . The abscissa of the tangent point O' shows the boundary between groups A and B, and the ordinate of this point shows the total part of the inventories of Group A according to the chosen classification criterion.

In order to continue dividing the stock inventory nomenclature into groups, it is necessary to connect the points O' and D with a straight line, and then construct a tangent l_2 to the curve $OO'D$, parallel to $O'D$. The abscissa of the tangent point O'' shows the boundary between groups B and C, and the ordinate of this point shows the total part of inventories of groups A and B according to the chosen classification criterion.

2. EXECUTING ABC ANALYSIS IN MS EXCEL

The classification of the inventory's nomenclature by empirical and graphical method in the MS Excel will be presented on the basis of sample data. A 25-items nomenclature has been intentionally chosen in order to be able to visualize the results of each stage.

Let's assume that information for the unit price and average inventory level of each type is available (fig.1). The total value of the inventories of each item in the nomenclature is defined as a product of price and quantity.

S30						
	A	B	C	D	E	F
1	Inventory					
2	No	Name	Price (lv.)	Quantity	Total value (lv.)	
3	1	Stock 001	0,29	45	13,05	
4	2	Stock 002	0,48	18	8,64	
5	3	Stock 003	6,15	150	922,5	
6	4	Stock 004	6,15	125	768,75	
7	5	Stock 005	4,00	175	700	
8	6	Stock 006	5,75	250	1437,5	
9	7	Stock 007	2,98	72	214,56	
10	8	Stock 008	8,32	18	149,76	
11	9	Stock 009	2,50	5	12,5	
12	10	Stock 010	3,10	25	77,5	
13	11	Stock 011	1,33	16	21,28	
14	12	Stock 012	1,30	3	3,9	
15	13	Stock 013	0,16	5	0,8	
16	14	Stock 014	0,43	107	46,01	
17	15	Stock 015	3,33	4	13,32	
18	16	Stock 016	2,31	9	20,79	
19	17	Stock 017	2,23	6	13,38	
20	18	Stock 018	1,17	10	11,7	
21	19	Stock 019	5,11	40	204,4	
22	20	Stock 020	1,05	32	33,6	
23	21	Stock 021	1,68	25	42	
24	22	Stock 022	4,05	31	125,55	
25	23	Stock 023	8,24	11	90,64	
26	24	Stock 024	0,40	22	8,8	
27	25	Stock 025	7,85	8	62,8	
28						

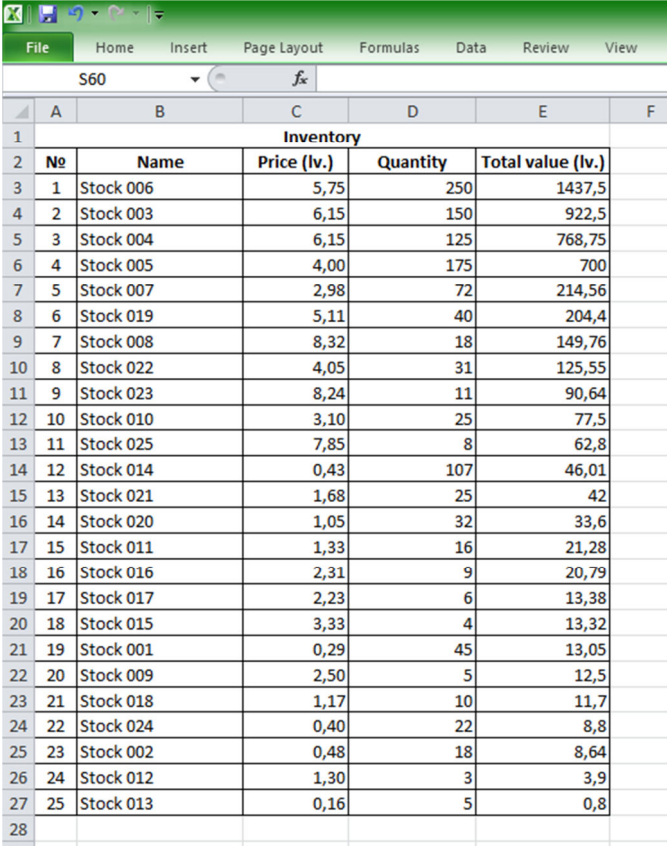
Figure 1. Inventory's nomenclature

As a classification criterion, we choose the total value of inventories of each type, i.e. the most important for the organization will be the inventories with the biggest investment, and the least important ones will be those that need the least financial resources.

The next step is to rearrange the inventory nomenclature in descending order by their total value. With MS Excel this is a simple procedure. The Sort function from the Data menu is used, and for the sort criterion (Sort by) Total Inventory is pointed. Sorting from the largest to smallest item must be set. One particular point here is that the entire

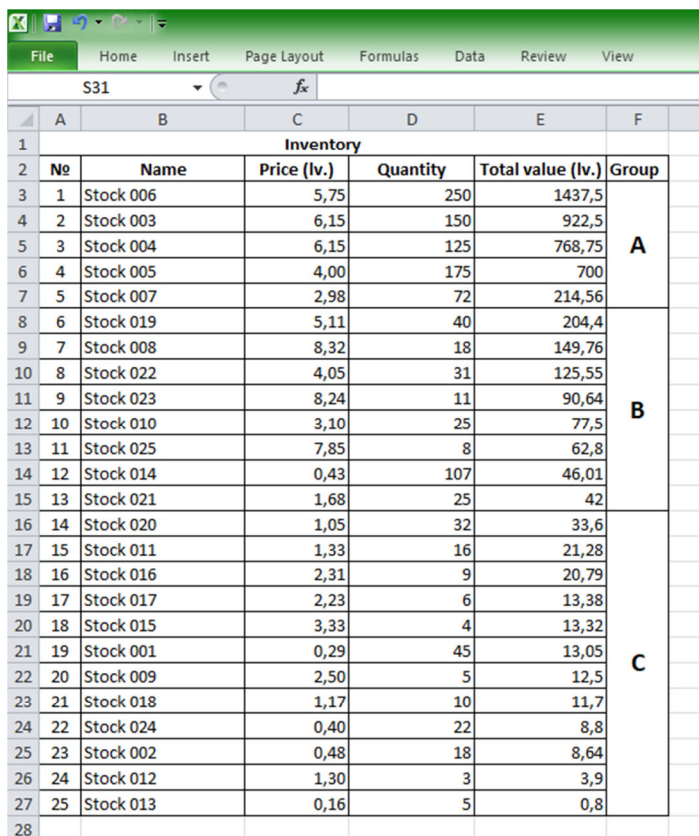
table (zone B2: E27) must be marked in order not to lose the link between the inventory name and its total value (fig. 1).

After ordering the inventories' nomenclature in descending order by total value, it becomes as following (Fig. 2).



	A	B	C	D	E	F
1						
2	No	Name	Price (lv.)	Quantity	Total value (lv.)	
3	1	Stock 006	5,75	250	1437,5	
4	2	Stock 003	6,15	150	922,5	
5	3	Stock 004	6,15	125	768,75	
6	4	Stock 005	4,00	175	700	
7	5	Stock 007	2,98	72	214,56	
8	6	Stock 019	5,11	40	204,4	
9	7	Stock 008	8,32	18	149,76	
10	8	Stock 022	4,05	31	125,55	
11	9	Stock 023	8,24	11	90,64	
12	10	Stock 010	3,10	25	77,5	
13	11	Stock 025	7,85	8	62,8	
14	12	Stock 014	0,43	107	46,01	
15	13	Stock 021	1,68	25	42	
16	14	Stock 020	1,05	32	33,6	
17	15	Stock 011	1,33	16	21,28	
18	16	Stock 016	2,31	9	20,79	
19	17	Stock 017	2,23	6	13,38	
20	18	Stock 015	3,33	4	13,32	
21	19	Stock 001	0,29	45	13,05	
22	20	Stock 009	2,50	5	12,5	
23	21	Stock 018	1,17	10	11,7	
24	22	Stock 024	0,40	22	8,8	
25	23	Stock 002	0,48	18	8,64	
26	24	Stock 012	1,30	3	3,9	
27	25	Stock 013	0,16	5	0,8	
28						

Figure 2. Inventories' nomenclature in descending order of total value



	A	B	C	D	E	F
1	Inventory					
2	No	Name	Price (lv.)	Quantity	Total value (lv.)	Group
3	1	Stock 006	5,75	250	1437,5	A
4	2	Stock 003	6,15	150	922,5	
5	3	Stock 004	6,15	125	768,75	
6	4	Stock 005	4,00	175	700	
7	5	Stock 007	2,98	72	214,56	
8	6	Stock 019	5,11	40	204,4	B
9	7	Stock 008	8,32	18	149,76	
10	8	Stock 022	4,05	31	125,55	
11	9	Stock 023	8,24	11	90,64	
12	10	Stock 010	3,10	25	77,5	
13	11	Stock 025	7,85	8	62,8	
14	12	Stock 014	0,43	107	46,01	
15	13	Stock 021	1,68	25	42	
16	14	Stock 020	1,05	32	33,6	C
17	15	Stock 011	1,33	16	21,28	
18	16	Stock 016	2,31	9	20,79	
19	17	Stock 017	2,23	6	13,38	
20	18	Stock 015	3,33	4	13,32	
21	19	Stock 001	0,29	45	13,05	
22	20	Stock 009	2,50	5	12,5	
23	21	Stock 018	1,17	10	11,7	
24	22	Stock 024	0,40	22	8,8	
25	23	Stock 002	0,48	18	8,64	
26	24	Stock 012	1,30	3	3,9	
27	25	Stock 013	0,16	5	0,8	
28						

Figure 3. Inventories' nomenclature classified by empirical method

The next step is to apply the empirical method of classifying inventories on the ordered nomenclature. According to it, the first 20% of the inventories, or $20\% \cdot 25 = 5$, i.e. the top 5 items in the rearranged nomenclature fall into Group A. In group B, the next 30% or $30\% \cdot 25 = 7,5$, rounded to an integer is 8, i.e. the next 8 elements of the nomenclature. The rest inventories from item 14 to item 25 belong to group C (fig. 3).

In the next step we will show how the graphical method of classifying the inventory's nomenclature can be applied, which is a bit more laborious and involves some additional calculations (fig. 4).

Firstly, the total value of all stocks is determined as the sum of the total values of the particular items in the nomenclature. The formula „=SUM(E3: E27)" entered in cell E28 was used. This value is used to calculate the relative proportion of each element in the nomenclature using formula (1). For this purpose, the column G in the table in MS Excel was used. In the cells from G3 to G27 by copying are entered the following formulas „=E3*100/\$E\$28“, „=E4*100/\$E\$28“, ..., „=E27*100/\$E\$28“.

The next step after this is to calculate the total relative shares of stocks in the nomenclature and use column H. The relative share of the first element of the nomenclature must be entered in box H3 and therefore the formula '= G3' is used. The sum of the relative shares of inventories from the first to the current position must be entered in each subsequent cell of this column. Therefore, the formulas „=SUM(\$G\$3:G4)“, „=SUM(\$G\$3:G5)“, ..., „=SUM(\$G\$3:G27)“ are entered in cells H4 to H27.

ABC analysis - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View								
Q30								
	A	B	C	D	E	F	G	H
1	Inventory							
2	Nº	Name	Price (lv.)	Quantity	Total value (lv.)		Relative share (%)	Total relative share (%)
3	1	Stock 006	5,75	250	1437,5		28,73	28,73
4	2	Stock 003	6,15	150	922,5		18,44	47,16
5	3	Stock 004	6,15	125	768,75		15,36	62,53
6	4	Stock 005	4,00	175	700		13,99	76,52
7	5	Stock 007	2,98	72	214,56		4,29	80,81
8	6	Stock 019	5,11	40	204,4		4,08	84,89
9	7	Stock 008	8,32	18	149,76		2,99	87,88
10	8	Stock 022	4,05	31	125,55		2,51	90,39
11	9	Stock 023	8,24	11	90,64		1,81	92,20
12	10	Stock 010	3,10	25	77,5		1,55	93,75
13	11	Stock 025	7,85	8	62,8		1,26	95,01
14	12	Stock 014	0,43	107	46,01		0,92	95,93
15	13	Stock 021	1,68	25	42		0,84	96,77
16	14	Stock 020	1,05	32	33,6		0,67	97,44
17	15	Stock 011	1,33	16	21,28		0,43	97,86
18	16	Stock 016	2,31	9	20,79		0,42	98,28
19	17	Stock 017	2,23	6	13,38		0,27	98,55
20	18	Stock 015	3,33	4	13,32		0,27	98,81
21	19	Stock 001	0,29	45	13,05		0,26	99,07
22	20	Stock 009	2,50	5	12,5		0,25	99,32
23	21	Stock 018	1,17	10	11,7		0,23	99,56
24	22	Stock 024	0,40	22	8,8		0,18	99,73
25	23	Stock 002	0,48	18	8,64		0,17	99,91
26	24	Stock 012	1,30	3	3,9		0,08	99,98
27	25	Stock 013	0,16	5	0,8		0,02	100,00
28	SUM				5003,73			
29								

Figure 4. Calculations in which is necessary to apply the graphical method for classifying the inventory's nomenclature

Column H values are used to construct the cumulative curve. In MS Excel, it is also easy to do this by entering a graph, such as Line or Scatter (fig. 5).

After the cumulative curve has been constructed, two tangent lines to the curve should be constructed in succession based on the algorithm described above, showing where the boundary between the different groups of inventories are.

Building a plan related to the graphical method can be accomplished without much difficulty using the Line drawing tool included in Shapes from the Insert menu.

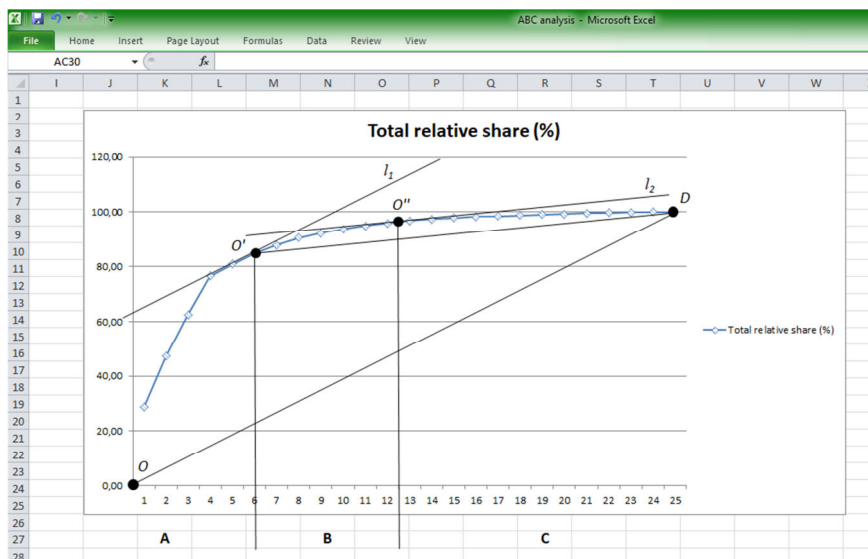


Figure 5. Graphical method for classifying inventory's nomenclature

The following result is obtained when working with the suggested sample data for inventory's nomenclature and drawing the plan according to the described graphical method: Group A must include the inventories which are in the first 6 positions of the rearranged nomenclature; Group B must include inventories in positions 7 to 12; Group C should include inventories from 13th to 25th position.

When implementing the ABC analysis method in practice it is a matter of manager's decision which method for determining of the groups A, B and C will be used. Usually the results are similar but not always identical and up to now there are no clear instructions which is the one to be preferred for application.

Some difficulties may occur when applying the graphical method if the volume of the inventory's nomenclature is relatively bigger. In this case an approximated function of the cumulative curve may be used which is obtained for example by the "least squares method". Then, without needing a graph, we may determine the coordinates of the points which mark the boundaries between the three groups inventories, using some calculus skills (straight line in plane and derivatives of function of

one variable). This can also be performed by using MS Excel which will reduce the volume of the calculation procedures and will give quite accurate results.

CONCLUSION

The method proposed for ABC analysis of the inventories in the logistical system by MS Excel is definitely reducing the number of the calculation procedures. It is an easy and applicatory tool and can be used by a wide range of consumers, having in mind how popular the software is. Its application in education is also profitable because it makes the method clear and comprehensible. It is immediately applicable in practice when working with real data and bigger nomenclatures.

The approach proposed in the paper is another confirmation of the necessity of applying the technologies in IT industry in all the areas of business and education in the contemporary society.

REFERENCES

1. DYBSKAYA, V. V. etc. (2008) Logistics. Integration and optimization of logistic business processes in supply chains. *Moscow: Eksmo*. [in Russian]
2. LUKINSKY, V. S. etc. (2007) Logistics in examples and tasks. *Moscow: Finance and statistics*. [in Russian]
3. LUKINSKY, V. S. (2003) Models and methods of the theory of logistics. *St. Petersburg: Peter*. [in Russian]
4. SERGEEV, V. I., etc. (2005) Corporate Logistics. 300 answers to questions from professionals. *Moscow: INFRA-M*. [in Russian]
5. STERLIGOVA, A. N. (2012) Inventory management in Supply Chain. *Moscow: INFRA-M*. [in Russian]
6. <https://mywestford.com/blog/fsn-analysis-in-inventory-management/>
7. <https://scialert.net/fulltext/?doi=ajava.2018.85.90>
8. <https://www.syntacticsinc.com/news-articles-cat/inventory-analysis-methods/>
9. <https://www.unleashedsoftware.com/blog/fsn-analysis-can-useful-tool-inventory-management>

CHOOSING STORAGE DEVICES WITH CONSIDERATION OF BACKUP AND RESTORE PERFORMANCE IN MS SQL SERVER

Ivan Kuyumdzhev¹, Radka Nacheva²

¹ *University of Economics – Varna/Informatics Department,
Varna, Bulgaria, ivan_ognyanov@ue-varna.bg*

² *University of Economics – Varna/Informatics Department,
Varna, Bulgaria, r.nacheva@ue-varna.bg*

Abstract

Importance of the process of restoring a database requires precise examination of all factors that affects it. Considering the decline of price of SSD and it's better read and write speeds it seems like logical choice when reducing downtime is primary concern for the organization. This paper categorizes storage devices in use as well as main backup types in MS SQL Server. Tests are conducted to evaluate the impact of the storage device on backup and restore efficiency. In this regard, **the purpose of the research** is to examine problems associated with selection of a storage device for efficient backup and restore in MS SQL Server. Conducted tests showed some surprising results regarding SSD's efficiency. The reason these findings are important is that they could be used as a guideline for choosing proper storage device for organization's needs.

Keywords: *backup and restore performance, storage media, databases performance tests, MS SQL server.*

INTRODUCTION

The volume of data worldwide is growing enormously, hence the need for storage capacity is increasing. Over five years, the generated data has more than tripled - from 12 ZB for 2014 to the expected 40 ZB by the end of 2019 (Statista, 2019c). Global data center IP traffic for 2018 is 952 EB and until the end of 2019 it is predicting it will be 997 EB only in traditional data centers but the traffic from the cloud data centers is over 10 times larger (Statista, 2019d). Especially considering the fact that more and more businesses are getting oriented towards online presence (Stanoeva, 2017), the data is increasing rapidly, it is important to choose a storage data device for better performance of backup and restore processes.

MS SQL Server is one of the most popular commercial relational database management system. It offers built-in backup and restore functionality, that can mitigate the risk of disastrous events. This functionality includes couple backup types that could be combined in many different strategies. Characteristics of each backup type is outlined by Microsoft regarding factors as database size and amount of changed data but there is no specific guidance on which storage device to be chosen.

In this regard, **the purpose of the research** is to examine problems associated with selection of a storage device for efficient backup and restore in MS SQL Server. On this basis the paper could be used as a guideline for choosing proper storage device for organization's needs.

1. COMPARISION OF DIGITAL DATA STORAGE MEDIA

As an integral element of the information technology sector, the global demand and supply of data storages by the end of 2019 will be around 31,000 EB and 20,000 EB compared to a period of five years ago, when they were approximately 6 times smaller. (Statista.com, 2019a). Worldwide spending on data storage has exceed 51 billion U.S. dollars in 2019 (Horst, 2018). According to other statistics global data will grow 61% to 175 zettabytes by 2025, with as much of the data residing in the cloud as in data centers (Patrizio, 2018). This shows that the need for using storage devices is increasing, including for private users.

Data storage refers to the devices which are used for retaining and archiving digital data that will be used by computer or another device. It could be divided to primary (main), secondary (auxiliary) or tertiary memory. The first type is random access memory (RAM) and its derivatives. The second type are the commonly used hard disks and solid-state drives. The third type are networked or cloud storages. Other bearers of digital information could be magnetic tapes, floppy disks, optical disks (CDs, DVDs, Blu-ray disks), flash drives and flash memory cards. Considering the purpose of this study and the capabilities of each of the digital data media, we are focusing on researching capabilities of secondary memory devices - hard disk drives (HDDs) and solid-state drives (SSDs).

Currently, many of the computers (home or enterprise machines) use both main types of secondary memory media. Computer configura-

tions equipped with both types of devices are often encountered, as home and corporate users prefer to take advantage of both technologies. Statistics show that around 360 million HDDs and 280 million SSDs are expected to sell by the end of 2019, compared to the previous three or four years, when SSDs sales were approximately 4 times less than HDDs (Statista.com, 2019b).

The answer to the question "Which device is better?" cannot be given unequivocally because it depends mainly on the purposes for which each of them will be used. The main difference between them is the technology they are designed for - hard disk drives are magnetic data carriers and solid-state drives are based on flash technology. Through Table 1, we compare the characteristics of the two main media types.

Table 1

Digital Data Storage Media¹

Characteristic	Hard Disk Drive	Solid-State Drive
Cost	buying a cheap 4TB model - around €0.02 / GB	buying a cheap 1TB model - around €0.09 / GB
Capacity ²	2 ÷ 10 TB	128 GB ÷ 4 TB
Power Consumption	7 Watt average	2 Watt average
Read Speed	approx. 150 MB/s	over 450 MB/s
Write Speed	approx. 100 MB/s	over 500 MB/s
Time to Access	approx. 2 sec. before it can read/write	read/write immediately
Annualized Failure Rate	1,7%	0,4%
Encryption	Full Disk Encryption	Full Disk Encryption
CPU Power	7%	0,7%

Still, the digital data storage capacity of SSDs is much smaller than HDDs'. However, the price ratio is inversely proportional - SSDs have a

¹ The data in Table 1 are averaged based on tests performed by (UserBenchmark, 2019a), (UserBenchmark, 2019b), (Xu, 2019), (Safford, 2019) and (Klein, 2019). The specifications vary from manufacturer to manufacturer.

² Depending if it is a laptop or desktop machine.

significantly higher price than HDDs. The flash based SSDs' technology, as seen in Table 1, enables faster reading and writing of data, faster access and a correspondingly lower annualized failure rate. CPU power consumption on SSDs is also less than on HDDs.

2. BACKUP AND RESTORE IN MS SQL SERVER

MS SQL Server is one of the most used commercial databases systems up to date. It offers a wide variety of options for continuity. To keep the business processes running, the organization can implement solutions such as a clustering, replication, mirroring the database, and log file transfer, so that even if the main server stops working, the requests are sent to failover server (Jorgensen, 2012). Main issue with these strategies is that they require extra financing – including software, hardware and professionals with high level of expertise in the area. This factor is more than enough for majority of the organizations to look for cheaper alternatives – one of which is using the built-in backup and restore functionality in MS SQL Server Express.

Main database backup types can be considered depending on whether clients have access to the data during the process. With offline (cold) backup, the database management system suspends access to the archived object and all applications that use it must wait for the process to complete. Online (hot) backup enables the database to remain operational and allow clients to perform the necessary operations without knowing that backup is currently underway. Considering the fact that organizations prefer to keep systems running, hot backup types are examined in the following tests.

Since MS SQL Server version 7 there are four primary database backup types (Desai, 2000):

- Full Backup – backups all data and objects part of the database.
- Differential Backup – it only backs up database changes that have occurred since the last full backup to date, and therefore runs in less time than a full backup would run at the same point in time.
- Log Backup – copies all operations recorded in the transaction log performed since the last log backup.
- Partial Backup - Backing up file groups or individual files is used when the database is so large that the full backup goes beyond ac-

ceptable organization timeframes. This option is added in MS SQL Server version 2005 and backups all file groups except marked as read-only.

Backup plan determines the intensity, sequence, and types of backup that apply to the database. It depends on various factors: what is the amount of information available, at what periods and how it enters the database, what part of it could be recovered from other sources, etc. Recovery plan depends on the backup strategy implemented. Time needed to execute a restore command is similar to time needed for corresponding backup type – fastest when using log backup and slowest with full backup.

Considering wide variety of backup types and strategies test should be conducted to measure their performance with different settings for used storage device.

3. COMPARISON OF THE PERFORMANCE OF DIFFERENT STORAGE MEDIA WITH MS SQL SERVER

To conduct the tests, we chose to use two of the fastest and most reliable storage devices examined in the first section – HDD and SSD. To properly compare their efficiency, we created a web-based application that executes commands to MS SQL Server instance and records the results of their execution. Database in use is a variant of Northwind sample database. To observe if the characteristics of each backup and restore type is preserved with different size of the database, the application executes insert queries with every iteration. The algorithm in use is increasing number of execution of steps 2 and 3 at every iteration. This means that at the last step there are 10 log backups, containing database incremented 10 times. After the end of this the experiment is repeated using an HDD instead of SSD storage device. The program logic takes the following steps:

- 1) create full database backup;
- 2) increasing database size by adding 3 000 records in tables production.TransactionHistory, person.person and sales.SalesOrderDetail ;
- 3) create log backup;
- 4) create differential backup;
- 5) create full backup;
- 6) restore full backup executed in step 1;

- 7) restore log backups executed in step 3;
- 8) restore full backup executed in step 1;
- 9) restore differential backup from step 4;
- 10) restore full backup executed in step 5;

Hardware testing environment is close to the one that a low budget organization would be using: CPU AMD FX 6100 3.30GHz, RAM 4GB DD3-1333, Samsung SSD 850 EVO and SATA III Hard Disk. Software in use is as follows: Windows 10 64bit, PHP 5.4.7, Apache Server 2.4.3.

Below are the results.

Table 2

HDD Log Backup and Restore results

Iteration	Log backup files	Log Backup			Log Restore	
		Time (s)	Backup Size	Speed (MB/s)	Time (s)	Speed (MB/s)
1	1	0,70	63,37	62,76	4,43	9,94
2	2	1,06	89,72	83,36	5,19	16,95
3	3	1,81	133,25	73,03	6,97	18,94
4	4	2,30	170,79	76,67	8,63	20,39
5	5	3,02	212,70	72,93	9,84	22,35
6	6	3,29	254,74	80,32	11,47	23,02
7	7	3,89	298,29	79,20	13,35	23,07
8	8	4,39	348,46	80,12	15,32	22,98
9	9	5,55	425,52	71,38	18,06	21,93
10	10	5,91	469,55	74,47	20,43	21,54

Table 3

SSD Log Backup and Restore results

Iteration	Log backup files	Log Backup			Log Restore	
		Time (s)	Backup Size	Speed (MB/s)	Time (s)	Speed (MB/s)
1	1	0,55	63,06	80,62	4,37	10,06
2	2	0,78	89,80	112,45	5,10	17,25

3	3	1,18	131,97	111,96	6,68	19,77
4	4	1,48	170,51	119,30	8,62	20,43
5	5	2,54	212,93	86,55	10,32	21,31
6	6	3,09	254,75	85,45	12,59	20,97
7	7	3,43	297,22	89,91	14,67	20,99
8	8	4,26	347,56	82,65	17,02	20,68
9	9	5,05	424,55	78,49	20,19	19,62
10	10	5,61	468,91	78,49	22,92	19,19

Table 4

HDD Differential Backup and Restore results

Iteration	Differential Backup			Differential Backup Restore	
	Time (s)	Backup Size	Speed (MB/s)	Time (s)	Speed (MB/s)
1	1,55	140,13	13,53	1,84	11,42
2	1,61	164,75	26,03	2,16	19,46
3	2,01	195,82	31,30	2,32	27,20
4	2,35	224,38	35,81	2,65	31,70
5	2,63	254,82	39,88	2,91	36,08
6	3,01	287,63	41,84	3,25	38,74
7	3,28	324,88	44,81	3,60	40,84
8	3,56	365,82	47,19	4,30	39,06
9	4,35	420,00	43,47	4,54	41,64
10	4,84	467,25	43,39	5,40	38,90

Table 5

SSD Differential Backup and Restore results

Iteration	Differential Backup			Differential Backup Restore	
	Time (s)	Backup Size	Speed (MB/s)	Time (s)	Speed (MB/s)
1	1,43	139,04	14,68	1,73	12,14

2	1,54	164,23	27,36	1,98	21,24
3	1,82	196,17	34,54	2,20	28,58
4	1,97	223,04	42,59	2,48	33,92
5	3,13	254,04	33,50	2,73	38,43
6	3,49	287,29	36,07	3,53	35,71
7	3,81	323,92	38,61	4,28	34,35
8	3,82	365,23	43,94	4,74	35,44
9	4,61	419,98	41,04	5,07	37,26
10	5,61	466,10	37,41	6,22	33,78

Table 6

HDD Full Backup and Restore results

Iteration	Full Database Backup from step 1			Full Database Backup Restore	
	Time (s)	Backup Size	Speed (MB/s)	Time (s)	Speed (MB/s)
1	1,85	190,23	102,63	1,92	121,73
2	2,19	234,13	106,78	2,49	109,77
3	2,54	273,63	107,61	2,72	121,47
4	3,20	330,94	103,42	3,34	120,27
5	3,77	401,57	106,61	3,94	124,17
6	4,51	489,32	108,57	4,51	131,48
7	5,45	593,25	108,90	5,64	127,06
8	6,57	717,13	109,19	6,92	124,08
9	7,71	858,25	111,29	7,80	132,23
10	9,46	1031,69	109,02	9,56	127,73

Table 7

SSD Full Backup and Restore results

Iteration	Full Database Backup from step 1			Full Database Backup Restore	
	Time (s)	Backup Size	Speed (MB/s)	Time (s)	Speed (MB/s)
1	1,32	190,21	144,28	1,59	146,81
2	1,84	233,23	126,60	2,02	133,45
3	2,08	273,17	131,13	2,43	135,91
4	2,59	330,98	127,68	2,93	136,96
5	2,96	400,98	135,58	3,51	139,12
6	3,63	488,67	134,45	4,03	147,03
7	5,02	593,10	118,03	5,55	128,98
8	5,61	716,54	127,68	7,40	115,95
9	6,92	857,85	123,88	9,06	113,86
10	8,89	1031,23	115,94	10,82	112,72

Calculating average speeds leads to interesting results. Table 8 shows that surprisingly speed of differential backup and restore is better when HDD storage type is used. Speed of both full backup and restore is slightly better with SSD but once again log backup and restore speeds are not undeniably faster with the more expensive configuration.

Table 8

Average results

Data storage	Operation Type	Log backup/ restore	Differential backup / restore	Full backup / restore
HDD	Backup	75,424 MB/s	36,725 MB/s	107,402 MB/s
HDD	Restore	20,111 MB/s	32,504 MB/s	123,999 MB/s
SSD	Backup	92,587 MB/s	34,974 MB/s	128,525 MB/s
SSD	Restore	19,027 MB/s	31,085 MB/s	131,079 MB/s

CONCLUSION

The publication researches the different types of storage media and their combination with backup and restore in MS SQL Server. The analysis of the theoretical capabilities and qualities of the carriers shows that the SSDs has more advantages than HDDs. For example, better read, write and access performance, lower failure rate and CPU power consumption. On the other hand, we found that the actual speed of different backup types is not as better as it should be expected with SSD technology.

That is why HDD storage type should be used when the database is as large as several gigabytes or backup strategy relies on usage of differential backups. SSD should be used with databases as large of several thousand gigabytes with backup strategy using many full and log backups.

REFERENCES

1. DESAI, A. (2000). *SQL Server 7 Backup & Recovery*. Mcgraw-Hill Osborne Media.
2. HOLST, A. (2018). Data Storage - Statistics & Facts. [Online] Available from: <https://www.statista.com/topics/3150/data-storage/> [Accessed 08/08/2019].
3. JORGENSEN, A., Wort, S., LoForte. R. and Knight, B. (2012). *Professional Microsoft SQL Server 2012 Administration*. Wrox.
4. KLEIN, A. (2019). Backblaze Hard Drive Stats Q2 2019. [Online] Available from: <https://www.backblaze.com/blog/hard-drive-stats-q2-2019/> [Accessed 08/08/2019].
5. PATRIZIO, A. (2018). IDC: Expect 175 zettabytes of data worldwide by 2025. [Online] Available from: <https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html> [Accessed 08/08/2019].
6. SAFFORD, M. (2019). How to Buy the Right SSD: A Guide for 2019. [Online] Available from: <https://www.tomshardware.com/reviews/ssd-buying-guide,5602.html> [Accessed 08/08/2019].
7. STANOEVA, I. (2017) Creating "Digital Marketing" Curriculum For "Art Management" Programmes. *Proceedings of The Second Conference on Innovative Teaching Methods (ITM 2017)*, 1(1), pp. 5-13.

8. STATISTA.COM. (2019a). Data storage supply and demand world-wide, from 2009 to 2020 (in exabytes). [Online] Available from: <https://www.statista.com/statistics/751749/worldwide-data-storage-capacity-and-demand/> [Accessed 08/08/2019].
9. STATISTA.COM. (2019b). Shipments of hard and solid state disk (HDD/SSD) drives worldwide from 2015 to 2021 (in millions). [Online] Available from: <https://www.statista.com/statistics/285474/hdds-and-ssds-in-pcs-global-shipments-2012-2017/> [Accessed 08/08/2019].
10. STATISTA.COM. (2019c). Volume of data/information created worldwide from 2010 to 2025 (in zetabytes). [Online] Available from: <https://www.statista.com/statistics/871513/worldwide-data-created/> [Accessed 08/08/2019].
11. STATISTA.COM. (2019d). Global data center IP traffic from 2012 to 2021, by type (in exabytes per year). [Online] Available from: <https://www.statista.com/statistics/227259/forecast-of-global-data-center-ip-traffic-by-type/> [Accessed 08/08/2019].
12. USERBENCHMARK. (2019a). Speed test your SSD in less than a minute. [Online] Available from: <https://ssd.userbenchmark.com/> [Accessed 08/08/2019].
13. USERBENCHMARK. (2019b). Speed test your HDD in less than a minute. [Online] Available from: <https://hdd.userbenchmark.com/> [Accessed 08/08/2019].
14. XU, E. et. al. (2019). Lessons and Actions: What We Learned from 10K SSD-Related Storage System Failures. *Proceedings of the 2019 USENIX Annual Technical Conference*. pp. 961-975.

PREDICTING STUDENTS PERFORMANCE IN MOODLE PLATFORMS USING MACHINE LEARNING ALGORITHMS

Yanka Aleksandrova¹

*¹ University of Economics Varna,
Varna, Bulgaria, yalexandrova@ue-varna.bg*

Abstract

E-learning platforms have become a widely used and advanced media to enhance the educational process. They bring benefits to all participants in this process – teachers, students and administration – in several different areas like teaching, learning, communication and sharing. The paper focuses on the application of machine learning algorithms for predicting students' performance based on their interaction with the e-learning platforms. The research hypothesis is that the success or failure on e-learn courses could be predicted using data from activity logs. To support the hypothesis several machine learning algorithms have been performed, such as logistic regression, random forest, gradient boosting decision trees (xgboost) and neural network. The results indicate that all algorithms perform the classification task satisfactory with accuracy above 0.84. The comparison of the evaluation metrics reveals a better performance for neural network and gradient descent boosting trees compared to logistic regression and random forest. The experiments have been performed using R programming language.

Keywords: *machine learning, e-learn, Moodle, students' performance, gradient boosting, random forest, R language.*

INTRODUCTION

The application of machine learning algorithms in educational electronic platforms has been a research topic for many scientists. Nespereira, Elhariri et al. (Nespereira, et al., 2016) compared the performance of random forest and support vector machines to predict students' tendency to pass/fail as a relationship with their past course interactions with Learning Management Systems (LMS). A software framework has been proposed by Olive, Huynh et al. (Olive, et al., 2018) with implemented predictive model to identify students at risk of abandoning a course. In another research these authors (Olive, et al., 2019) presented a case study model which predicts students at risk of dropping out a Moodle course. The average accuracy achieved with a neural network was 88.81%.

Conijn, Snijders, Kleingeld and Uwe (Conijn, et al., 2017) analysed data from 17 blended Moodle courses to predict student performance from LMS variables using multi-level and standard regressions. Their results showed that predictive modelling strongly vary across courses and suggest low portability of the models across courses. They conclude that a more specific theoretical argumentation is needed to complement the LMS data.

Gamie et al. (Gamie, et al., 2019) developed a model with new features grouped in dimensions according to their relativeness to the teaching style and students' activities on an e-learn system. The prediction analysis has been performed with the best fitting classifier reaching accuracy of 87%. Other authors (Hussain, et al., 2018) implemented machine learning classification and clustering techniques to detect the low-performance students prior the examination. Their experimental results identified that the higher accuracy in identifying the inactive students could be achieved by fuzzy unordered rule induction algorithm (FURIA). K-means clustering could be also implemented to separate active from inactive and poorly performed users.

Nguyen et al. (Nguyen, et al., 2018) proposed a forecast model based on students' interaction with e-learn systems. They tested their model in the Moodle LMS system and achieved an accuracy of over 50% with 75% of students showing outcomes close to the predicted results.

1. METHODOLOGY

The research hypothesis is stated as follows: the success or failure of e-learn courses could be predicted using data from activity logs. For the purpose of the research the students' performance is defined as a success or failure on the final exam after the completion of the course. In order to support the hypothesis several research stages have been defined as depicted on figure 1. The Moodle based platform (e-learn.ue-varna.bg) has been used for blended learning with all the learning resources provided only through the platform. There are two main limitations of the current research. The first one is regarding the stated hypothesis which only explores the success or failure of the exam. The second limitation is connected to the input dataset. It contains relatively small

number of observations (112) from only two bachelor programs – “Informatics” and “Business Information Systems”.

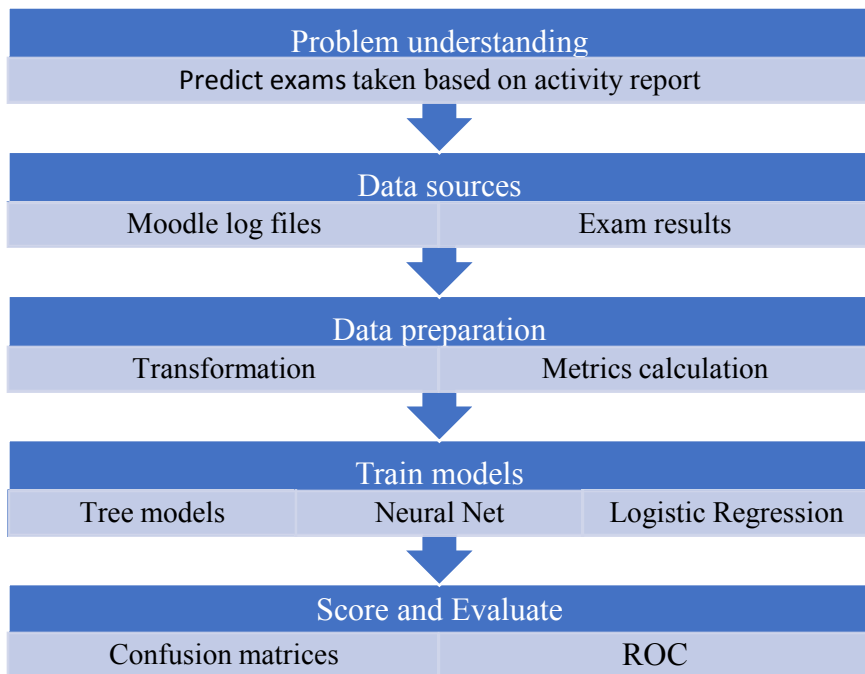


Figure 1. Stages of the research process

The problem is defined according to the research hypothesis. Data sources include activity log from two completed courses in a Moodle based platform and exam results. The structure of the activity log file is as follows: activity log (time, user full name, affected user, event context, component, event name, description, origin, IP address). Data from the log file has been transformed and summarized at a student level to form the structure of the dataset shown on table 1. In order to eliminate differences between courses due to course structure and number of resources provided, some metrics have been calculated as a ratio to the average values. These include tasks, files viewed and files uploaded.

The definition of the students’ performance as a success or failure leads respectively to the chosen algorithms for binary classification –

logistic regression, tree models and neural network. The models have been fitted using cross-validation due to the relatively small number of observations. The trained models have been scored and evaluated using performance metrics as accuracy, sensitivity, specificity, etc.

Table 1

Dataset structure

Variable	Type	Description
Students ID	integer	sequential number
Gender	category	2 levels - F and M
Program	category	currently with two states
Enrolment days after start of semester	integer	[Enrolment date] - [Semester start date]
Days from first to last session	integer	[Date of last session] - [Date of first session]
Unique days count	integer	Count distinct of date
Actions taken	integer	Actions count
Average actions per day	num	$[\text{Actions count}] / \text{Avg}([\text{Actions count for the course}])$
Actions taken from university network (ratio)	num	$[\text{Actions count taken from university}] / [\text{Actions count}]$
Actions taken from outside university network (ratio)	num	$[\text{Actions count taken from outside the university}] / [\text{Actions count}]$
Tasks count compared to average	num	$[\text{Task actions}] / \text{Avg}([\text{Task actions}])$
Files viewed compared to average	num	$[\text{Files viewed actions}] / \text{Avg}([\text{Files viewed actions}])$
Files uploaded compared to average	num	$[\text{Files uploaded actions}] / \text{Avg}([\text{Files uploaded actions}])$
Exam taken	category, target	2 levels - 0 (not taken) and 1 (exam taken)

2. PRELIMINARY ANALYSIS

The dataset contains 112 examples of students completed a course of “Design of Information Systems” from two bachelor programs – “Informatics” (69) and “Business Information Systems” (43). The target variable (exam_taken) is relatively equally distributed with 52.68% from

the students successfully passed the exam and 47.32% failed. As a preliminary analysis several boxplots charts of some numerical variables by target have been drawn as depicted on figure 2.

The boxplot charts suggest that some of the variables could be better predictors as there are clear differences in their distribution among the different groups. Such variables are number of actions taken, unique days, tasks from average and files viewed by average. At the same time there is no clear difference in the distribution of actions per day and days from start of the semester to the enrolment date.

The target distribution between the two categorical variables – gender and program is shown using stacked bar chart (figure 3 and figure 4). The target variable is unequally distributed in the two gender groups with 66% from female students successfully passed the exams related to 41.94% from the male students. The students from the two programs performed almost similarly with a slightly greater percentage of students from Informatics taken the exam (56.52%) compared to those from “Business Information Systems” (46.51%).

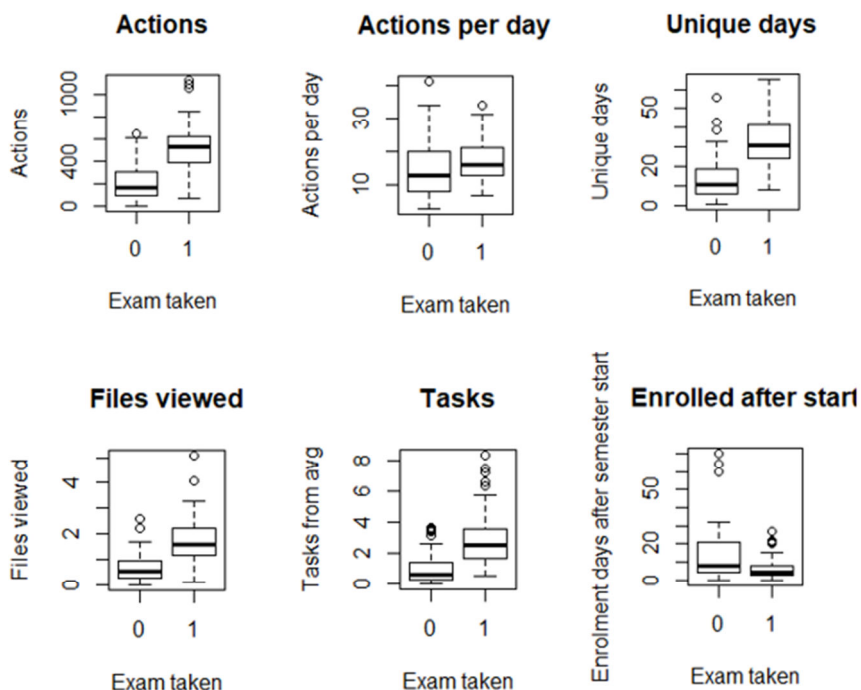


Figure 2. Boxplots of some numeric variables by target

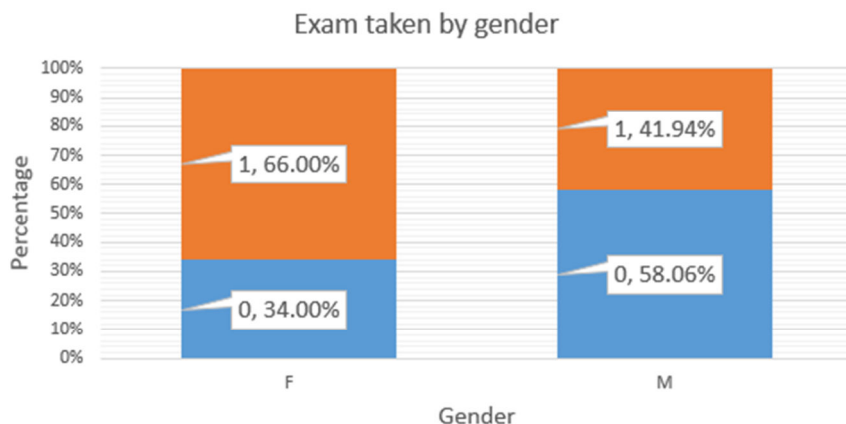


Figure 3. Exam taken by gender (percentage)

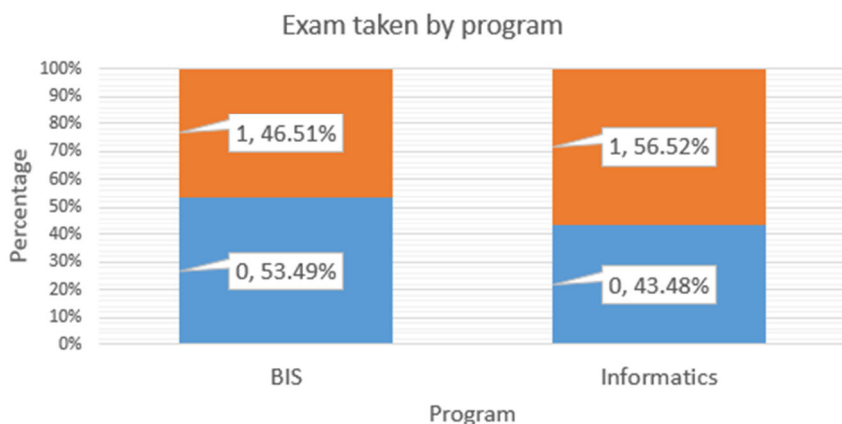


Figure 4. Exam taken by program

3. RESULTS AND FINDINGS

First, a logistic regression has been performed to set the benchmark level for model performance. The variable importance is shown on table 2 with significant variables ($p < 0.05$) flagged by “*”. The most important variables identified by the model are “files uploaded from avg”, “actions uni” and “files viewed from avg”. The logistic regression achieved accuracy of 0.8750 with specificity of 1.0000 and sensitivity of 0.7647.

Table 2

Logistic regression – variable importance

Variable	Overall	Significant
files_uploaded_from_avg	2.8990461	*
actions_uni	1.9134639	*
files_viewed_from_avg	1.7894404	*
programInformatics	1.4277667	
tasks_from_avg	1.3805153	
days_unique	1.1362720	
days_frst_last	0.6005918	
actions	0.4044596	
avg_actions_per_day	0.1968169	
genderM	0.1596336	
enr_days_after_sem_start	0.1322345	

The second applied machine learning algorithm was random forest. The model has been fitted with a 10-fold cross-validation. The parameters have been tuned to number of trees = 2500 and mtry=4. The variable importance derived from random forest algorithm is shown on figure 5. According to MeanDecreaseAccuracy and MeanDecreaseGini the most significant variables are “tasks from avg”, “actions” and “files uploaded from avg”. The model achieved an accuracy of 0.8438, sensitivity 0.8824 and specificity 0.8000.

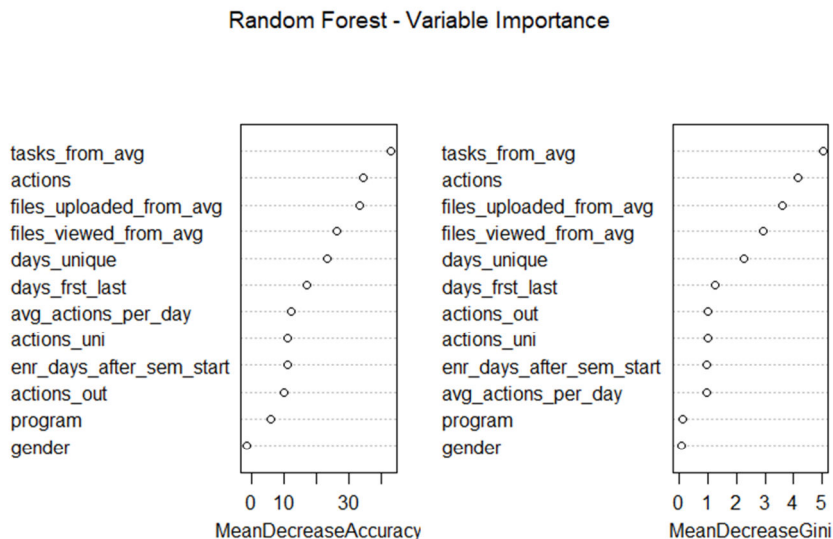


Figure 5. Variable importance derived from Random Forest model

The gradient descent boosting trees based on xgboost package and neural net (12-7-1) performed better than the first two algorithms achieving both accuracy of 0.9062. The xgboost model was hypertuned to nrounds=3, max_depth=20, eta=0.1, gamma=0.2, colsample_bytree=0.6.

Table 3

Evaluation results of applied machine learning algorithms¹

Metric	Random Forest	Logistic Regression	Neural Net	XG Boost
Accuracy	0.8438	0.8750	0.9062	0.9062
95% CI	(0.6721, 0.9472)	(0.7101, 0.9649)	(0.7498, 0.9802)	(0.7498, 0.9802)
No Information Rate	0.5312	0.5312	0.5312	0.5312
p-value [ACC > NIR]	0.000216	4.15E-05	6.19E-06	6.19E-06
Kappa	0.6850	0.7529	0.8110	0.8125
Mcnemar's Test P-Value	1.0000	0.1336	1.0000	1.0000
Sensitivity	0.8824	0.7647	0.9412	0.8840
Specificity	0.8000	1.0000	0.8667	0.9333
Pos Pred Value	0.8333	1.0000	0.8889	0.9375
Neg Pred Value	0.8571	0.7895	0.9286	0.8750
Prevalence	0.5312	0.5312	0.5312	0.5312
Detection Rate	0.4688	0.4062	0.5000	0.4688
Detection Prevalence	0.5625	0.4062	0.5625	0.5000
Balanced Accuracy	0.8412	0.8824	0.9039	0.9078

CONCLUSION

The research results confirm the hypothesis that the success or failure on exams could be predicted using machine learning algorithms on data extracted from students' interaction with e-learning platforms. All from the applied algorithms show satisfactory accuracy with xgboost and neural net with accuracy of 0.9062. Some of the significant variables identified by the models include number of actions taken, tasks from average and files uploaded from average.

¹ Positive class = 1

Related future works would focus on implementing machine learning algorithms on bigger datasets derived from activity log files. Another dimension of future improvement is to apply regression machine learning algorithms to predict the grade mark or exam points. Machine learning algorithms could be also used to predict students with greater probability of failure long before the exam date which could optimize the educational process.

REFERENCES

1. CONIJN, R., SNIJDERS, C., & KLEINGED, A. (2017). *Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS*. IEEE Transactions on Learning Technologies.
2. GAMIE, E., EL-SEOUD, M., SALAMA, M., & HUSSEIN, W. (2019). *Multi-dimensional analysis to predict students' grades in higher education*. International Journal of Emerging Technologies in Learning, 14(2), 4-15. doi:10.3991/ijet.v14i02.9905
3. HUSSAIN, S., HUSSAIN M., ZHANG, W., ZHU, W., THEODOROU, P., & ABIDI, S. (2018). *Mining moodle data to detect the inactive and low-performance students during the moodle course*. 2nd International Conference on Big Data Research, ICBDR 2018, (pp. 133-140). doi:10.1145/3291801.3291828
4. NESPEREIRA, C., ELHARIR, E., EL-BENDARY, N., & VILAS, A. (2016). *Machine learning based classification approach for predicting students performance in blended learning*. Advances in Intelligent Systems and Computing, 407, 47-56. doi:10.1007/978-3-319-26690-9_5
5. NGUYEN, V., NGUYEN, Q., & NGUYEN, V. (2018). *A model to forecast learning outcomes for students in blended learning courses based on learning analytics*. 2nd International Conference on E-Society, E-Education and E-Technology, ICSET 2018, (pp. 35-41). doi:10.1145/3268808.3268827
6. OLIVE, D., HUYNH, D., REYNOLDS, M., DOUGIAMAS, M., & WIESE, D. (2018). *A supervised learning framework for learning management systems*. First International Conference on Data Science, E-learning and Information Systems. doi:10.1145/3279996.3280014

7. OLIVE, D., HUYNH, D., REYNOLDS, M., DOUGIAMAS, M., & WIESE, D. (2019). *A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC*. Journal of Computing in Higher Education. doi:10.1007/s12528-019-09230-1

M-LEARNING APPLICATIONS

Latinka Todoranova¹, Bonimir Penchev²

*¹ University of Economics – Varna/Informatics, Varna, Bulgaria,
todoranova@ue-varna.bg*

*² University of Economics – Varna/Informatics, Varna, Bulgaria,
b.penchev@ue-varna.bg*

Abstract

The expansion of mobile technologies and their wider usage are transforming the process of e-learning into m-learning. The purpose of the article is to analyze the popular m-learning applications, to highlight their basic features and to outline the basic requirements for their development. Based on the current state of mobile phone market share the object of the research is Android learning apps. The applied approach includes analysis of the ten most popular m-learning applications. The results of the study show that all of the reviewed applications have a version for Android and for iOS operating systems. However, most of them are not typical m-learning applications, but rather they are e-learning libraries. In this regard, there is a need to find a way that more quickly introduce new learning approaches including the active use of mobile devices and applications in the learning process. First step must be to define a comprehensive framework for development of m-learning applications.

Keywords: *mobile learning, mobile applications, mobile technologies.*

INTRODUCTION

The use of mobile technologies in education is not a new concept. The active development in this area is for almost 20 years. One of the first publications in this scientific field, which is cited by a number of authors – “m-learning: Mobile, Wireless, In-Your-Pocket Learning”, has been published in 2000 (Quinn, 2000). This year is also associated with the generation Z¹. The representatives of this generation are surrounded by technology throughout their whole life and it is neither logical, nor useful to exclude this technology from the learning process. Therefore, these days the number of e-learning resources provided through content management platforms is very high. The access to these platforms is

¹ Generation Z – those born from the mid-1990s to the early 2000s.

established not only from desktop computers and laptops, but also from mobile devices. The learners are provided with educational materials in different formats – text, images, audio, video, etc. But this approach is not interactive enough and leads to limited interaction between learners and educators. Forums, chats, polls and tests, which are usually part of mobile learning platforms, are commonly used to improve communication and to provide better feedback. With the development of technology, mobile devices have improved their technical parameters. This in turn is a prerequisite for the creation of m-learning applications with rich functionality, attractive design and easy-to-use interface.

1. M-LEARNING APPLICATIONS DEVELOPMENT

Back in 2005, Traxler gave the following definition of m-learning: “any educational provision where the sole or dominant technologies are handheld or palmtop devices” (Traxler, 2005). With the development of technology, this definition still remains relevant: “m-learning is the use of mobile technology to make learning materials accessible and effective at a time and place convenient to the learner” (E-learning center, 2019). From the definition given nearly 15 years later, it becomes clear that the learner is already at the center of mobile learning, that he is getting used to working with mobile devices and to using more intuitive mobile applications.

In this paper, mobile applications are considered as application software intended for use on mobile devices such as smartphones and tablets. This type of software is tailored to the specific features of mobile devices and its market share is growing significantly. On the other hand, the development of mobile learning is leading to the development of m-learning applications. They are more often used in the learning process and are inherently the introduction of new learning methods. M-learning application advantages are generally:

- easier access to learning materials;
- synthesized learning materials²;
- provision of learning materials in a more accessible way – with many examples, illustrations, effects;

² The volume of learning units in m-learning is significantly smaller than that of e-learning materials.

- no need of print resources;
- significantly reduced time and effort required to search the needed information;
- facilitated communication between learners and educators;
- significantly better awareness of the learners.

The usage of such applications makes the learning process more accessible, more interesting and more engaging for the learners.

The development of m-learning applications is not an easy task. Lots of researchers, such as (Pastore, 2014), look in detail and analyze the advantages and disadvantages of the basic development approaches – native, web or hybrid applications. Generally, native applications are developed for a specific mobile platform (operating system). This allows the application to use the full functionality of the device on which it works but limits it to work only on the platform for which it is created. In contrast, web applications are used through a mobile web browser and have no restrictions on the type of the platform on which they will be used. However, they do not fully exploit the capabilities of the mobile devices. To overcome these drawbacks, hybrid applications could be used. They transform web applications into native applications. All three approaches have both advantages and disadvantages. In order to select the most suitable one for the development of a particular application, it is necessary to conduct a survey among the potential users. It should start with the type of mobile devices and their operating system.

The statistical data (Newzoo, 2018) shows that there is an increasing use of mobile devices, and in particular of mobile phones – in 2016 worldwide smartphone users are 2.5 billion. It is estimated that by 2021 this number will reach 3.8 billion. On the other hand, the results of different mobile phone market researches show a steady dominance of Android OS over iOS devices. The tendencies are these two operating systems to be the most used (fig. 1).

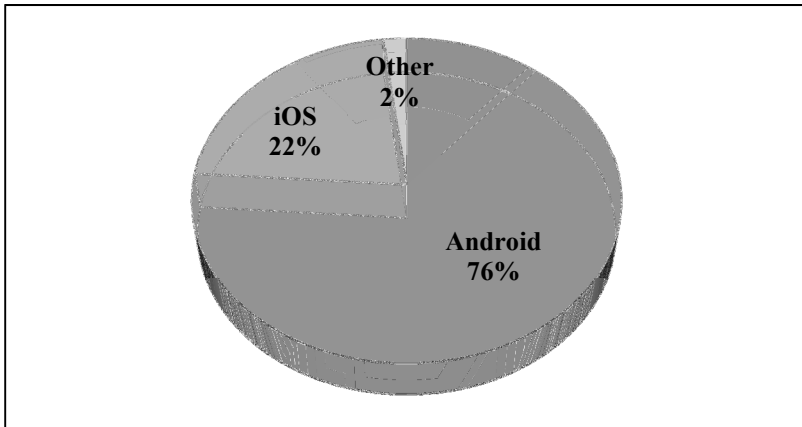


Figure 1. Mobile Operating System Market Share Worldwide (July, 2019)

Source: Statcounter, *Mobile Operating System Market Share Worldwide (July 2018 – July 2019)*, <https://gs.statcounter.com/os-market-share/mobile/worldwide/> [Accessed 26/08/2019]

These data indicate that m-learning applications should be created in such way that they can function equally well on these both operating systems. Examples of development tools that can be used for this purpose are PhoneGap, Appcelerator Titanium and RhoMobile suite.

Currently, mobile application users could choose to download between 2.46 million applications from Google Play (Statista, 2019), and almost 1.96 million available applications in Apple's App Store (Statista, 2019). With such a large number of applications, it is important to meet user expectations in order to convince a certain user to use one application instead another. If the application is not meeting user expectations, he can easily remove it from his device and download another one. These expectations are related both to the functionality of the applications and to the user experience. Therefore, creating useful and user-friendly applications is a difficult task. In their study Sulova and Nacheva (Sulova, Nacheva, 2018) conclude that the successful implementation of the UX strategy depends directly on the study of the target audience of the digital service or the software product as a whole. This is

due to the fact that the potential users are very different – they differ in gender, age, education, interests, nationality, etc.

2. BASIC FEATURES OF M-LEARNING APPLICATIONS

In this research paper we analyze existing m-learning applications in the market and explore their basic features. Undoubtedly, all of them have features such as flexibility, collaboration, motivation, accessibility, portability (Color, Atagan, 2009), efficiency, learnability, satisfaction, effectiveness and accessibility (Alotaibi, 2017) because they are crucial for the quality of this kind of software applications.

In order to facilitate any analysis, it is important to classify in advance the studied objects. However, this is a difficult task in terms of m-learning applications. One possible classification could be based on the right of use – free or paid, but it does not help to better analyze this kind of applications. Another very general classification is proposed by Mary Brown (Brown, 2014) and it has the following groups:

- Science Educational Apps – applications related to education in physics, biology or mathematics;
- Apps for Language Learning – support foreign language learning;
- Apps for Dyslexic students – for learners with Dyslexia and other learning disabilities;
- Apps for classroom education – support classroom learning;
- Apps for educators – help the development of educators;
- Learning apps for kids and elementary school – target the youngest users of mobile devices.

However, the available internet rankings do not include a breakdown by the aforementioned categories. There, applications are most often divided by subject area or mobile platform for which they are developed.

Most of the applications listed in a certain ranking, as well as those with similar functionality, are also found in many other similar rankings.

Table 1 presents the leading applications (Android Authority, 2019), the operating system for which they are developed and the area of specialization (subject area for which they are intended).

Table 1

Features of leading m-learning applications

Application	Mobile OS	Specialization
Amazon Kindle	iOS, Android	No
Coursera	iOS, Android	No
Khan Academy	iOS, Android	No
LinkedIn Learning	iOS, Android	No
Udemy	iOS, Android	No
Wolfram Alpha	iOS, Android, Windows ³	No
YouTube	iOS, Android	No
Duolingo	iOS, Android, Windows	Language learning
PhotoMath	iOS, Android	Math Problems
SoloLearn	iOS, Android	Coding learning

Back in 2013, a study conducted by Setiabudi and Tjahyana (Setiabudi, Tjahyana, 2013) among students from the Petra Christian University from the Informatics Department showed that 100% of the respondents use their mobile phones to download study materials. This indicates that mobile phones are used in education mostly for this purpose. It is also evidenced by the analysis of the main features of the presented applications – 70% is the share of m-learning applications that provide educational materials in the form of books, courses, videos. These are *Amazon Kindle*, *Coursera*, *Khan Academy*, *LinkedIn Learning*, *Udemy*, *Wolfram Alpha*, *YouTube*. The learning materials available in them are not limited to a specific subject area.

Coursera is a platform that provides a large number of courses accessible through mobile devices. The courses are organized into 11 main categories. Another popular platform similar to *Coursera* is *edX*. The number of courses in it is over 2500, organized in 30 thematic areas. It is also available for iOS and Android. In both platforms, the provided courses are developed by leading educational institutions.

³ Some of the courses do not support all operating systems.

The mobile application *Udemy* provides over 100,000 educational courses in 13 general categories. Since the number of courses is very large, after selecting a course category, the user can test his knowledge and on the basis of the result, he is offered with the most suitable courses in the selected category.

The application *Khan Academy* provides access to courses created by specialists in their respective subject areas. They are more specialized than those in *Coursera* and *edX*. *Khan Academy* is similar to the application *LinkedIn Learning*, which gives access to over 13,000 courses.

Wolfram Alpha contains a number of mobile applications providing deep computational knowledge in specific educational, professional, and personal areas.

Amazon Kindle is an application that provides access to over six million Kindle books. It has a very similar functionality to this in *Google Books*.

The application *SoloLearn* is more specific than the ones discussed above. It includes a set of 13 programming training courses. The lessons in every course are organized into modules. Each module ends with a test of the acquired knowledge.

However, with the development of technology, the share of specialized m-learning applications such as *Duolingo* and *PhotoMath* is expected to grow. They have features such as personalized experience, higher engagement level and notifications.

Applications related to language education (such as *Duolingo*) are being actively developed. They have specific functions that are required when learning a foreign language: determining the level of the learner; providing lessons (tailored to this level), exercises, task assessment and feedback.

PhotoMath is an application that not only provides math tutorials through many examples but also gives learner the opportunity to solve problems that are printed or handwritten. It is an example of an application that uses the basic functionality of mobile devices – the camera.

The widespread usage of mobile devices running Android and iOS has inevitably influenced the creation of m-learning applications. All of the applications reviewed have a version for both operating systems.

However, most of them are not typical m-learning applications, but rather they are e-learning libraries.

The real m-learning applications should have the following capabilities:

- online and offline access to learning units;
- optimized learning materials for small screens;
- bite-size tutorials;
- automatically synchronizes user activity across devices;
- enhanced interaction;
- immediate feedback;
- track course histories and progress;
- students must be engaged before, during and after class;
- entertainment.

These features are rather technological and are crucial to the imposition of one application over another. Therefore, they are more often taken into account.

The educational value of the learning content included in the m-learning applications should also not be forgotten. In the development of e-learning units it is not enough the participation of IT professionals. The importance of the educators, the people who prepare the curriculum and the regulation of intellectual property rights is enormous. For example, reducing the volume of the offered material compared to e-learning units should not be at the expense of missing out on important topics, but rather by synthesizing the main points. This cannot be accomplished without the help of the educators leading the certain course.

Research shows also that developers of modern m-learning applications are increasingly relying on the inclusion of artificial intelligence and augmented reality in their applications. In general, their purpose is to generate innovation in schools, offer new and different possibilities in the teaching and learning process, and help students to better understand content and information (Sonego, Machado, Torrezan, Behar, 2016).

CONCLUSION

The development of technology and the active usage of smartphones in everyday life require working towards a major change in the traditional education process. The millennium generation that is now starting university is "addicted" to mobile devices and they cannot be

excluded from their daily routine, in which is included the learning process. Therefore, there is a need to find a way that more quickly introduce new learning approaches including the active use of mobile devices. To this end, educators' training should be considered in order to take full advantage of the capabilities of modern mobile devices (in particular smartphones). In addition, educators should be actively involved in the process of creating m-learning applications in specific courses.

REFERENCES

1. ALOTAIBI, A. (2017) Increasing Levels of Interaction and Usability of Mobile Learning based on Augmented Reality. *International Journal of Computer Science and Mobile Computing*, Vol.6 Issue.3. Available from: <https://www.ijcsmc.com/docs/papers/March2017/V6I3201742.pdf> [Accessed 15/08/2019].
2. ANDROID AUTHORITY, 10 best Android learning apps! (Updated 2019), <https://www.androidauthority.com/best-android-learning-apps-566227/> [Accessed 26/08/2019]
3. BROWN, M. (2014) Mobile Apps in Education. Available from: <http://www.schoolmoney.org/mobile-apps-in-education/> [Accessed 08/08/2019].
4. E-LEARNING CENTER, Mobile Learning in 2019 – The Complete Guide to mLearning, <https://www.e-learningcentre.co.uk/mlearning> [Accessed 10/08/2019]
5. NEWZOO, Global mobile market report 2018, <https://newzoo.com/insights/trend-reports/newzoo-global-mobile-market-report-2018-light-version/> [Accessed 26/08/2019]
6. PASTORE, S. (2014) Developing mobile educational apps: development strategies, tools and business models. *Advances in Computer Science: an International Journal*, Vol. 3, Issue 1, No.7, January 2014.
7. QUINN, C. (2000) m-learning: Mobile, Wireless, In-Your-Pocket Learning. Available from: <http://www.linezine.com/2.1/features/cqmmwiyp.htm> [Accessed 26/08/2019].
8. SETIABUDI, D., TJAHYANA, L. (2013) Mobile Learning Application Based On Hybrid Mobile Application Technology Running On Android Smartphone and Blackberry. Available from: <https://core.ac.uk/download/pdf/11852245.pdf> [Accessed 14/08/2019]

9. SONEGO, A., MACHADO, L., TORREZZAN, C., BEHAR, P. (2016) Mobile learning: pedagogical strategies for using applications in the classroom. Available from: <https://files.eric.ed.gov/fulltext/ED571451.pdf> [Accessed 10/08/2019].
10. STATCOUNTER, Mobile Operating System Market Share Worldwide (July 2018 – July 2019), <https://gs.statcounter.com/os-market-share/mobile/worldwide/> [Accessed 26/08/2019]
11. STATISTA, Number of apps available in leading app stores as of 2nd quarter 2019, <https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/> [Accessed 26/08/2019]
12. STATISTA, Number of available apps at Google Play from 2nd quarter 2015 to 2nd quarter 2019, <https://www.statista.com/statistics/289418/number-of-available-apps-in-the-google-play-store-quarter/> [Accessed 26/08/2019]
13. SULOVA, S., NACHEVA, R. (2018) Approach to exploring users' expectations of digital services' functionality. Available from: https://www.researchgate.net/publication/325570164_Approach_to_exploring_users%27_expectations_of_digital_services%27_functionality [Accessed 2/08/2019].
14. TRAXLER, J. (2005) Defining mobile learning. Available from: https://www.researchgate.net/publication/228637407_Defining_mobile_learning [Accessed 10/08/2019].

THE IMPACT OF SOCIAL MEDIA ON VIDEO GAME COMMUNITIES AND THE GAMING INDUSTRY

Boris Bankov¹

*¹ University of Economics – Varna/Department
of Computer Science, Varna, Bulgaria,
boris.bankov@ue-varna.bg*

Abstract

The video game industry has grown immensely in the last three decades and one major driving force has been social interaction among gamers offline and online. The purpose of this paper is to present an overview of the evolution of the gaming industry and the different implementations of social channels in video games, the impact of integrating social media subsystems on gaming communities, their responses and feedback. Observations are made on the different trends in gaming in years between 1990 and 2019 and how communication and social interaction affect players. Income models throughout the period are explored. The study includes a research into a few significant examples of the symbiotic relationship between social media platforms and the gaming industry. Some examples are shown of the financial strength and revenue possibilities in gaming. As a result it is concluded that gaming communities are strongly tied to the use of platforms, oriented towards social interaction outside of games.

Keywords: *gaming industry, video games, social media, social platform integration.*

INTRODUCTION

Computer video games and gaming begins as an experimental hobby to ostracized groups of young individuals and programmers in the 1980s and 1990s. In the late 2000s it became a medium that attracted major interest in community-hosted events at first on a local and then national level. This past decade games have become a career path both for software engineers and players, a billion dollar investment option for shareholders and a driving force for innovation in hardware processing power and life-like graphics algorithms. As a means to reach a wider audience gaming corporations are using social channels to promote their products and attract new clients. The evolution of gaming as a social

activity has affected the tech world and has changed the global market. In this paper we take a look at the modern state of social interactions within gaming communities as well as how the impact of moving from a traditional single player offline games to multiplayer online ones.

Game theory and gamification has been used by several authors to implement models and methods that can be applied to other information systems (Stoyanova, 2015), (Vasilev and Stoyanova, 2016), (Liu, Santhanam and Webster, 2017), (Koivisto and Hamari, 2019).

We present an overview of significant events in the gaming industry in the past 30 years and a personalized point of view on the direction of social marketing for electronic entertainment hardware and software. The rest of the paper is structured in two sections. The first follows the evolution of games and gaming studios and trends in income models. Section two presents an overview of different social platforms for communication between players in and outside of computer games. Here we discuss the effects of using Facebook, Twitter, Discord and Twitch.tv as external channels for socializing and advertisement.

1. GAMING INDUSTRY EVOLUTION

Computer games have existed since the 1950s. Although not commercially available for a considerable amount of time till 1971, video games fascinated the minds of programmers and regular computer enthusiasts. Now almost 50 years later games run on computers, laptops, tablets, smartphones and variety of consoles. It is estimated that roughly $\frac{1}{3}$ of the world's population or around 2.4-2.5 billion people has played a video game more than once in the past five years.

The evolution of the gaming industry is an interesting topic for discussion. Our main goal is to share observations and present data to back up some claims drawn from said observations. The author's experience in the field started at around the end of the last century. At the time there were considerably fewer computer games.

At the end of the 1990s game developers are looking for funding and an audience. This era of gaming can be marked as the experimental beginning. Different games are being made with the goal to spark interest in potential players. It is at this time that a handful of games would dictate the changes in the gaming industry. Popularity of computer video

games is now not only among developers and players. The first game-oriented Electronic Entertainment Expo (E3) is hosted in 1995 and it is the biggest event in video game history with over 40000 attendees (Buckley, 2013). Titles such as Warcraft, Half-Life, Grand Theft Auto, StarCraft and Heroes of Might and Magic and their respective game studios are about to change history. For the next two decades these and other games shape software and hardware technology, carve the path for gaming as a sport and inspire many to build their own careers, franchises and corporations. The games mentioned above are genre-defining and first titles in their respective long series of sequels, expansions and remakes.

In the first decade of the 21st century computer games are starting to draw attention from a wider audience. This is the era of illegal distribution of games and the creation of mods. While in the West it is possible to buy or acquire a modern game, in Eastern Europe games that require a purchase are illegally distributed and pirated. Around this time internet cafes appear, allowing visitors to pay an hourly fee and play on public computers. Pirate copies of games required a “crack” – an edited or hacked version of key files and configurations. This also allowed for the rise of “mods” or modifications. Modifications are possible due to either open-sourced core game engines or source tool-kits released by the game developers themselves. Mods include new components, levels or maps for existing games which sometimes end as the basis for completely separate titles or even the establishing of new sub-genres in games. Popular mod for Blizzard Entertainment’s Warcraft III is a map called Dota, which would end up being one of the most successful modern day game (Dota 2) for a different company called Valve. Valve also is one of the pioneers in the first person shooter genre when in 1998 it releases Half-Life. The game in turn would inspire mods such as Team Fortress and Counter Strike and hundreds of similar titles, which 20 years later are still played by millions of people.

In 2003 Valve also takes part in creating the first platform for player interactions outside of games. Steam is a software program that is first planned as a channel for distributing updates and bug fixes to Valve’s game titles. Later it became a digital distribution platform or an online

store for many game developers and organisations. The platform sells primarily software, computer games and Valve hardware. By 2019 Steam has attained over 1 billion registered users according to services such as Steamfinder.

Up until 2003-2004 games are a onetime purchase of a physical disc. In 2004 two games launched that required an additional fee or **subscription** each month. Lineage 2 becomes incredibly popular in Asia, while World of Warcraft is sweeping both Europe and the Americas. Both games are still played to this day with World of Warcraft reaching 100 million unique accounts in 2014.

Apart from monthly subscriptions, in 2006 in a free-to-play game called ZT Online a new revenue model appears, which is currently referred to as **loot boxes** (Martinsen, 2007). Loot boxes are considered a gambling mechanic, where for a small fee the player “opens” a box and receives a random prize from a wide collection of known or unknown in-game items or perks. This model alongside microtransactions becomes one of the biggest revenue streams of companies in the next decade. Microtransactions are often season passes, downloadable content, items, abilities, skins or perks that are available for purchase with real money.

Towards the end of 2009 games have become more accessible worldwide. Gaming computers and laptops are sought after for their optimal hardware specifications and build quality. Moderate-sized tournaments are being organized and teams are formed. Sponsors of such events are now looking at opportunities to sign contracts with players. Samsung and Korea Telecom are making deals with South Korean StarCraft players and eventually the game becomes an iconic trademark of the entire country and a pastime for the generations.

Now nearing the end of 2019 gaming has significantly changed. Microtransactions are the primary income source for multiplayer and free-to-play games. Obtaining a rare novelty item or skin brings notoriety to players and the gambling effect to the human brain from opening loot boxes has become a real issue and countries are putting laws in place to protect the youth. In 2017 Belgian Minister of Justice stated that opening loot boxes is mixing money and addiction, which is gambling. In China loot boxes cannot be purchased with real money but can be

given as a reward to players for achieving different goals in-game¹ (Tang, 2018). In Bulgaria only betting websites are monitored by the government and online computer or mobile games are not subject to such regulations.

Another difference between the 2000s and the 2010s in the gaming industry that ties into the previous matter is the popularity of mobile games. With the increased storage and processing power of today's smartphones more and more developers are investing resources into creating mobile games. They are traditionally free-to-play. A way to make revenue from mobile games is to put advertisements in the game and a way to remove them by purchasing in-game passes, premium accounts, etc.

Advances in hardware, internet networks and software also introduced a new way for different types of devices to connect. Consoles, computers and mobile phones now can use cross-play technology. One particular game can attribute its massive success to cross-play. Fortnite is a free-to-play game that was released in 2017 and in 2018 it had amassed more than 100 million players and broke records for most annual earnings for a video game in history with revenue of \$2.4 billion, thanks to microtransactions.

The current decade can also be described as the **game launcher** era. Thanks to Steam's success most big corporations have built their own digital distribution networks. Organisations like Blizzard Entertainment, Electronic Arts and Epic Games have made their games only available via their trademarked game launchers, respectively Battle.net, Origin and Epic Game Store. Platforms like Steam and Good Old Games (GoG) are friendlier towards independent developers (indie developers), although they also sell major game titles.

With the rise of these networks, physical copies of games are now rare and are only really available with expensive collector's editions. The 2010s introduced digital activation codes, which can be physically bought but to obtain the game, purchasers need to input the game code

¹ <http://www.mondaq.com/china/x/672860/Gaming/A+MiddleGround+Approach+How+China+Regulates+Loot+Boxes+and+Gambling+Features+in+Online+Games>

and download the product from platforms such as Steam, Uplay or Epic Games Store.

These past 10 years are also closely tied to the rise of competitive e-sports. Revenue from sales of game copies, merchandise and e-sports is breaking records constantly. One particular game has been going strong for 6 year since its official release in 2013. Dota 2 is a multiplayer online battle arena or MOBA that is now being cited as on par in terms of financing and prestige as traditional sport tournaments. Below is comparison between the prize pools of the biggest Dota 2 tournament – the International and Wimbledon in the last 4 years. Even though Wimbledon's prize pool is larger, the winning team of the International and each member won more than the winner of Wimbledon in 2019. Dota 2 is a team game and the prize is split between 5 players and the award for finishing at 1st place this year was \$15,607,638 (see Table 11).

Table 1

Comparison of Dota 2 International and Wimbledon prize pools by years

Year	Dota 2 International Prize pool²	Wimbledon Prize pool³	Dota 2 International 1st place prize	Wimbledon Prize 1st place prize
2019	\$34,316,118	\$46,263,723	\$3,121,527	\$2,983,748
2018	\$25,532,177	\$41,393,857	\$2,246,831	\$2,976,230
2017	\$24,787,916	\$38,471,938	\$2,172,536	\$2,884,489
2016	\$20,770,460	\$34,211,755	\$1,827,800	\$2,589,231

2. IMPACT OF SOCIAL MEDIA ON VIDEO GAMES

The social aspect is one of the key factors in a game's success. The more discussions are had for a particular game, the more it is likely that it can reach new audience. While single player games thrived in the decades before the Internet was widely spread, online communication and

² Dota 2 Prize Pool Tracker, <https://dota2.prizetrac.kr> [Accessed 31 Aug. 2019]

³ Wimbledon Prize Money, <https://www.perfect-tennis.com/prize-money/wimbledon/> [Accessed 31 Aug. 2019]

chat rooms were becoming very popular in the 1980s (Internet Relay Chat or IRC) and 1990s (Instant Messaging or IM). Multiplayer games push the era of social interaction between gamers further. In 2004 Lineage 2 was among the first documented integrations of a multiplayer online game and an instant messaging software. Lineage 2 allows players to log in while playing into their Microsoft Network Messenger (MSN Messenger) and connect with friends and family outside of the game.

With the growth of the gaming industry, games becoming more popular and richer in content as well as more complex and challenging the need for better communication is apparent. A number of software programs that allow voice over Internet Protocol or VoIP are developed in the early 2000s. Popular programs at the time are Ventrilo, TeamSpeak and Mumble, released around 2002-2005 and are mainly used in gaming. Nowadays one of the most widely used software for VoIP is Discord. Discord is released in 2017, reporting 130 million users in 2018 and now in 2019 doubling that number to 250 million (Kerr, 2019). The platform is free to use and it offers its users instant text messaging with their friends, creating and participating in multiple separate communities and as mentioned already – voice chat. In late 2018 Discord began selling games and acting as a digital distribution network, similar to Steam, with its primary focus still on group text messaging.

In the last decade due to the exponential increase in available games, hard core fans are diversifying their game time with different titles. Due to the massive size of a few big gaming corporations, the sustainability from yearly or biyearly releases of new games is becoming fragile. They are looking into ways to promote their games and attract new customers. Blizzard Entertainment is one such example where during a major patch to its most played game a new feature is introduced aiming to utilize social networks. In early 2015 World of Warcraft receives an update allowing players to sign-in with their Twitter accounts and be able to automatically share and post in-game screenshots directly to the social platform. The initial reaction to this update is met with mixed responses, due to the fact that the game was becoming stale and during major patches players expect in-game content or gameplay balance changes.

In 2016 Twitter integrated its mobile development source kit Fabric with the popular game development engine Unity. Fabric allows software engineers to attach different events happening in games to an API that can create posts and share them on Twitter. Mobile game development is on the rise and one of the most efficient ways to garner interest and reach a wide audience is through advertisements on social media platforms.

It is around the same time that a new platform for video game coverage and discussions really takes off. Twitch.tv is created in 2011 and due to the fact that it is drawing an increasingly big audience after being acquired by Amazon in 2014 for \$1 billion it becomes the biggest website for live video streaming of people playing games. Initially and primarily used as to stream video games to viewers, now the platform offers channels which are marked “IRL” or In Real Life in its Just Chatting subsection, where one streamer would engage in video and audio chat with his viewers, who in turn use instant messaging to communicate back. Twitch.tv has affected computer games in various ways. Major gaming tournaments are being streamed either on Twitch.tv or on various other similar platforms. Some Twitch Streamers have earned more than \$600,000 in a single night due to their popularity and gaming events happening live (Tassi, 2018). Viewers can subscribe to a streamers channel for \$5.00 or use Twitch Prime which is a bonus feature included in Amazon Prime and Prime Video services. Streamers can also receive donations while they are broadcasting live. To date the biggest single donation from a viewer to a streamer amounts to \$75,000 (Puschina, 2019). Although it is difficult to garner a healthy audience that subscribes and donates regularly, more and more people are considering streaming as a part-time or full time job. Streaming has also helped smaller gaming studios promote their games and reach more people, either as potential clients or as investors.

Following Twitch.tv's initial success and the rise of streaming services, Blizzard Entertainment partnered with Facebook in 2016 to bring Facebook Streaming to its game launcher Battle.net, which did not have a big impact on the corporation's players count. Usually older generations of gamers are less keen to engage on general purpose social media

platforms such as Twitter and Facebook. Mobile games currently are popular amongst young children and teens that are more likely to use social media to share video game moments.

To summarize, playing video games has become more and more of a social dynamic for young people with the previous generations being able to afford more expensive games on various digital distribution networks and spending money on in-game microtransactions or to support their favourite streamer or Internet personality. Social interactions have grown from talking with friends face to face to instant text messaging and voice and video chats.

CONCLUSION

The gaming industry has grown rapidly in the last 30 years. Several gaming corporations have managed to build massive followings and continuously provoke audiences with new and exciting technology achievements. Ranging from traditional single player offline games to massively multiplayer online games played on different consoles and devices has created an efficient job force in software, hardware, networking, digital marketing and more. Social platforms and channels have been a welcome addition for gaming communities although traditional marketing on Facebook and Twitter is rather not suited for the mature generations. Platforms such as Discord and Twitch are garnering good results by giving the most modern approaches to social interaction to their users. Digital distribution networks and online stores for games are thriving and pushing gaming as a service further than ever before. Communication and social interaction are one of the key components that drive the popularity of games and they will continue to impact how games are marketed and played.

REFERENCES

1. BUCKLEY, S. (2013) *Then there were three: Sony, Microsoft, Nintendo and the evolution of the Electronic Entertainment Expo*. [online] Available at: <https://www.engadget.com/2013/06/06/sony-microsoft-nintendo-and-the-evolution-of-the-electronic-entertainment-expo/>, [Accessed 31 Aug. 2019].

2. VASILEV, J. and STOYANOVA, M. (2016) Model for communication between an ERP system and a system for gamification. *Computer science and technologies*, 2(1), pp. 18 - 25. [online] Available at: http://eknigibg.net/spisanie-br1-2016_pp.19-26.pdf [in Bulgarian], [Accessed 31 Aug. 2019].
3. KERR, C. (2019) *Discord surpasses 250 million registered users in four years*. [online] Available at: https://www.gamasutra.com/view/news/342526/Discord_surpasses_250_million_registered_users_in_four_years.php, [Accessed 31 Aug. 2019].
4. KOIVISTO, J., and HAMARI, J. (2019) The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, pp. 191-210.
5. LIU, D., SANTHANAM, R., AND WEBSTER, J. (2017) Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS quarterly*, 41(4).
6. MARTINSEN, J. (2007) *Gamble your life away in ZT Online*. [online] Available at: http://www.danwei.org/electronic_games/gambling_your_life_away_in_zt.php, [Accessed 31 Aug. 2019].
7. PUSCHINA, D. (2019) *Twitch Streamer receives donation of US\$75,000 marking a new record*. [online] Available at: <https://www.notebookcheck.net/Twitch-Streamer-receives-donation-of-US-75-000-marking-a-new-record.394481.0.html>, [Accessed 31 Aug. 2019].
8. STOYANOVA, M. (2015) Gamification process in information systems. *Special Issue Proceedings of ETAEc 2015 Conference, Scientific Bulletin – Economic Sciences*, 14, pp. 174-180.
9. TANG, T. (2018) *China: A Middle-Ground Approach: How China Regulates Loot Boxes And Gambling Features In Online Games*. [online] Available at: <http://www.mondaq.com/china/x/672860/Gambling/A+MiddleGround+Approach+How+China+Regulates+Loot+Boxes+and+Gambling+Features+in+Online+Games>, [Accessed 31 Aug. 2019].

10. TASSI, P. (2018) *Twitch Comments On The Record-Breaking Drake-Ninja 'Fortnite' Stream*. [online] Available at: <https://www.forbes.com/sites/insertcoin/2018/03/15/twitch-comments-on-the-record-breaking-drake-ninja-fortnite-stream/#1c02cc4246c6>, [Accessed 31 Aug. 2019].

SOCIAL MEDIA AND THE RECRUITMENT OF IT PROFESSIONALS IN BULGARIAN IT COMPANIES

Velina Koleva¹, Svetoslav Ivanov²

¹ University of Economics Varna/departement "Management and Administration", Varna, Bulgaria, vlkoleva@ue-varna.bg

² University of Economics Varna/departement "Informatics", Varna, Bulgaria, svetoslav_ivanov@ue-varna.bg

Abstract

Digitalization in society drives to the widespread penetration of social media in business. Given their specificity and advantages, they are also gaining popularity in staff management in Bulgarian organizations. The aim of this article is to present and analyze the main sources for the recruitment of computer specialists within the organization. The focus is on social media, their applicability, and their effectiveness. The research is based on an example from some companies in the information technology sector in Bulgaria. An online survey and in-depth interviews are carried out. The survey found that the most effective recruitment channels are job posting sites, social media LinkedIn, personal contacts, database of previous competitions, recruitment agencies and employee referrals.

Keywords: *social media, recruitment, IT companies, HR management, Bulgaria.*

INTRODUCTION

Digitalization in society has led to the widespread penetration of social media in the business. They began to define the way companies work, allowing information and knowledge to be shared with a huge audience. This innovative way of working reflects the management of people in the organization and in areas as recruitment and selection of talents, increasing employee engagement, administration, collaboration, and improving the public image of the company.

One of the most important functions in human resources (HR) management is related to the formation of the human resources of the organization. It aims to provide the necessary people in quantitative and qualitative terms. Its main activities include recruitment, selection, and appointment. Very often the recruitment and the selection are combined,

where recruitment is the first selection procedure (Bach and Edwards, 2013 p.126; Guest, et.al, p.203). Other authors ignore the recruitment and give importance to selection procedures, only (Sims and Ronald, 2014, p.139; Rakowska and Babnik, 2015, p.168). In our opinion, recruitment is a very important activity in the formation of HR. The recruitment's quality depends on its effectiveness. While recruitment is about attracting candidates, the selection focuses on selecting the most suitable candidate for the vacant position. The properly selected employees provide a competitive advantage to the organization.

The recruitment of human resources is connected traditionally with identifying, searching and attraction of potential candidates for a vacancy in the organization. After "screening", those who are as close as possible to the requirements of the position are selected. The more candidates are attracted, the more likely is to have suitable ones among them, i.e. the potential staffing base will be richer for the organization.

The sources of human resources for the organization are two types - internal and external. The internal recruitment allows vacancies to be filled by current employees of the organization. It is performed public announcement of the vacancies, which is an invitation to all company employees. Occupation of vacant positions by internal employees can be accomplished in two ways - hierarchy lifting and staff rotation.

The external recruitment provides many more alternatives. The external sources are diverse and provide different labour in characteristics, potential qualities, and cost. Each organization should formulate own recruitment policy depending on its strategy and capabilities, as well as from the job characteristics, which is looking for suitable candidates. The channels for recruitment of external sources are:

- Employment agencies.
- Universities and other educational institutions.
- Professional journals.
- Consulting companies.
- Database of candidates from previous competitions/applications.
- Industry organizations and competitors (incl. headhunting).
- Recruitment exhibitions (Career Days, scientific events, forums).
- A network of professional contacts and acquaintances (networking).
- Other external sources.

1. SOCIAL MEDIA AND RECRUITMENT

The increasing accessibility to the Internet and high-level information technologies make the recruitment increasingly digital, giving it a new look. Professional job posting sites are widely used. The most popular job posting sites for IT companies in Bulgaria are in Table 1. The list is sample and does not pretend to be full or the order to bring any preference.

Table 1

**List of the most popular job posting websites in Bulgaria
for IT job positions**

jobs.bg	itjobs.bg	back2bg.com
zaplata.bg	jobspartner.bg	obiavi.bg
rabota.bg	buljobs.bg	jobco.bg
jobtiger.bg	bgtrud.com	computer- world.bg/careerzone
karieri.bg	sloth.works/bg	work.bg

Source: *Horizons*, available from <https://www.horizons.bg/bg/portali-zarabota-v-balgariya/>

Also, organizations use their own organization's website, where they publish the vacancies. Another alternative is to use social media sites for company job postings and company-related information. A sample list of social media sites is listed in Table 2.

Table 2

List of social media sites used for company-related postings

Facebook	Linkedin	Google+
Instagram	Twitter	Pinterest

Other places to publish job postings are discussion forums, focused groups, micro-blogs, social bookmarking (for more information see Veleva, 2015), and others.

Our specific interest in the current research is social media. According to Kaplan and Haenlein social media are "applications that give users access to connect by creating profiles with personal information, inviting friends and colleagues to access that information, and sending emails and messages to each other" (Kaplan and Haenlein, 2012, pp.60-62). Gastelum and Whattam state that "Social media consists of tools, techniques, and technologies that use the Internet to support communication in an open environment. Types of social media include social networking sites (Facebook), microblogs (Twitter), blogs (Blogspot), chat (AIM), open-source mapping (Wikimapia), and photo and video sharing (Flickr, Picasa, YouTube)" (Gastelum and Whattam, 2013, p.4). According to Parusheva, Alexandrova and Petrov social media are part of contemporary concepts Web 2.0 and Web 3.0, which offer many possibilities, including new methods of communication, interactivity, connections, learning, and education (Parusheva, Alexandrova, Petrov, 2017, p.19). For the purpose of this article, we maintain the view that social media is a means of communication, connectivity and exchange via the Internet that contains one or more of the main types of media mentioned - social networks, blogs, microblogging, messaging, content sharing.

According to Berkowitch (2010, p.10), the recruitment of candidates for social media vacancies has two aspects:

- A leading participant in this process is the organization seeking candidates for vacancies that meet certain parameters required by the position - education, skills, experience.
- The active role is taken by the job seeker who applies for the vacancies posted by the organizations in company social profiles.

It is important to note that social media tools and techniques in use cannot replace fully the traditional processes of recruitment and selection. These tools and techniques are additional to the process. The social media have a number of advantages over traditional recruitment sources. Some of these advantages are related to:

- Access to a large number of candidates and the ability to quickly communicate with candidates - social media is an excellent channel for reaching a large number of people in a short time. In addition, a high percentage of organizations use them to find "passive candidates" - the candidates, who do not seek a new job actively but could be attracted to a good offer.

- Saving time and costs - social media require significantly less cost than traditional sources.

- Opportunity to generate interest in jobs - it is done by publishing interesting information and initiating discussions. It is conditioned by the interactive way of working in social media.

- Opportunity for pre-candidate / post-survey of the applicant - social media enables the employer to receive information about the candidates on the basis of a shared profile and content. There are cases of candidates being dropped from the selection process due to their negative presence on the network - inappropriate statements, non-compliance with media rules, illegal content, etc.

At the same time, social media have drawbacks. Some of these are listed below:

- Ethics and protection of personal data.
- Human resource specialists need training - the entry of social media as a form of recruitment requires qualified and trained HR specialists.

- Shared information on social media is not completely controllable. It can be difficult to filter the audience reached and non-eligible people may apply for a vacancy.

The main conclusion that could be drawn from the above is that recruitment is an essential element of the HR formation process in an organization. It is implemented through two main channels - internal and external. Our goal is to analyse them, focusing primarily on external recruitment sources in one of their modern forms - social media, following the example of organizations in the information technology sector in Bulgaria.

2. THE SURVEY AND RESPONDENTS

To achieve the stated goal, specific research methods and tools are applied. In theoretical and methodological terms, it is used as a systematic and comparative approach, a descriptive method. Empirically, a survey and in-depth interviews are conducted. The empirical research is pilot and serves as a basis for further in-depth study of the topic. To determine the sample size, the Classification of Economic Activities –

2008, with code 62 “Activities in the field of information technology” is used. Of interest to us is to share the companies according to The European Commission definition for small and medium enterprises (SME) according to the number of employees, namely: micro enterprises (under 10), small enterprises (between 10 and 50), medium enterprises (between 50 and 250), and large enterprises (above 250) (European Commission, 2019). For the purposes of the survey, we also interviewed several specialized agencies for IT professional's recruitment. Online surveys in the Bulgarian language are sent to 102 respondents from Bulgarian organizations who take or participate in taking decisions about the company staff. 38 respondents participated in the survey by the respondent method. Data collection is done with an online survey created through Google Forms. The collected empirical data are processed with the use of Google Sheets and Excel 2016 software. The list of survey questions and responses is given in Annex 1 to this article and translated from Bulgarian.

The respondents in the survey are tested for the location of surveyed respondent's organization activity, company size, company activity, and respondent's position.

The highest percentage of respondents (and surveyed organizations) are in Sofia (73%). 22% of the respondents are companies with the location of activity in multiple big cities in Bulgaria. Usually, these organizations have a head office in Sofia, but there are offices in Sofia, Plovdiv, Varna, Ruse, Burgas, Veliko Tarnovo as well. It is more realistic to distribute this percentage between cities. 5% indicated the city of Varna, only.

Organization size is tested with the number of working employees according to the company classification. The largest share is for small companies (37%), then medium-sized companies (about 32%), micro firms (just 21%) and large organizations last (10%).

The main percentage (share) of the companies surveyed are staffing agencies (nearly 47% of respondents) and IT outsourcing (26%). The share of companies developing own software products is smaller (21%). The interviewed respondents reported their position held in the company. Most of them represent the position of Consultant Recruitment and Selection (37%), and then the Human Resources Specialist - 32%. These are the professionals who are involved in the recruitment process within

the organization and have a real look at the challenges that accompany it. Human Resources Managers are 21% and up to 10% of others. The latter can be explained by the fact that in small companies there is no specialist for the “recruitment and selection” activity and the role is performed by the manager, office assistant, marketing specialist or other employees.

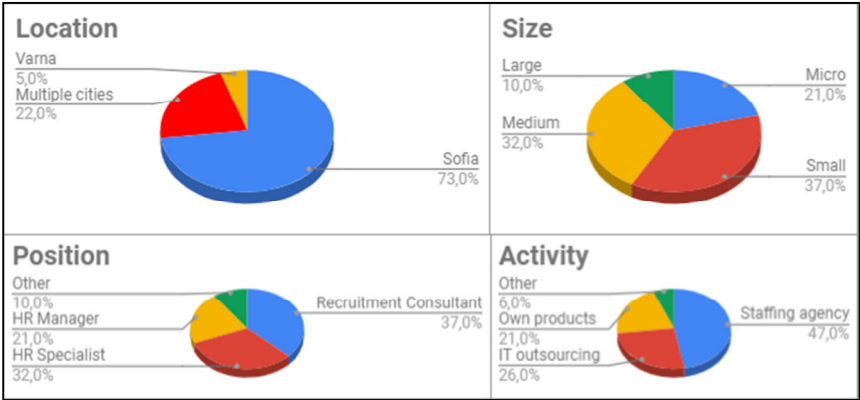


Figure 1. The respondent's profile - location, organization size, activity, position

3. THE SURVEY RESULTS ABOUT RECRUITMENT SOURCES

The basic question which the respondents had to answer was related to the *types of channels for recruitment* of candidates in the surveyed organizations (see question 5 in Annex 1). Respondents were entitled to more than one answer. A large part of the respondents cites internal sources as an opportunity to find potential employees for vacant positions (53%). In-house recruitment actually has a number of advantages over external recruitment: it is a quick and inexpensive method that involves fewer risks, the adaptation period is minimal and there is less chance of errors, as applicants are internal employees who are well versed in the organization. At the same time, the in-house kit is a great tool for the performance appraisal system and provides opportunities for job diversity and career development.

Another source is contacts of people working in the company - personal contacts (89%) and recommendations from company employees (74%). This high percentage is explained by the fact that personal trust is valuable to employers. On the one hand, employees know the company, its needs, and its culture. On the other hand, they may have acquaintances, friends, or former colleagues who they deem appropriate for a vacancy. In addition, these recruitment sources imply low financial resources and low time consuming, attracting candidates who are not actively looking for a new job but are interested in change. In IT companies, where there is a strong staff shortage, bonuses are given for finding and hiring a candidate found with the help of a company employee.

The use of advertisements, whether through own site (68%) or ad sites (68%), has already become a classic form of recruitment. Here, the percentages for both sources are the same because they are likely to be used in parallel by employers. The sources for the selection of qualified Career Days (Forums) and young professionals (Universities) are less represented (53% and 32%). This is probably because of the fact that the specialists here do not have enough experience and the aim is to build an employer brand, through example and conversations with colleagues in the company. The use of recruitment agencies (47%), and database of candidates from previous competitions (63%) ranks near the forums. In both cases, it relies on available contacts. In some cases, agencies replace and often complement the company's "Recruitment and selection" specialist. A small number of respondents (16%) rely on the "theft" of talent from competing organizations.

In the course of the study, we raised the next question of the *effectiveness of the recruitment channels* in the organizations surveyed (see question 6 in Annex 1). Internal sources have a low score (5%), which is explained by the specifics of the IT business and the dynamics of the processes. Personal contacts and recommendations from employees are an extremely strong factor (26% and 16% respectively). As they are also cited as the most commonly used factor, they are objectively the most effective. Job sites are also the preferred source (26%). Forums and universities are not an effective source or are lacking as such. Rather, they are used to generate interest and to inform applicants, but without direct recruitment. Therefore, the respondents did not cite them as a strong

factor. It appears that the databases of previous competitions and selection agencies also have similar results, but significantly lower than the other methods (16% and 11%). These results may be lower and due to the participation of respondents from selection agencies who did not indicate this method as most effective and by default excluded it.

Finally, we compared the types of channels used to recruit candidates and their effectiveness. The results are presented in Table 3. The last column reflects the conversion rate in which the channel indicated by the respondents was also identified as the most effective. The comparative analysis of the data in Table 3 allows us to make a conclusion that there is a discrepancy in the degree of use of the collection channels and their efficiency. The most used channels are not always the most effective. Calculating the conversion rate from the percent of the used channel to the percent of the effective channel, it is found that there are discrepancies in the extent of recruitment channels usage. The most effective recruitment channels are job posting sites, personal contacts, the database of candidates from previous competitions, recruitment agencies and recommendations from employees in that order.

Table 3

List of recruitment channels, their use, effectiveness, and conversion rate

Source/Channel	Used channel	Effective channel	Conversion rate
Internal sources	53%	5%	9%
Company website	68%	0%	0%
Educational institutions	32%	0%	0%
Personal contacts	89%	26%	29%
Database of candidates	63%	16%	25%
Competitors	16%	0%	0%
Employee recommendations	74%	16%	22%
Recruitment exhibitions	53%	0%	0%
Job posting websites	68%	26%	38%
Recruitment agencies	47%	11%	23%

4. THE SURVEY RESULTS ABOUT SOCIAL MEDIA

The study of social media used for recruitment has been examined with three questions about usability, effectiveness, and reason for effectiveness (see questions 2, 3 and 3.2 in Annex 1). The results of the usability show that over 80% of the respondents use social media for recruitment. A preferred social media for this purpose is LinkedIn as a major channel for searching for computer professionals. Half of the respondents, on the other hand, cited Facebook (53%) as an opportunity to search for and attract potential candidates for vacant positions. There are other options such as job posting sites and communications software, most of which are not social media (less than 10% of respondents). This may be due to a lack of understanding of the term "social media" or the respondents use the job posting site or communication software is the recruitment media. Summarized data are given in Table 4.

Table 4

Summary of responses for social media and reasons for effectiveness

Social Media	Count of candidates	Professional auditory	Professional media	None or incorrect	Most effective
LinkedIn	21%	21%	31%	31%	84%
Facebook	5%	21%	0%	5%	21%
Other media	0%	5%	5%	5%	10%
None/incorrect	5%	0%	0%	5%	5%
Total	31%	47%	36%	46%	-

The research of effectiveness requires respondents to answer the question of which are the most effective media. The additional answers were changed to a section "Other media" and the wrong one to "None/incorrect". Then the received results are 84% for LinkedIn as the most effective media, 21% for Facebook (in most cases in combination with LinkedIn) and 10% for "Other media" such as Opportunity, Codes Rank, GitHub listed as additional. Along with the issue of channel effectiveness, we also looked at the reasons for this performance. Respondents indicated individual responses to this question. Their semantic anal-

ysis indicated that a summary could be made of a key trait expressed by various means of expression in the answers. The first criterion is the number of candidates given by the media. The second criterion is quality, that is, the media audience is suitable for the search for candidates. The third criterion relates to whether the media itself is intended for professional purposes and contains professional information for the candidates (professional experience, projects, examples). According to these criteria, the results could be summarized and present them in Table 4. It is important to clarify that there are respondents with an invalid answer for the media or a missing reason, so there is an option "None or incorrect". The amount of percent on the horizontal row is different from the final and 100% due to the provision of more than one reason for a media. Also, the respondents were asked whether they use different social media for different positions (see question 4, Annex 1). 78% of the respondents answered that they use different social media. The analysis of the data above allowed us to draw several important conclusions from the data in Table 4:

- LinkedIn is the most preferred social media because it has a professional audience and represents professional expertise.
- For many respondents, being a media professional is understood by default as a cause of effectiveness (and therefore the high % missing reason for LinkedIn).
- Facebook is second, but there the focus is only on specific Facebook Groups because of the audience.
- The strong focus on presenting professional expertise is leading for other media.

CONCLUSION

The current article examines the main sources for the recruitment of computer professionals in Bulgarian organizations. In particular - social media in theoretical and practical terms. The practical part consisted of a survey. The survey found that the most effective recruitment channels are job posting sites, personal contacts, the database of previous competitions, recruitment agencies and employee referrals in this order. Social media is of the highest importance, with 80% of the respondents using it precisely to recruit candidates. The most preferred social media is LinkedIn because of its professional focus.

REFERENCES

1. BACH, S., EDWARDS, M. (2013) *Managing Human Resources: Human Resource Management in Transition*. John Wiley and Sons.
2. BERKOWITCH, A. (2010) *What Does Success Look Like for Your Company: Social Media Starting Points with Measurable Returns. People & Strategy*. Vol. 33, Issue 3
3. EUROPEAN COMMISSION (2019) *What is an SME?* Available from: https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en [Accessed: 2/08/2019].
4. GASTELUM, Z., WHATTAM, K. (2013) *State-of-the-art of social media analytics research*, Available from <http://www.pnnl.gov/main/publications/external/technical_reports/PNNL-22171.pdf> [Accessed 8/08/2019]
5. GUEST, D., et al. (2012) *HRM and Performance: Achievements and Challenges*. John Wiley and Sons.
6. HORIZONS (2019) *Portali za rabota v Bulgaria* [in Bulgarian] [Online]. Available from <https://www.horizons.bg/bg/portali-za-rabota-v-bulgariya/> [Accessed 8/08/2019]
7. KAPLAN, A., HAENLEIN, M. (2010) *Users of the world, unite! The challenges and opportunities of Social Media*. *Business Horizons*. 53(1). pp.59-68, Available from https://www.researchgate.net/publication/222403703_Users_of_the_World_Unite_The_Challenges_and_Opportunities_of_Social_Media [Accessed 28/08/2019]
8. KAPLAN, A., HAENLEIN, M. (2012) *Social media: Back to the roots and back to the future*. *Journal of Systems and Information Technology*. 14(2). Available from: https://www.researchgate.net/publication/263388197_Social_media_Back_to_the_roots_and_back_to_the_future [Accessed: 15/08/2019]
9. PARUSHEVA, S., ALEKSANDROVA, Y., PETROV, P. (2017) *A Study of the Use of Social Media in Higher Education Institutions in Bulgaria*. 4th International Multidisciplinary Scientific Conferences on Social Sciences & Arts SGEM 2017, Albena, SGEM, 4, 2017, Book 1. pp.19-26. Available from <https://www.researchgate.net/publication/319354330_A_STUDY_OF_THE_USE_OF_SOCIAL_MEDIA_IN_HIGHER_EDUCATION_INSTITUTIONS_IN_BULGARIA> [Accessed 8/08/2019]

10. RAKOWSKA, A., Babnik, K. (2015) Human Resources Management Challenges: Learning & Development. ToKnowPress
11. SIMS, R. (2014) Organizational Success through Effective Human Resources Management. Westport, Connecticut: QUORUM BOOKS.
12. VELEVA, M. (2015) Modernite tehnologii v savremennoto upravlenie na choveshkite resursi v turizma: Vazmozhnosti i perspektivi. Sbornik dokladi ot jubilejna mezhdunarodna nauchna konferenciya. Varna: Universitetsko izdatelstvo Nauka i ikonomika. pp.774-782 (Bulgarian)

ANNEX 1: THE SURVEY QUESTIONS AND ANSWERS

Title: Social media and recruitment of IT professionals in Bulgarian IT companies

Dear colleagues,

In connection to a scientific research and our desire to analyse the topic "Social Media and the Recruitment of Computer Professionals in Bulgarian IT Companies", we would like to invite you to participate in this survey. It will take about 5 minutes to complete. More than one answer is possible.

Your opinion is very important to our project. Thank you for your cooperation!

1. Do you use social media and what is the purpose you use it for? *
 - ☐ Presentation of the organization activity
 - ☐ Building an employer brand
 - ☐ Recruitment of candidates
 - ☐ Demonstration of work and working conditions
 - ☐ Communication with business partners and other organizations (customers, partners, suppliers)
 - ☐ We do not use social media
 - ☐ Something else:
2. If you use social media as a recruitment tool, which social media do you use predominantly in your organization? *
 - ☐ Facebook
 - ☐ LinkedIn

- ☐ Google+
- ☐ Instagram
- ☐ Twitter
- ☐ Pinterest
- ☐ Something else:

3. Which social media are most effective in recruitment of candidates in your organization: *

- ☐ Facebook
- ☐ LinkedIn
- ☐ Google+
- ☐ Instagram
- ☐ Twitter
- ☐ Pinterest
- ☐ Something else:

3.2. Why do you think the media you specify is more effective than the others? (free text)

4. Do you use one or different social media to recruit candidates for different positions in the organization? *

- YES (Different)
- NO (One)
- We do not use social media

5. What other recruitment channels do you use? *

- ☐ Internal sources – a hierarchy promotion/rotation
- ☐ Own site
- ☐ Universities and other educational institutions
- ☐ Personal contacts (friends, recommendations from friends and family)
- ☐ Recommendations from company employees
- ☐ Database of candidates from previous competitions in the company
- ☐ Competitive companies
- ☐ Selection Fairs (Scientific Events, Career Days, Other Forums)
- ☐ Ad Sites (jobs.bg, zaplata.bg, stackoverflow.com and more)
- ☐ Recruitment agencies

6. Which channel is the most effective one in recruiting candidates in your organization? *

- Internal sources – a hierarchy promotion/rotation
- Own site
- Universities and other educational institutions
- Personal contacts (friends, recommendations from friends and family)
- Recommendations from company employees
- Database of candidates from previous competitions in the company
- Competitive companies
- Selection fairs (Scientific Events, Career Days, Other Forums)
- Job posting sites (jobs.bg, zaplata.bg, stackoverflow.com and more)
- Recruitment agencies

6.2. Why do you think the channels you specify are more efficient than the others? (free text)

7. How big is the company you work for? *

- up to 10 employees
- up to 50 employees
- up to 250 employees
- more than 250 employees

8. Type of business of the company you represent: *

- IT Outsourcing
- Process outsourcing
- Recruitment
- Own software products
- Something else:

9. Your position: *

- Human resources manager
- Human resource specialist
- Recruitment and Selection Consultant
- Something else:

10. Company localization *

- Sofia
- Plovdiv
- Varna
- Veliko Tarnovo
- Ruse
- Burgas
- In several cities from Sofia, Plovdiv, Varna, Veliko Tarnovo, Ruse, Burgas
- Something else:

Legend:

* – mandatory question

□ – multiple answers per question are allowed

○ – a single answer per question is allowed

(free text) – allows an answer in free text

“Something else:” – allows to enter an additional response in free text.

MODELING AND SIMULATION OF SOME FUNCTIONS OF TWO INDEPENDENT RANDOM VARIABLES

Deyan Mihaylov¹

¹ University of Economics-Varna/Department "Statistics and Applied Mathematics", Varna, Bulgaria, dgmihaylov@ue-varna.bg

Abstract

This paper represents a way to analyze the duration of some processes, which have little complexity. They can be decomposed into limited number of elementary activities. It is supposed that the activity durations are random variables. Hence, the total duration can be represented as a function of random variables. The functions Sum, Maximum and Minimum of two independent random variables are discussed. One way to obtain the distribution functions of results is represented. The mathematical models are tested by simulations. The results, obtained by mathematical and simulation models are similar. The represented method can be used in analysis of processes or objects which have non-deterministic parameters.

Keywords: *Modeling and Simulation; R-language; Time-measured random variables.*

INTRODUCTION

Every complex process can be decomposed into a logical sequence of subprocesses. The subprocess is a process of less complexity and it can be decomposed, too. The decomposition can be continued until reaching of elementary activities. These activities can run sequentially or in parallel.

The processes which have little complexity will be discussed in this paper.

If activities can run sequentially, each activity starts after the previous one is completed. In this case the total duration of these activities is a sum of the individual durations.

If activities can run in parallel, there are two basic cases. If the condition of completion of the process is all included activities to be completed, the total duration of the process is the maximum of individual durations. If the condition of completion of the process is at least one of

included activities to be completed, the total duration is the minimum of individual durations.

Three processes, which are the least complicated are represented in Figure 1.

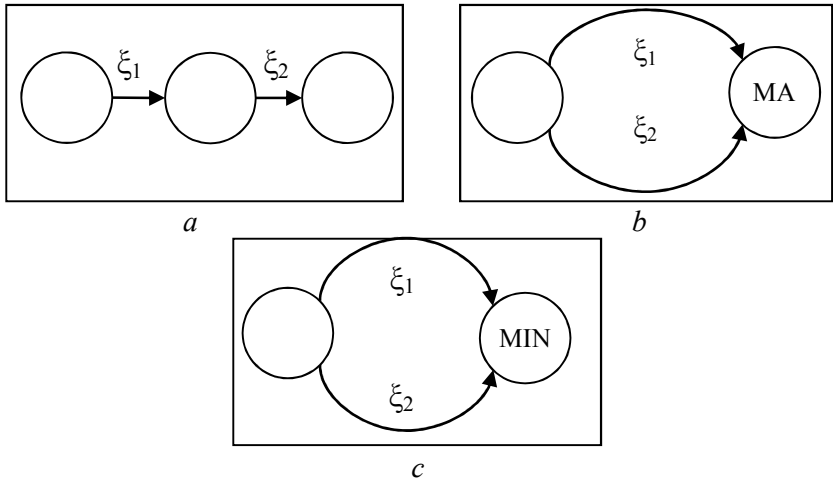


Figure 1. Processes, which are the least complicated

Many factors influence the length of every activity and, therefore, the time for its execution is a random variable. The duration of each process is a function in terms of the durations of elementary activities, i.e. a function of random variables.

The aim of this paper is to represent a way to study the duration of the subprocesses. This is just one of characteristics. It is not necessary to do a full study, but it is enough to construct a mathematical model. It represents behaviour of the object through equations. In addition, the model is a testable representation (Dubois, 2018) (p. 1). The possible test of mathematical model is the simulation.

Two tasks will be done. The first one is to represent formal (by equations) description of the duration of process (subprocess) as a function of random variables. The second one is to test this model by computer simulation.

1. METHOD

Let the activity durations are the independent continuous random variables ξ_1 and ξ_2 with cumulative distribution functions (CDF) F_1 and F_2 and probability density functions (PDF) f_1 and f_2 . The argument of each function is time. According to Kolmogorov's definition (Kolmogorov, 1950, p23).

$$F_j(t) = P(\xi_j < t) . \quad (1)$$

The CDF is equal to the probability that the process has completed before t .

The PDF is the first derivative of CDF:

$$f_j(t) = F'_j(t) . \quad (2)$$

Every random variable is characterized fully by its distributed function. Hence, the obtaining of distribution function gives full information of the process duration.

Total duration of the process of Figure 1.a is equal to $\xi_1 + \xi_2$, of Figure 1.b is equal to $\max(\xi_2, \xi_2)$, of Figure 1.c is equal to $\min(\xi_2, \xi_2)$.

Let $\eta = \xi_1 + \xi_2$. The probability density function of η is

$$f_\eta(t) = \int_{-\infty}^{+\infty} f_1(z) \cdot f_2(t - z) dz . \quad (3)$$

The random variables ξ_1 and ξ_2 are time measured, so they are non-negative. In this case the equation (2) reduces to (Feller, 1971, p. 7).

$$f_\eta(t) = \int_0^t f_1(z) \cdot f_2(t - z) dz . \quad (4)$$

Let $\theta = \max(\xi_1, \xi_2)$. The cumulative distribution function of θ is (Venttsel' and Ovcharov, 2000, p.377).

$$F_\theta(t) = F_1(t)F_2(t) . \quad (5)$$

Let $\nu = \min(\xi_1, \xi_2)$. The cumulative distribution function of ν is (Venttsel' and Ovcharov, 2000) (p. 374)

$$F_\nu(t) = 1 - [1 - F_1(t)][1 - F_2(t)] . \quad (6)$$

The probability density functions of θ and ν can be obtained by using equation (2).

The way to construct mathematical models of given cases is represented above. The first task is solved.

The computer simulation can be performed using suitable software, as R-language. R is available as Free Software (<https://www.r-project.org/>). It provides built-in random-number generators for many known distributions and graphical techniques.

2. EXAMPLE

Let ξ_1 and ξ_2 have exponential distributions with parameters λ_1 and λ_2 , i.e. $\xi_1 \sim \text{Exp}(\lambda_1)$, $\xi_2 \sim \text{Exp}(\lambda_2)$. The CDFs and PDFs are

$$F_1(t) = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-\lambda_1 t}, & t > 0 \end{cases};$$

$$f_1(t) = \begin{cases} 0, & t \leq 0 \\ \lambda_1 e^{-\lambda_1 t}, & t > 0 \end{cases};$$

$$F_2(t) = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-\lambda_2 t}, & t > 0 \end{cases};$$

$$f_2(t) = \begin{cases} 0, & t \leq 0 \\ \lambda_2 e^{-\lambda_2 t}, & t > 0 \end{cases}.$$

From the equation (4) we obtain

$$f_\eta(t) = \int_0^t \lambda_1 e^{-\lambda_1 z} \cdot \lambda_2 e^{-\lambda_2(t-z)} dz = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (e^{-\lambda_2 t} - e^{-\lambda_1 t}).$$

Hence, the PDF of η is

$$f_\eta(t) = \begin{cases} 0, & t \leq 0 \\ \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (e^{-\lambda_2 t} - e^{-\lambda_1 t}), & t > 0 \end{cases}. \quad (7)$$

This is the well-known generalized Erlang distribution.

From the equation (5) we obtain

$$F_\theta(t) = (1 - e^{-\lambda_1 t})(1 - e^{-\lambda_2 t}) = 1 - e^{-\lambda_1 t} - e^{-\lambda_2 t} + e^{-(\lambda_1 + \lambda_2)t}.$$

According to equation (2)

$$\begin{aligned} f_\theta(t) &= (1 - e^{-\lambda_1 t} - e^{-\lambda_2 t} + e^{-(\lambda_1 + \lambda_2)t})' = \\ &= \lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)t}. \end{aligned}$$

The PDF of θ is

$$f_{\theta}(t) = \begin{cases} 0 & , t \leq 0 \\ \lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} - (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)t} & , t > 0 \end{cases} \quad (8)$$

From the equation (6) we obtain

$$F_v(t) = 1 - [1 - (1 - e^{-\lambda_1 t})][1 - (1 - e^{-\lambda_2 t})] = 1 - e^{-(\lambda_1 + \lambda_2)t}.$$

According to equation (2)

$$f_v(t) = (1 - e^{-(\lambda_1 + \lambda_2)t})' = (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)t}.$$

The PDF of v is

$$f_v(t) = \begin{cases} 0 & , t \leq 0 \\ (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)t} & , t > 0 \end{cases} \quad (9)$$

Let $\lambda_1 = 0,4$ and $\lambda_2 = 0,7$. The mathematical models are calculated using equations (7), (8) and (9). The simulation models are obtained using R-language. 100000 simulations are done. The comparisons between mathematical and simulation models are presented in Figures 2, 3 and 4.

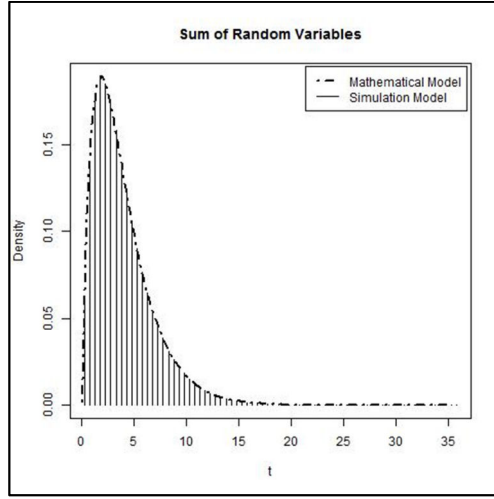


Figure 2. Probability Density Function of $\eta = \zeta_1 + \zeta_2$
 $\zeta_1 \sim \text{Exp}(0,4), \zeta_2 \sim \text{Exp}(0,7).$

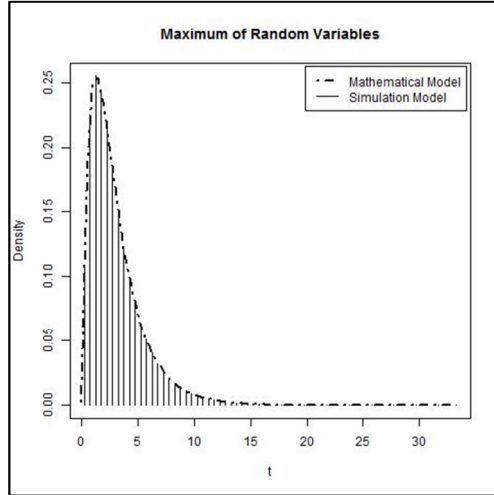


Figure 3. Probability Density Function of $\theta = \max(\zeta_1, \zeta_2)$.
 $\zeta_1 \sim \text{Exp}(0,4), \zeta_2 \sim \text{Exp}(0,7).$

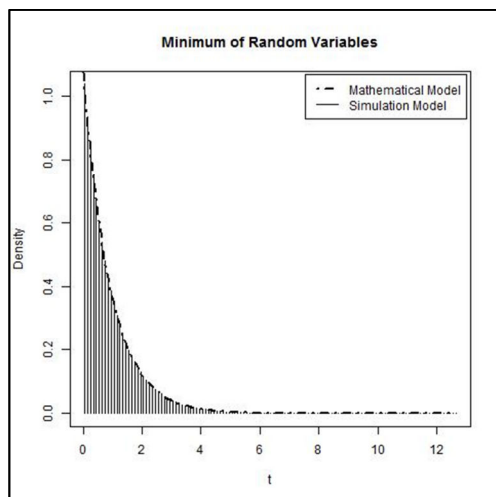


Figure 4. Probability Density Function of $v = \min(\zeta_1, \zeta_2)$.

$$\zeta_1 \sim \text{Exp}(0, 4), \zeta_2 \sim \text{Exp}(0, 7).$$

The graphs show that the results, obtained by mathematical and simulation models are similar. Hence, these two kinds of models are equivalent.

CONCLUSIONS

Some simple structures are studied above. But the method to analyze more complex structures is the same. For example, the total duration of process in Figure 5 is equal to $\max[(\zeta_1 + \zeta_2), \zeta_3]$

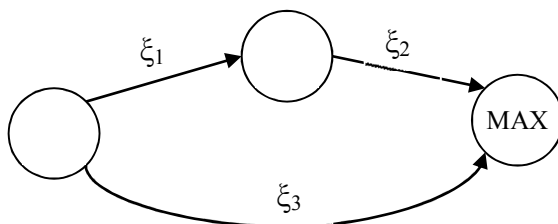


Figure 5. Example of process, which has higher complexity

The estimation obtained from the mathematical models is accurate, but it will be noted, that the creation of mathematical models of processes, which have very high complexity can be very difficult, or impossible. In this case only possible way to analyse the process duration is the simulation.

A comparison between models should be made for simple cases. In this way, the researcher can verify better that the simulation tools are good. Only then the simulation models can be used to study more complex processes.

The comparison between mathematical and simulation models can be applied not only to study duration of processes. For example, (Kuyumdzhiyev, 2016, pp. 51-54) uses mathematical and simulation models to investigate the probability of generating test combinations which satisfy given conditions.

At last, the comparison between models can be useful in education. The simulation can be an experiment that confirms the theory.

REFERENCES

1. DUBOIS, G. (2018). Modeling and simulation: challenges and best practices for industry. Boca Raton: Taylor & Francis
2. FELLER, W. (1971). An Introduction to Probability Theory and Its Application 2th ed., Vol. 2. New York: Wiley&Sons.
3. KOLMOGOROV, A. N. (1950). Foundations of the theory of probability. New York: Chelsea Publ. Co.
4. KUYUMDZHIEV, I. (2016). Developing an Algorithm for Generating Computerized Test Combinations. Izvestiya. Journal of University of Economics-Varna , 60 (1), pp 47-60.
5. VENTTSEL', E. S. and OVCHAROV, L. A. (2000). Teoriya veroyatnostey i ee inzhenernye prilozheniya. Moskva : Vysshaya shkola.

USING ASSOCIATION RULES AS SEMANTICS FOR DOMAIN ONTOLOGIES

*Oumar Sy*¹

¹ *Université Gaston BERGER/U.F.R Sciences Appliquées et de Technologie (Comp. Sc.), Saint-Louis, Sénégal, oumar.sy@ugb.edu.sn*

Abstract

Association rules and ontologies both are domain-based knowledge. However, the first is an unexpected discovered knowledge from databases, while the second is a priori knowledge. In this paper, based on a generic meta-schema as common referential and theoretical foundation, we show how a set of computed and pruned association rules can be useful to enriching a domain ontology. To this end, the meta-schema is scanned with the itemsets appearing in each association rule. Then, according to the formal links between the concepts or the attributes involved, a semantic-based check constraint is built. As a result, the ontology and the database can continuously be monitored with a new semantics.

Keywords: *Domain ontology, Association Rule, Semantics, Meta-Schema, Relational database, SQL.*

INTRODUCTION

Association rules (AR) and ontologies both are domain-based knowledge. Yet, while ontologies are a priori background knowledge, AR are unexpected implicative and interesting tendencies discovered in databases (Fayyad, Piatetsky-Shapiro, Smyth and Uthurusamy, 1996). On the other hand, ontology learning is concerned with knowledge acquisition (Buitelaar, Cimiano and Magnini, 2005). Moreover, ontologies' creation and maintenance helped in the emerging of many tools (Noman, Siddiqui and Shaikh, 2010). Thus, AR, which are knowledge obtained by the Knowledge Discovery from Databases (KDD), can be evaluated and interpreted as new semantic-based knowledge in the underlying domain ontology development process. Indeed, following (Tomic, Drenjanac, Lazendic, Hörmann, Handler, Wöber, Schulmeister, Otte and Auer, 2014), we emphasize that a rule constraint may represent a predicate that specifies, e.g., a relationship, the instantiation of variables, or a SQL aggregate function. In our case, as examples, (V.FGERM

≥ 90) AND (V.FGERM=98), V.CATEGORY= 'JAYA' are rule constraints, where V represents the concept of variety of rice seed, and FGERM the attribute describing the faculty of germination of the plant. Intuitively, we can use these two rules to build a single rule of the form if V.CATEGORY = 'JAYA' then (V.FGERM ≥ 90) AND (V.FGERM=98), or conversely. Yet, such a rule can undoubtedly be understood as an AR of the form V.CATEGORY= 'JAYA' \Rightarrow V.FGERM $\in [90, 98]$. In other words, on one side, "variety of rice", V in the above example, which is a concept described, among others, by the attribute CATEGORY, is knowledge being of the domain and, on the other side, certain values of these attributes calculated, e.g., by an KDD's algorithm can be applied as additional semantics to enhance the ontology-based background knowledge.

The posed problem is the useful combination of these two knowledge enriching each other for decision-making. Such a problem stated, according to (Dou, Wang and Liu, 2015), ontologies have been introduced to semantic data mining for three purposes, namely: (i) To bridge the semantic gap; (ii) To provide data mining algorithms with a priori knowledge; (iii) To formalize the representation of the data mining flow.

Now, most of the works focus for the improving of the data mining's tasks (Furletti, 2009; Brahimi, Araour and Yahya, 2012; Vidya, 2012; Saravanam, Vijayalakshmi and Joseph, 2017), especially the pre-processing, e.g., (Antunes, 2007, pp. 43-54), and post-processing tasks, e.g., (Marinica and Guillet, 2009, pp. 76-80; Marinica and Guillet, 2010; Albert, Fayaz and Babu, 2014). These approaches are widely reviewed in (Lina Zhou, 2007; Hazman, El-Beltagy and Rafea, 2011; Dou, Wang and Liu, 2015). Facing to the same problematic, Rajan and Dhas (Rajan and Dhas, 2012) used an ontology-based classification of web pages through association rules. In (Tatsiopoulos and Boutsinas, 2009) association rules were used for ontologies mapping. In (Galárraga, Teflioudi, Hose and Suchanek, 2013), the authors mine AR in an ontological knowledge base. Various attempts to combine ML techniques and ontologies, semantically annotated data or both are reviewed in (Blöhdorn and Hotho, 2009, pp. 637-661).

To the best of our knowledge, only (d’Amato, 2014) and (Idoudi, Ettabaa, Solaiman, Hamrouni and Mnif, 2016) have dealt with the problem concerning the enrichment of ontologies by association rules. However, data and domain ontology are not linked, and these authors did not specify how the ontology was built, in order to improve the semantics of the sources. Yet, intuitively, as new knowledge-based constraints, the discovered association rules can be used for tuning and enriching the schemas of the database and of the domain ontology.

In this paper, our interest is the strengthening of ontology-based data with a set of discovered AR. Additionally; our approach contributes to reduce the search space for further mining tasks because of the availability of prior knowledge (Dou, Wang and Liu, 2015) in the Knowledge-based semantic meta-model (KBSM) (Sy, Duarte and Dal Bianco, 2018) that is the basis of our approach, and whose main feature is that it is a domain-independent abstract structure. Moreover, the KBSM can stand as a referential for ontology learning because it contains the same concepts as defined in (Cimiano, 2006) for the ontology learning layer cake. Furthermore, the new constraint-based semantics can be exploited by the standard database’s structured query language, and or an advanced programming language, depending on the database technology that is used.

The paper is organized as follows: In Sections 1 and 2 we expose the basic concepts of AR and ontologies. Section 3 presents the closed related works. Section 4 presents a brief discussion on semantics vs. constraints. In Section 5, we expose our proposal. Section 6 concludes the paper and points out the future work.

1. ASSOCIATIONS RULES

Association rules are obtained by the task of data mining of the KDD's process. Traditionally, the algorithms of extraction of association's rules were applied to the transactions of the basket data where a transaction T is described by its identifier TID and a set I of items (Agrawal and Srikant, 1994). More formally, given $I = \{i_1, i_2, \dots, i_m\}$ and $D = \{T_1, T_2, \dots, T_n\}$, with $T_i \subseteq I$, an association rule is an implication of the form $X \Rightarrow Y$ where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$ (Agrawal and Srikant, 1994).

In our case, the mining sources are databases and ontologies. Thus, a transaction is a tuple or an individual; items are attributes of the relations that describe real world, and rules are in the form $\langle A_{i1} = v_1 \wedge A_{i2} = v_2 \wedge \dots \wedge A_{ik} = v_k \Rightarrow A_{i0} \in E \rangle$, where A_{i0} is the predicted attribute and E is an interval or a set of values. As example, the rule $\langle V.CATEGORY = 'JAYA' \wedge V.PS = 'AVERAGE' \Rightarrow V.FGERM \in [90, 98] \rangle$ means that if the category of the rice is JAYA and its specific purity is AVERAGE then the value of the germinal faculty is in the interval [90, 98]. We emphasize that the AR we used were computed upon a unique relational table based on the measures of support and confidence as defined in (Agrawal and Srikant, 1994; Feno, 2007; El Ayyadi, Ouiziri, Benbernou and Younas, 2015).

2. ONTOLOGIES

The term ‘‘Ontology’’ appeared for the first time in 1967 in the computer and information science literature (Guizzardy, 2007). In the Semantic Web community, an ontology is defined as a ‘‘formal, explicit specification of a shared conceptualization’’ (Studer, Benjamins and Fensel, 1998, pp.161-197), namely an artifact designed for a specific purpose and represented in a specific language. Following (Guizzardy, 2007), we defined an ontology as ‘‘an organized structure, constrained by a set of axioms’’ (Sy, Duarte and Dal Bianco, 2018), more formally a set $O^D = \{\Sigma, \tau_{\leq}, A, \Omega, \rho, \phi\}$ where: Σ is a set of concepts $\{C_1, \dots, C_n\}$ assigned with a taxonomic partial order relation $\tau_{\leq} C_i, C_j$; A is a set of attributes for the description of the concepts in Σ ; Ω is a set of unary and binary relations; ρ a relation over Ω assigning to each $R_i \in \Omega$ a domain $pdom: R_i \rightarrow \Sigma \times \Sigma$ and a range $\rho_{range}: R_i \rightarrow \Sigma \times \Omega$; and ϕ a set of axioms that Σ and Ω must respect.

3. RELATED WORK

As emphasized above, to the best of our knowledge, the reuse of AR for ontology-based semantics enrichment have only been studied in (d’Amato, 2014) and (Idoudi, Ettabaa, Solaiman, Hamrouni and Mnif, 2016). In (d’Amato, 2014) the work focused on ontology enrichment as a pattern discovery problem by exploiting the ML methods in the aim to

extend existing ontologies with formal rules, and to suggest new knowledge axioms. Each pattern is a Horn-like clause in the form of $B \rightarrow H$ called “relational association rule”. Given, respectively, C_i and R_i the concept and role names of the ontological knowledge base, the discovered rules are in the form:

$$C_1(x) \wedge R_1(x, y) \wedge \dots \wedge C_n(z) \wedge R_l(z, a) \rightarrow R_k(y, z) \quad (1)$$

$$C_1(x) \wedge R_1(x, y) \wedge \dots \wedge C_n(z) \wedge R_l(z, a) \rightarrow C_h(y) \quad (2)$$

In our case, knowing the type of plot lands, their state, and the varieties of rice, for each variety of seed of rice an ontologist can deduce the plot land on which it is/must be sowed. Such kind of knowledge can be expressed as following:

$$\text{Plot}(x) \wedge \text{Variety}(y) \wedge \text{hasState}(x, s) \rightarrow \text{sowOn}(x, y) \quad (3)$$

The main difference between our approach, sustained by the Knowledge-Based Semantic Model (KBSM) (Sy, Duarte and Dal Bianco, 2018), and the approach proposed by d’Amato (d’Amato, 2014) resides in the manner by which data and ontologies are linked and processed. First, based on the KBSM, any domain ontology schema can be merged with its canonical database schema. Secondly, the business-data and the domain ontology are integrated such that their meta-schemas are logically linked in a single database catalog. As a result, the ontology and the database complement the semantics of each other according to the knowledge domain. As illustration, in the example above, the instances of the roles, `hasState` and `sowOn`, and these of the ontology-based concepts, `Plot` and `Variety`, can be stored in a unique database catalog in distinct schemas. Thus, having a set of association rules computed over a given database state, these rules can be applied as new semantics over the schemas. Moreover, any rule of the type (3) can be expressed through a Select-Project-Join (SPJ) query using SQL. Finally, no transformation under the form of (1) and (2) of the computed association rules is necessary. Association’ rules-based ontology enrichment is also discussed in (Idoudi, Ettabaa, Solaiman, Hamrouni and Mnif, 2016). However, unlike ours, the ontology and the database are separated, and it is not clear how their schemas are built. Moreover, based on the KBSM

(Sy, Duarte and Dal Bianco, 2018), itemsets' mapping to the ontology's concepts is not necessary. Shortly, one of the main advantages of our approach is the use of a single algorithm for building semantic-based rules.

4. CONSTRAINTS VS. SEMANTICS

When describing databases schemas, according to the considered real world, and the users' points of view, the designers add constraints e.g., the well-known functional dependencies allowing schemas normalization, primary keys and referential constraints definition. Additionally, for data consistency, business-based and knowledge-based constraints are also added to the schemas. For example, a seed breeder have a yearly agreement, the rice gardens and the research stations are plots of land for rice seed reproduction.

Yet, in (Han, Lakshmanan and Raymond, 1999), in a multidimensional data mining context, five types of constraints have been distinguished: knowledge-based constraints, data constraints, dimension constraints, rules constraints, and interestingness-based constraints. The first type of constraints is used to extract some type of knowledge. From our point of view, the three other types of constraints are business-based constraints. Thus, we distinguish the functional constraints that serve to handle schemata consistency, the business-based constraints for data and knowledge filtering and/or for the classification, and the knowledge-based constraints, i.e. the constraints that carry semantics, e.g. a triggering for a taxonomic relationship tuning. In this order, we claim that AR can be viewed and exploited as meta-rules. Indeed, each assertion belonging to a body X is a relation of the type $R_i(x, y)$ we will write, from now on, $R_i(A_i, v)$ where A_i is an attribute and v a literal ; precisely, in our case, v is a numeric. Accordingly, for each AR, we compute a conversion by performing a cross-scan between the KBSM and the set of rules. As example, given the rule $\langle V.CATEGORY = 'JAYA' \wedge V.PS = 'AVERAGE' \Rightarrow V.FGERM \in [90, 98] \rangle$, thanks to the taxonomic hierarchy, the discovered knowledge is applied to each category of rice seed no matter the type of the plot of land where it has been sowed (see rule (3)). Furthermore, the disjunction property suffices to

determine the plot of land where the seed has been produced through a SPJ query.

5. ASSOCIATION RULES AS SEMANTICS

According to Curé (Curé, 2012, p.20), an AR ($X \rightarrow Y$) expresses a functional dependency or a conditional functional dependency between the sets of attributes X and Y . Yet, both these features carry semantics for the data and the ontology schemas. Thus, we consider AR as semantics that is the cornerstone for domain ontologies. On the other hand, each domain ontology is related to the database of a given Information System, and conversely. Such database contains the minimal canonical ontology (Sy, Duarte and Dal Bianco, 2018) of the considered domain, namely a subset of N -ary relations ($N \geq 2$). Moreover, the database's repository is the mining source of association rules computing. Hence, computed rules can be used for the enrichment of the domain ontology by exploiting the semantics captured through the underlying meta-model KBSM (See Figure 1).

In this spirit, we show that rules of type (3) are conceptualized by the KBSM that adds much semantics such as multiplicities that cannot be expressed by Horn-like clauses, but useful for knowledge retrieval. Each rule $R_i (A_i, v)$ represents a role name, where A_i belongs either to a concept or to a relationship between several concepts, respectively named CConcept and Rel.

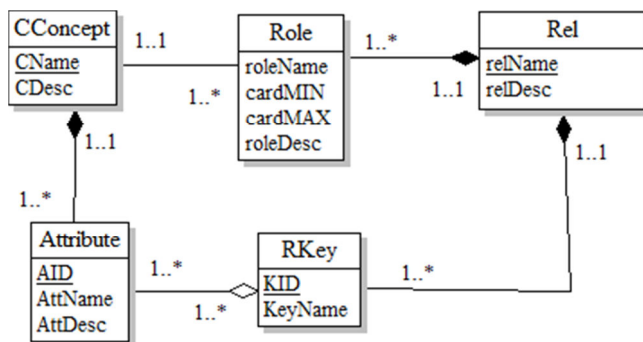


Figure 1. A sub-schema of KBSM for Semantics conceptualization

Such conceptualization takes into account the both axioms, namely the relations $R_i(x, y)$ and $R_i(A_i, v)$, which are the core components of the ontology and of the association rules, as well. Intuitively, if $A_i \in \text{Rel}$, then A_i is either a key or a part of key, hence the rule at which it belongs does not bring any new knowledge. In the other case, the concept at which it belongs is checked by using the item name as search value through the table Attribute. Then, the rule is applied for either to add a default value constraint or a values domain constraint. If the concept is involved in a role where the relationship is a taxonomy, the constraint is propagated to all the subsumed concepts. To apply the rules, tables CConcept and Rel are explored with the help of our semantic-based algorithm. To this end, our first material is the discovered AR computed by an Apriori-like algorithm launched upon a relational table of certified rice seed. These AR are stored by the mining algorithm in a formatted text file ending with the dot symbol (.), where each line contains a rule of the form $\langle A_{i1} = v_1 \wedge A_{i2} = v_2 \wedge \dots \wedge A_{ik} = v_k \Rightarrow A_{i0} \in E \rangle$ (Jamy, Jen, Laurent, Loizou and Sy, 2004, pp. 103–124). In this work, the computed AR are stored in a relational table. Our second material is the KBSM (See Figure 1), and which is the basis of our rules-driven semantic-based algorithm (RDSB), enhanced by the data model of figure 2.



Figure 2. Rules driven semantic data model

Below is depicted the RDSB algorithm.

Algorithm RDSB

Input: RHeads //Rules heads table

RBodies //Rules bodies table

Begin

For each r in RHeads do

Select * Count (rid) as nbItems From RBody Where RBodies.rid=

```

RHeads.rid
  If nbItems=1 Then
    If RBodies (1).name  $\notin$  Rel Then
      item  $\leftarrow$  RBodies (1).name
    Search (CConcept, item)
    Alter table CConcept add constraint CC check (RBodies (1).value=
RHeads (1).value)
  End if End if
  If nbItems>1 Then
    While not eof (RBodies) do
      If RBodies (1).name  $\notin$  Rel Then
        item  $\leftarrow$  RBodies (1).name
        Search (CConcept, item)
        //concept containing the item
        AddItem (item, t_cons, RBodies (1), RHeads (1))
      End if End while alter table CConcept add <t_cons> End if
    End For Each
  End.

```

Data processed by the RDSB Algorithm are illustrated at Table 1 and Table 2. For the sake of space, we consider here only five rules computed from a relational view of four columns. The RHEADS table column represents the germination faculty that is predicted, and that we abbreviate FGERM.

Table 1

RHEADS data table

Nº	RID	RHEADS
1	1	[90, 95]
2	2	[90, 93]
3	3	[90, 93]
4	4	[92, 98]
5	5	[92, 95]

Table 2

RBODIES data table

№	RID	ITEMID	ITEMNAME	ITEMVALUE
1	1	1	variety	jaya
2	2	2	category	base
3	3	1	variety	jaya
4	3	2	category	base
5	4	3	puriy	average
6	5	1	variety	jaya
7	5	3	purity	average

For each rule head in Table 1, the data of Table 2 is processed for building a semantic check constraint. For example, from rule 5, we built the semantic check constraint: CHECK (variety = 'jaya' AND purity = 'average' AND fgerm > 92 AND fgerm < 95).

Finally, the proposed approach can easily help to manage data and knowledge, in the context of a decision-making environment. Thereby, let us consider the problem of the missing-values handling, assuming that the completion by imputation and the mean-value are retained. Thus, in the absence of known value, the semantic check constraint implemented thanks to rule 5 above allows us to fix the mean value of the predicted range [92, 95] as default-value for the germination purity. Indeed, the lack of abstract representation of the knowledge requires the collaboration of the experts of the domain for the construction of the ontology from the existing databases or conversely. Additionally, the data actually stored in the databases are not all processed, and not all of the attributes are used in the queries, as well.

CONCLUSION

In this paper, we highlighted the relevance of the reuse of a set of already computed association rules for the enrichment of the underlying domain ontology, and of the database, as well. Actually, more than a simple taxonomy, the ontology includes a set of constraints semantic-based on the concepts of the domain. Thus, the main interest of this contribution is the

proposed RDSB Algorithm through which we described the feasibility of schemas tuning upon an expressive conceptualization for associations rules reusability as semantics. Finally, we emphasize that the proposed generic meta-schema as common referential, we called semantics conceptualization and the RDSB Algorithm, constitute the theoretical foundations of this work, and as well, for ontology learning because it contains the same concepts as those defined for the well-known ontology learning layer cake.

The next step is to demonstrate the applicability of the algorithm by its implementation in the aim to show experimental results after the use of the set of rules over the data augmented with semantic abstractions in the underlying database catalog.

REFERENCES

1. BARBARA FURLETTI (2009). *Ontology-Driven Knowledge Discovery*, PhD Program in Computer Science and Engineering XXI Cycle, IMT Institute for Advanced Studies, Lucca, Italy.
2. BUITELAAR, P., CIMIANO, P., and MAGNINI, B. (2005): *Ontology learning from text: An overview*. *Ontology learning from text: Methods, evaluation and applications*. *Frontiers in Artificial Intelligence and Applications Series* 123.
3. C. TATSIPOULOS and B. BOUTSINAS (2009). *Ontology Mapping Based on Association Rule Mining*, *Iceis 2009 - International Conference on Enterprise Information Systems*.
4. CLAUDIA ANTUNES (2007). *Onto4AR: a framework for mining association rules*, *International Workshop on Constraint-Based Mining and Learning at ECML/PKDD 2007 CMILE'07*, September 21, 2007, Warsaw, Poland, pp. 43-54.
5. CLAUDIA D'AMATO (2014). *Machine Learning for Ontology Mining: Perspectives and Issues*, *OWLED 2014*, Riva del Garda, October 18, 2014.
6. CLAUDIA MARINICA and FABRICE GUILLET (2009). *Improving Post-Mining of Association Rules with Ontologies*, *The XIII International Conference "Applied Stochastic Models and Data Analysis" (ASMDA-2009)* June 30-July 3, 2009, Vilnius, LITHUANIA, ISBN 978-9955-28-463-5 L. Sakalauskas, C. Skiadas and E. K. Zavadskas (Eds.): *ASMDA-2009 Selected papers*. Vilnius, 2009, pp. 76-80 © Institute of Mathematics and Informatics, 2009 © Vilnius Gediminas Technical University.

7. CLAUDIA MARINICA and FABRICE GUILLET (2010). Knowledge-Based Interactive Postmining of Association Rules Using Ontologies, *Ieee Transactions on Knowledge and Data Engineering*, vol. 22, NO. 6.
8. D. SARAVANAM, S. VIJAYALAKSHMI, DENNIS JOSEPH (2017). Finding of Interesting Rules from Association Mining by Ontology and Page Ranking, *Proceedings of National Conference on Recent Advances in Computer Science & Engineering (NCRACSE-2017)* - International Journal for Modern Trends in Science and Technology Volume: 03, Special Issue No: 01, ISSN: 2455-3778, <http://www.ijmtst.com>
9. D. WILLIAM ALBERT, K. FAYAZ, D. VEERABHADRA BABU (2014). A Survey of Post Mining of Association Rules and High-Speed Association Rule Mining, *International Journal of Computer Engineering and Applications*, Volume V, Issue II.
10. DANIEL RAJAONASY FENO (2007). Mesures de qualité des règles d'association : normalisation et caractérisation des bases. *Mathématiques [math]*. Université de la Réunion, 2007. Français. <tel-00462506>
11. DEJING DOU, HAO WANG, HAISHAN LIU (2015). Semantic Data Mining: A Survey of Ontology based Approaches, February 2015 DOI:10.1109/ICOSC.2015.7050814
12. FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMYTH, P. and UTHURUSAMY, R. (1996). *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press.
13. GUIZZARDI, G (2007). On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta)Models, *Databases and Information Systems IV*, O. Vasilecas et al. (Eds.), IOS Press, 2007.
14. HANEN BRAHMI, NADIA ARAOUR, SADOK BEN YAHIA (2012). La fouille intelligente des règles d'association à partir des données décisionnelles, EGC-31, Atelier AiDE.
15. ISAM EL AYYADI, MOURAD OUIZIRI, SALIMA BENBERNOU, MUHAMMAD YOUNAS (2015). La découverte des règles d'association dans un contexte distribué avec des données manquantes : Décomposition tensorielle. EGC-2015, Jan 2015, Luxembourg, Luxembourg. <hal-01116734>.

16. JIAWEI HAN, LAKS V.S. LAKSHMANAN, RAYMOND T. NG (1999). Constraint-Based, Multidimensional Data Mining, Computer 0018-9162/99/\$10.00 © 1999 IEEE.
17. LINA ZHOU (2007). Ontology learning: state of the art and open issues, Inf Technol Manage (2007) 8:241-252, DOI 10.1007/s10799-007-0019-5, Springer Science+Business Media, LLC.
18. LUIS GALÁRRAGA (2016). Rule Mining in Knowledge Bases, Thèse, Spécialité "Informatique ", TELECOM ParisTech, 2016.
19. LUIS GALÁRRAGA, CHRISTINA TEFLIOUDI, KATJA HOSE, FABIAN M. SUCHANEK (2013). AMIE: Association Rule Mining under Incomplete Evidence in Ontological Knowledge Bases, WWW, '13 Rio de Janeiro, Brazil, ACM 978-1-4503-2035-1/13/05.
20. MARYAM HAZMAN, SAMHAA R. EL-BELTAGY, AHMED RAFEAA (2011). A Survey of Ontology Learning Approaches, International Journal of Computer Applications (0975 - 8887), Volume 22- No.9.
21. MÄS, STEPHAN; WANG, FEI; REINHARDT, WOLFGANG (2005). Using Ontologies for Integrity Constraint Definition, Proceedings of the 4th International Symposium On Spatial Data Quality'05.
22. N. RADHIKA, K. VIDYA (2012). Association Rule Mining based on Ontological Relational Weights, in International Journal of Scientific and Research Publications, Volume 2, Issue 1, ISSN: 2210-3153
23. NOMAN ISLAM, MUHAMMAD SHAHAB SIDDIQUI, ZUBAIR A. SHAIKH (2010). TODE: A Dot Net Based Tool for Ontology Development and Editing, DOI: 10.1109/ICCET.2010.5486292 · Source: IEEE eXplore.
24. OLIVIER CURÉ (2012). Improving the Data Quality of Drug Databases using Conditional Dependencies and Ontologies. Journal of Data and Information Quality, 2012, 4 (1), pp.20.
25. OUMAR SY, DENIO DUARTE, GUILHERME DAL BIANCO (2018). Ontologies and Relational Databases Meta-Schema Design: A Three-Level Unified Lifecycle, 5th International Conference on Control, Decision and Information Technologies, CoDIT'2018 10-13 April, Thessaloniki - Greece.

26. P SIVA BHAVANI, M N HIMA BINDU, K. SYAMA SUNDARARAO (2015). Ontology Based Post Mining of Association Rules, International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 7.
27. P. CIMIANO. ONTOLOGY LEARNING and Population from Text (2006): Algorithms, Evaluation and Applications. Springer, 2006.
28. R. HUBERT RAJAN, JULIA PUNITHA MALAR DHAS (2012). A Method for Classification based on Association Rules using Ontology in Web Data, in International Journal of Computer Applications (0975 - 8887), Volume 49- No.8.
29. RAKESH AGRAWAL, RAMAKRISHNAN SRIKANT. Fast Algorithms for Mining Association Rule, Proceedings of the 20th VLDB Conference, Santiago, Chile, 1994.
30. RIHAB IDOUDI, KARIM SAHEB ETTABAA, BASEL SOLAIMAN, KAMEL HAMROUNI, NAJLA MNIF. Association rules-based Ontology Enrichment, in International Journal Web Applications Volume 8 Number 1 March 2016.
31. RUDI STUDER, V. RICHARD BENJAMINS, DIETER FENSEL (1998). Knowledge Engineering: Principles and methods, Data & Knowledge Engineering 25 (1998) 161-197, © 1998 Elsevier.
32. S. DANA K. TOMIC, DOMAGOJ DRENJANAC, AND GORAN LAZENDIC, SANDRA HÖRMANN, FRANZ HANDLER, WILFRIED WÖBER, KLEMENS SCHULMEISTER, MARCEL OTTE, WOLFGANG AUER (2014). AgriOpenLink: Semantic Services for Adaptive Processes in Livestock Farming, International Conference of Agricultural Engineering, Zurich 6-10 July, Ref: CO274.
33. STEPHAN BLÖHDORN AND ANDREAS HOTH (2009). Ontologies for Machine Learning, in Handbook on Ontologies, Steffen Staab Rudi Studer (Eds.) pp. 637-661, 2nd Edition, ISBN 978-3-540-70999-2, Springer, DOI 10.1007/978-3-540-92673-3.
34. SYLVIE IRÈNE JAMY, TAO-YUAN JEN, DOMINIQUE LAURENT, GEORGES LOIZOU, OUMAR SY (2004). Extraction de règles d'association pour la prédiction de valeurs manquantes, Revue Africaine de la Recherche en Informatique et Mathématique Appliquée ARIMA, vol. Spécial CARI04, pp. 103-124.

SOCIAL MEDIA POSSIBILITIES FOR IMPROVING ARTS MARKETING PERFORMANCE

Ina Stanoeva¹

¹ *Varna University of Management/School of Business and Marketing,
Varna, Bulgaria, ina12_bg@yahoo.com, ina.stanoeva@vumk.eu*

Abstract

The aim of this paper is to investigate the social media presence of the performing arts institutions and the freelance artists from the sector in Bulgaria.

Their presence on Facebook, Instagram, Youtube and Twitter, the state of their websites, the existence of ticket buying systems and usage of chat bots are compared. The methodology used in this article combines secondary research, content analysis and comparative analysis. Some recommendations for managerial implication are given at the end of the paper.

Keywords: *social media, advertisement, performing arts, content analysis, presence.*

INTRODUCTION

More and more the social media are becoming a tool for informing the audience about different events in the past few years. Social media are used broadly by digital marketing experts to inform, advertise and engage audiences. In traditional marketing communications theory exists the concept that they include advertising, PR, direct marketing, promotions and personal selling. Social media can combine all these activities and can be used in the marketing strategy and tactics.

Performing arts are traditionally more related to offline and live connection with their customers which are the audience of the performing arts product. A few years ago many of the Bulgarian performing arts institutions still counted on traditional marketing tools for attracting the audience such as leaflets and brochures, posters, word of mouth and the genuine interest of the public (Stanoeva, 2013). They didn't consider that other types of entertainment also fight for the attention of the customers.

However, the institutions and the freelance performance realise more and more that online connection with the public is also very important.

The aim of this paper is to investigate the social media presence of the performing arts institutions and the freelance artists from the sector.

The methods of content analysis, secondary research and comparative analysis have been used to investigate that.

1. LITERATURE REVIEW

Performing arts

Performing arts include theatres, music and opera, dance and circus. There are 66 theatres, 9 opera and ballet institutions in Bulgaria according to Bulgarian national statistical institute (NSI, 2019). There is no such information regarding the circuses. The information is presented in table 1.

Table 1

Performing arts institutions in Bulgaria in 2019

№	Institution	Number	Audience
1	Drama theatres	36	1 282 000
2	Puppet theatres	23	436 000
3	Drama-puppet theatres	7	268 000
4	Opera and ballet	7	313 000
5	Operettas	2	91 000
6	Philharmonies, ensembles, orchestras	51	780 847
	Total	126	

Source: National statistic institute of Bulgaria (2018)

The NSI doesn't present information about the types of the performing arts institutions in terms of their status – state owned or private.

It was searched for information on the website of the Bulgarian Ministry of Arts. There is information about the state and municipality owned performing arts institutions. Some interesting issues have been found. The information from NSI about the total number of performing arts institutions is presented in the table 2, in the left column and the information from the Ministry of Arts is presented on the right column. It

can be seen that at some rows there is a difference in the information. State drama-puppet theatres are more than the total number presented by the NSI. The same situation is related to opera and ballet institutions. There is no information about the operettas.

This may be due to different methods of classification of the institutions or it is a mistake. This should be cleared for future researches on the topic.

Table 2

State and municipality owned Performing arts institutions in Bulgaria in 2019

№	Institution	Total Number (according to NSI)	State and municipality owned (according to Ministry of arts)
1	Drama theatres	36	16
2	Puppet theatres	23	11
3	Drama-puppet theatres	7	8
4	Opera and ballet	7	10
5	Operettas	2	N/A
6	Phil harmonies, ansambles, orchestras	51	6
	Total	126	51

Source: National statistical institution of Bulgaria (2018); Ministry of Arts of Bulgaria (2018)

According to publications in the press the circus art is now private based and in Bulgaria exist 9 circuses (Kitanova, 2017; Vaseva, 2017).

Social media and social networks

Elefant (2011) describes the term social media as a “technology that helps interactive information, user-created content and collaboration”. There are different classifications of social media. Kaplan and Haenlein (2010) classify social media such as: collaborative projects (e.g. Wikipedia), blogs and microblogs (e.g. Twitter), content communities (e.g. Youtube), social networking sites (e.g. Facebook), virtual game worlds

(e.g. World of Warcraft) and virtual social worlds (e.g. Second Life). Its technologies include: blogs, picture-sharing, vlogs, wall-postings, email, instant messaging, music-sharing, crowd sourcing and voice over IP, to name a few.” According to Dencheva (2016) social networks differ from social media. Social networks are part of social media and they present Internet based platforms which gather different groups: people and companies in common virtual reality”.

Social media marketing includes digital marketing tools using “digital technologies (e.g. internet, smart phones, digital promotions, etc.)” (Thomas & Housden, 2011; Johnsen, 2017; Rowan, 2002; Wind & Mahajan, 2002). It combines with traditional marketing and traditional marketing communication tools, but it may have some new features such as content marketing, SEO, Pay-per-click, affiliation marketing, website, email marketing and blogs.

2. METHODOLOGY

The methodology used in this article combines secondary research, content analysis and comparative analysis. The websites, Facebook, Instagram, Youtube and Twitter profiles of the Bulgarian state performing arts institutions have been searched. To find them, it was used the complex assessment for the performing arts institutions made by Bulgarian Ministry of arts (2019). The research was conducted in the period 5-9.08.2019. While doing the research it was found out that some of the institutions had chat bots or their own ticket and reservation system. The results are presented in the next section.

Limitations

The research includes information about the state performing arts institution in Bulgaria excluding circuses since they are private owned nowadays. It is narrowed only to Facebook (since it is the most used social network platform in 2019 (similarweb.com, 2019; napoleon-cat.com, 2019) and Instagram, Twitter and YouTube. Information about some of the most popular Bulgarian artists in the field which have their online presence in the social media is also included.

3. RESULTS AND DISCUSSION

The results from our research are presented in the table 3 below.

Table 3

Social media presence of the Performing arts institutions in Bulgaria in 2019

№	Institution	Website	Facebook page	Twitter	YouTube	Instagram	Online ticket system	Chatbot	Points from Ministry assessment
1	National theatre-Sofia	Yes	52962	330	396	2136	Own	-	121
2	Drama theatre - Plovdiv	Yes	27856	-	202	-	Own	-	126
3	Comedy theatre-Sofia	Yes	10706	-	-	-	Own	-	117
4	Bulgarian Army theatre	Yes	16641	87	168	815	Own	Yes	117
5	Drama theatre - Rousse	Yes	3114	-	-	-	D.W.	-	117
6	Drama theatre-Burgas	Yes	8923	-	-	-	-	-	113
7	Drama theatre-Yambol	Yes	3135	-	-	-	-	Yes	107
8	Drama theatre-Gabrovo	Yes	2953	C.O.	C.O.	517	-	Yes	106
9	Drama theatre – Stara Zagora	Yes	2383	-	-	-	-	-	106
10	Drama theatre - Targovishte	Yes	2509	-	-	-	-	-	102
11	Sfumato	Yes	10915	-	-	-	Theatre.art. bg	-	102
12	Drama theatre-Dobrich	Yes	1308	27	1	-	Theatre .bg	-	95
13	DT Sliven	Yes	916	-	-	-	-	Yes	93
14	DT Blagoevgrad	I.D.	4554	-	-	-	-	-	58
15	DT-Lovech	-	1805	-	-	341	-	-	54
16	DT Smolyan	Yes	4045	-	-	-	-	-	52
17	Drama-puppet theatre 199-Sofia	Yes	9225	1175	59	792	Eventim.bg	Yes	127

18	DPT Pazardzhik	Yes	3935	-	-	-	main- tainan ce	-	115
19	DPT Haskovo	Yes	4831	-	-	596	-	Yes	109
20	DPT Shumen	Yes	2255	-	-	-	-	Yes	104
21	Youth theatre-Sofia	Yes	12123	227	36	756	Yes	Yes	104
22	DPT Pleven	Yes	6258	-	4	550	-	Yes	104
23	DPT Vratza	Yes	5243	-	-	-	-	Yes	76
24	DPT Silistra	Yes	864	-	-	-	-	-	64
25	Puppet theatre - Sliven	Yes	2409 friend profile	D.W.	D.W.	-	-	-	132
26	PT Burgas	Yes	2468	-	-	-	-	-	122
27	PT Plovdiv	Yes	2892	187	36	835	-	-	120
28	PT Yambol	Yes	-	-	-	-	-	-	116
29	PT Varna	Yes	3402	D.W.	-	D.W.	-	-	114
30	PT Rousse	Yes	5390	-	-	-	-	-	112
31	PT Stara Zagora	Yes	3590	-	-	-	Thea- tre.bg	Yes	109
32	PT Vidin	Yes	1024	-	-	-	-	-	108
33	PT Dobrich	Yes	-	-	-	-	-	-	105
34	PT Gabrovo	Yes	1154	-	-	-	-	-	99
35	PT Targovishte	Yes	1006	-	-	-	-	-	99
36	Theatre-music cen- tre Razgrad	Yes	1837	-	-	-	-	-	80
37	TMC Kardzhali	Yes	256 unnof- ficial page	-	-	-	-	-	76
38	Music drama thea- tre-Veliko Tarnovo	Yes	131	-	-	-	-	-	76
39	Sofia opera and ballet	Yes	7330	-	1349	D.W.	D.W.	-	121
40	Music and ballet centre Sofia	Yes	3299	-	35	-	Own	-	117
41	TMPC Varna	Yes	1763	-	194	-	Bgbile ti.com	-	115
42	Opera Plovdiv	Yes	675+1 006 ballet page	-	-	-	-	-	111
43	Opera Burgas	Yes	6094	D.W.	1	-	-	-	108
44	Opera Rousse	Yes	2027	-	D.W.	-	-	Yes	69
45	Opera Stara Zagora	Yes	5433	-	224	-	Own	-	69
46	Sofia philharmony	Yes	18651	74	152	1200	-	-	99
47	Symphoniette Vidin	Yes	972	-	-	-	-	-	99

48	Symphoniette Shoumen	Yes	-	-	-	-	-	-	87
49	Pleven Phylharmony	Yes	2198	-	-	-	-	Yes	62
50	Ensemble Philip Koutev	Yes	-	-	-	-	-	-	55
51	Symphoniette Vratza	Yes	948	-	-	-	-	-	35

Source: Facebook, Instagram, Twitter and YouTube profiles of the institutions and their websites; *Complex assessment of the performing arts institutions – Ministry of Arts - Bulgaria (2019)*

D.W. – Doesn't work

I.D. – in development

C.O. – can't be opened

As it can be seen 49 of 51 institution have their own sites. The presence in Facebook is also good – 47 of them have a page. This is very good since Facebook is the most popular media in Bulgaria and it is a good tool to reach and attract young audiences (Vasileva, 2017). Only 7 institutions have Twitter profiles, 14 have Youtube presence and 10 have Instagram profiles. We can see that these 3 platforms also don't have so many followers. For example, Opera Burgas and Drama theatre Dobrich have only 1 follower in YouTube. Despite that fact, we think that there are possibilities for improving the presence in these platforms since younger generations use them more.

7 institutions have their own reservation and ticket buying system. And 5 use external platform for tickets such as eventim.bg, theatre.bg, theatre.art.bg, bgbileti.com and it should be noted that these institutions are predominantly from Sofia.

It is impressive that some of them have chatbots on their Facebook page since this is relatively new and unpopular in Bulgaria (13).

Finally, the top 20 of performing arts artists in Facebook are presented in table 4.

Table 4

Top 20 Bulgarian performers in Facebook

No	Name	Total fans	Profession
1	Nina Dobrev	5 923 608	Actress inUSA
2	Preslava	690 145	Pop Folk singer
3	Galena	408 148	Pop Folk singer
4	Azis	391 157	Pop Folk singer
5	Andrea	388 630	Pop Folk singer
6	Desislava	387 449	Pop Folk singer
7	Maria Ilieva	331 976	Pop singer
8	Kamelia	324 406	Pop Folk singer
9	Tsvetelina Yaneva	276 392	Pop Folk singer
10	Emanuela	264 894	Pop Folk singer
11	Krisko	245 556	Rap singer
12	Emilia	237 386	Pop Folk singer
13	Vasko Vassilev	231 730	Musician
14	DJ Kink	183000	DJ
15	Malina	178 302	Pop Folk singer
16	Konstantin	163 449	Pop Folk singer
17	Elena Parisheva	160 474	Pop Folk singer
18	Anelia	148 215	Pop Folk singer
19	Toni Dacheva	136 897	Pop Folk singer
20	Galin	122 978	Pop Folk singer

Source: Combined classification of the Bulgarian performers using data from <https://www.socialbakers.com/statistics/facebook/pages/total/bulgaria/celebrities/>

As it can be seen 15 of the performers from top 20 in Facebook are pop folk singers. Nina Dobrev is on the first place, but she works in USA and she is worldwide known, that is why she has so impressive number of followers. In the classification place have also 1 pop singer, 1 musician, 1 DJ and 1 rap singer.

CONCLUSION

The social media presence of Bulgarian performing arts institutions and performers in the field was presented in this paper. The institutions are well presented with websites and Facebook, but they need more improvement in terms of other social platforms such as Instagram, Twitter,

YouTube. Some of them also have own system for ticket selling and reservations and chatbots. This gives them many possibilities for improvement in terms of their social media presence and this way they can improve their marketing activities and audience engagement.

Some recommendations for future research can be given. Bigger analysis could be performed adding other issues such as e-mail marketing and affiliation marketing. More performers from the different art genres could be added. Detailed and deeper statistical analysis could also be performed.

REFERENCES

1. CELEBRITIES in Facebook, statistics for Bulgaria (2019) [Online] Available at: <https://www.socialbakers.com/statistics/facebook/pages/total/bulgaria/celebrities/> [Accessed on 8/08/2019]
2. DENCHEVA, S. (2016) *Virtual communities in tourism*. Varna University of Management
3. ELEFANT, C. (2011). The "power" of social media: Legal issues & best practices for utilities. *Energy Law Journal*, 32, (1)
4. JOHNSEN, M. (2017). *Multilingual Digital Marketing: Managing for Excellence in Online Marketing*. Maria Johnsen.
5. KAPLAN, A. AND HAENLEIN, M. (2010). "Users of the world, unite! The challenges and opportunities of Social Media". *Business Horizons* 53 (1)
6. KITANOVA, D. (2017) *Bulgarian circus art is not presented in the world classifications* [Online] Available at: <http://focus-radio.net/%D0%B1%D1%8A%D0%BB%D0%B3%D0%B0%D1%80%D1%81%D0%BA%D0%BE%D1%82%D0%BE-%D1%86%D0%B8%D1%80%D0%BA%D0%BE%D0%B2%D0%BE-%D0%B8%D0%B7%D0%BA%D1%83%D1%81%D1%82%D0%B2%D0%BE-%D0%BD%D0%B5-%D0%BF%D1%80%D0%B8%D1%81/> [Accessed on 9/08/2019]
7. MINISTRY OF ARTS OF BULGARIA (2019) Complex assessment of Bulgarian performing arts institutions [Online] Available at: <http://mc.government.bg/page.php?p=58&s=81&sp=73&t=251&z=0> [Accessed on 5/08/2019]

8. MUSIC institutions in 2018 – statistics (2019) [Online] Available at: <https://nsi.bg/bg/content/3659/музикални-колективи> [Accessed on 9/08/2019]
9. ROWAN, W. (2002). *Digital Marketing: Using New Technologies to Get Closer to Your Customers*. London: Kogan Page.
10. SOCIAL media statistics for Bulgaria –(2019) [Online] Available at: <https://www.socialbakers.com/resources/reports/bulgaria/2019/june/> [Accessed on 8/08/2019]
11. STATISTICS for Facebook users in Bulgaria (2019) [Online] Available at: <https://napoleoncat.com/stats/facebook-users-in-bulgaria/2018/09> [Accessed on 5/08/2019]
12. STATISTICS for social media usage in Bulgaria (2019) [Online] Available at: <https://www.similarweb.com/> [Accessed on 5/08/2019]
13. THEATRES in 2018 – statistics (2019) [Online] Available at: <https://nsi.bg/bg/content/3657/%D1%82%D0%B5%D0%B0%D1%82%D1%80%D0%B8> [Accessed on 9/08/2019]
14. THOMAS, B. AND HOUSDEN, M. (2011). *Direct and Digital Marketing in Practice* (2nd ed.). London: A&C Black.
15. VASEVA, TS. (2017) 120 years – Bulgarian circus arts continues [Online] Available at: <https://truestory.bg/120-години-българско-цирково-изкуство-п/> [Accessed on 9/08/2019]
16. VASILEVA, B. (2017) Consumer Activities and Reactions to Social Network Marketing. *Management* 12(2). Pp. 133-144
17. WEB pages and social media profiles of the Bulgarian performing arts institutions
18. WIND, Y. AND MAHAJAN, V. (2002). *Digital Marketing: Global Strategies from the World's Leading Experts*. Hoboken: John Wiley & Sons.

PERSPECTIVES ON MOBILE DEVICES ADOPTION IN HEALTHCARE SECTOR

Iskren Tairov¹, Vladislav Vasilev²

¹ *D.A. Tsenov Academy of Economics/Department of Business Informatics, Svishtov, Bulgaria, i.tairov@uni-svishtov.bg*

² *D.A. Tsenov Academy of Economics/Department of Business Informatics, Svishtov, Bulgaria, badzealot1107@gmail.com*

Abstract

Despite the many recognized benefits and potential that mobile devices provide to the healthcare sector, their widespread use is still lagging behind. The fundamental problems related to information systems can be identified as a key factor in this - the conservative nature of the healthcare system, contradictory incentives and inappropriate control measures. In this situation, it is only possible to generate a limited vision of the potential success of mobile technologies, using approval frameworks supplemented by a strategic view of the development of the systems.

The paper focuses on mobile devices and the opportunities for their application in the healthcare sector.

Keywords: *mobile devices, healthcare, adoption model.*

INTRODUCTION

Mobile devices offer enormous opportunities to improve service levels, productivity, and cost savings in the healthcare sector. Mobile applications can be used for accessing and updating patient records, introducing review results, monitoring patients, and providing clinical information to support doctors and other healthcare workers. A number of trends indicate that mobile technology adoption needs to increase at a high pace, driven by the increased use of personal mobile devices, which is caused by reduced costs and considered as a more reliable and usable technology and telecommunication infrastructure. Although this potential is recognized, massive use of mobile devices and applications has not occurred in leading health systems.

In this paper, we look at the role mobile devices can play in the healthcare sector, explore their potential, the barriers to their widespread adoption, and offer models for their deployment.

1. POSSIBILITIES FOR APPLICATION OF MOBILE DEVICES IN THE HEALTHCARE SECTOR

The development of mobile technologies provides new opportunities for implementing and fulfilling the duties of health workers.

Authors like Hameed (Hameed, 2003) draw attention to the evolution of mobile wireless telecommunications, which is directly proportional to the decreasing maintenance costs. With the development of global telecommunication infrastructure, the importance of providing and using information is growing. Taking into account the then UK National Health Service's strategy to re-introduce better patient care technologies (Hameed, 2003), notes the following mobile-related activities:

- Commitments – maintenance of patient services, removal of the obligation to maintain written records;
- Achievements – Providing access and sharing of information to the necessary medical establishments, complying with preliminary requirements, and allowing October 2000 to maintain an electronic version of a patient file;
- Benefits – maintenance and completion of clinical history, provision of electronic services to hospitals.

Although the author's research was done at a much earlier period, his findings can be applied to the already existing strategy for the development of healthcare in Bulgaria 2020, and in particular to the application of mobile and wireless technologies to the collection, keeping and providing important medical information.

Also, considering the potential of cloud technologies (Sultan, 2014) with their ability to provide computing resources and store data without being personally administered by the user, they can be used as the basis for mobile technology deployment. Concepts of platform as a service and software as a service provide a great deal of processing of the required medical information. As an option, this provides more mobility, decentralized information infrastructure and less financial support. On the other hand, this raises a number of challenges, especially in security. Building a fully internal information infrastructure is a costly undertaking, but it provides more opportunities for control and, if necessary, for further development.

Based on this, we can make a proposal for the use of mobile devices in healthcare facilities on two principles:

- A system that has part of its infrastructure and activities is exported to the cloud;
- A system based entirely on the treatment facility and all adjacent compartments.

2. CHALLENGES RELATED TO THE USE OF MOBILE DEVICES IN HEALTHCARE FACILITIES

Along with the benefits of mobile devices, there are many problems that accompany them. Noticing this Varbanov (Varbanov, 2014), draws attention to the risks to information safety. If access is granted to new devices, this means that there are still more devices to be managed by the given IT departments. Both desktops and mobile devices are exposed to the same risks:

- Harmful programs;
- Data leakage due to violation of access control and security policy;
- Illegal traffic from or to the Internet;
- Violation of rules for using them.

At the same time, there are other types of obstacles related to legal rules and regulations. Considering European Union legislation and the General Data Protection Regulation (GDPR) (Anon., n.d.) . This is an additional measure introduced to protect personal data. This also has a great burden especially for healthcare establishments because they handle extremely important information. They also need to create their own rules for dealing with classified information.

From a technical point of view, the deployment of mobile technologies in the healthcare sector requires tackling a number of challenges that generally include (Mehta, 2019):

- Complexity of system integration - Developers should strive to design applications that are compatible with existing systems and do not require much intervention in them;
- Interoperability issues - the ability of different medical systems to communicate with each other and to seamlessly exchange patient health information in a protected environment;

- Risks associated with data;
 - Creating firewalls and activating data encryption;
 - Provide protection for healthcare and healthcare systems in line with HIPAA guidelines;
 - Forming a strong BYOD frame;
 - Creating a mobile-safe storage and retrieval system;
 - Creating a comprehensive employee training plan to raise awareness and follow-up on best practices;
- Failure to meet patient commitments - Mobile solutions must be developed with the key aim of being able to provide patients with full information on their health and treatment;
- Unsatisfactory user experience - the application must have an intuitive interface and easy navigation. Special attention should be paid to user experience and patient education on how to use the applications;
- Personalization needs - personalized cloud environments must be selected while developing an organization's mobility strategy. The cloud-based mobile system will not only provide seamless mobility but also reduce the complexity of integrating mobility with the existing IT environment;
- Unreliable solutions - a thorough technical assessment is required to ensure that the solution is reliable.

3. MODELS OF MOBILE DEVICES IMPLEMENTATION IN THE HEALTHCARE SECTOR

The implementation of mobile devices in organizations is done according to the specifics of the processes in the respective information system and the working environment, and departments are directed to the Bring Your Own Device – BYOD, Choose Your Own Device – CYOD and other concepts. Several models and approaches based on the Technology acceptance model - TAM (Riad, 2013) apply to the adoption of mobile technologies in the healthcare sector. The model seeks to explain consciously the planned behavior of a wide range of end user technologies and user groups. According to TAM, the beliefs of innovation and the resulting attitude lead to the individual's decision to use or not to use innovation. Beliefs about perceived utility and perceived ease of use are

considered to be the most important constructs for predicting technological acceptance in TAM. The attitude and intent to use technology are the variables that help to conceptualize and explain causality relations.

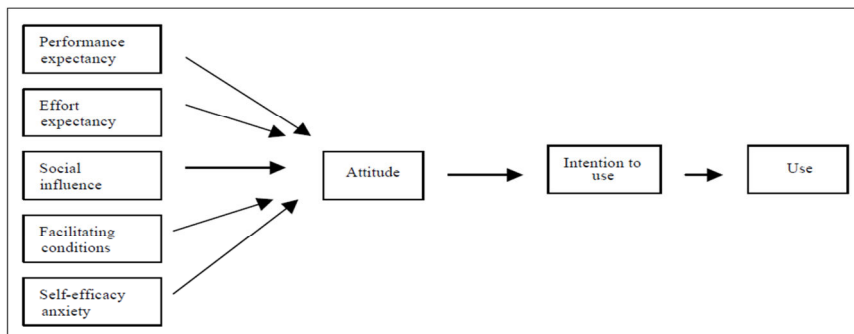


Figure 1. Conceptual adoption model (Riad, 2013)

While technology adoption can help developers and managers consider the key success factors of acceptance, they focus on the user and do not take a strategic perspective.

CONCLUSION

The adoption of mobile devices in the healthcare sector is a complex and multifaceted process that requires serious consideration and selection of appropriate tools. With the use of mobile reception frameworks, some aspects of the problems associated with the complexity of receiving these devices in the healthcare sector can be explained. However, systems are often described as complex structures of multiple components and defends the view that systemic problems can not be solved by solving the problems of a separate system component. In order to make sustainable improvements, executives and senior managers need to identify deeply troubled system problems and explore the potential of technology adoption from a systems perspective and also learn about the capabilities of the TAM model. In other words, a strategic perspective is needed that encompasses systemic thinking.

REFERENCES

1. ANON., n.d. *General Data Protection Regulation*. [Online] Available at: <https://gdpr-info.eu/>, [Accessed 15 07 2019].
2. HAMEED, K., 2003. The application of mobile computing and technology to health care services. *Telematics and Informatics* 20 (2003) 99–106.
3. MEHTA, S., 2019. *Key Challenges in Mobile Healthcare Implementation*, Edison: s.n.
4. RIAD, M. J. M. Z. S., 2013. Applying the technology acceptance model to the introduction of mobile healthcare information systems. *International Journal of Behavioural and Healthcare Research*, 4(2).
5. SULTAN, N., 2014. Making use of cloud computing for healthcare provision: Opportunities and challenges. *International Journal of Information Management*, april, 34(2), pp. 177-184.
6. VARBANOV, R., 2014. Applications of the Byod Conception – Benefits, Risks and Approaches. *Business management*, Issue 2.

DEEP NEURAL NETWORKS FOR DETECTION OF CREDIT CARD FRAUD

Angelin Lalev¹

*¹ D.A. Tsenov Academy of Economics, Svishtov, Bulgaria,
lalev@uni-svishtov.bg*

Abstract

The purpose of this paper is to present preliminary results from ongoing study, concerning application of Deep Neural Networks (DNN) to the detection of credit card fraud. The main approach is testing the change of different architectural parameters of the neural network. The influence of the change on the performance and convergence of the network is measured. Some of the chosen parameters are the depth of the network, the width of each layer and the chosen activation functions in each layer. The main results of this study indicate that deep neural networks with moderate number of hidden layers in combination with techniques of oversampling are performing markedly better in detecting fraudulent transactions than networks with only one hidden layer. The used sample dataset is imbalanced, which means that the DNN training on the dataset has tendency to overfit. The results of this study are useful for other researchers and practitioners who try to analyze real datasets in order to detect card frauds.

Keywords: *credit card fraud, deep neural networks, DNN, imbalanced dataset.*

MOTIVATION

Deep Neural Networks (DNNs) are rapidly evolving towards new architectures and solutions. These solutions are driven by availability of new, faster hardware and the vast volumes of data, available for analysis. The penetration of these new solutions is rather uneven and favors exactly the fields where the data is most abundant. Such fields are for example image recognition and natural language processing. Some of the architectures for building DNNs, developed for problems in these fields, are yet to be tested on problems in the area of business and economics. The goal of our study is to evaluate some of these architectures on inherently business-related problem – detection of credit card fraud.

The practical importance of detecting fraudulent transactions is beyond question. The steady increase of online transactions in the last

years has driven up the prevalence of Card-Not-Present fraud and has forced retailers, banks and credit card companies around the world to seek all possible solutions to combat this alarming trend (Parusheva & Atanasova, 2016). Data science is also important part of this effort and is relied upon to develop analytical techniques and models, which help identify as many as possible of these transactions without interfering with the activities of the legit credit card owners.

Due the specifics and sensitivity of the data being processed, development of working classifier models that separate legit from fraudulent transactions mostly happens behind closed doors. In these conditions it's very hard to apply the usual process of scientific discourse and to arrive at certain conclusions about the efficiency of different methods and approaches. Fortunately, in the recent years some datasets about credit card fraud have become publicly available. This fact makes our stated goal much more attainable. Even though these datasets are heavily anonymized, they contain enough information about real transactions to allow comparison of the efficiency of DNNs against more traditional techniques and methods for binary classification.

1. DATASET AND INITIAL ASSUMPTIONS

Our study uses publicly available dataset¹ with 280 000 transactions, which happened in the course 48 hours. 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 have introduced two further transformations of the dataset. Based on the hypothesis that time of day

¹ The dataset is available through Kaggle at <https://www.kaggle.com/mlg-ulb/creditcardfraud>

may be very slightly related to the likelihood 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 don't 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. In fact, only 55 transactions in the entire dataset are over \$5000.

To measure the performance of each DNN architecture, the dataset has been split into training and testing sets, and $\frac{1}{4}$ of the records in the original dataset have been assigned to the testing set, based on random selection.

2. METRICS

In essence, 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).

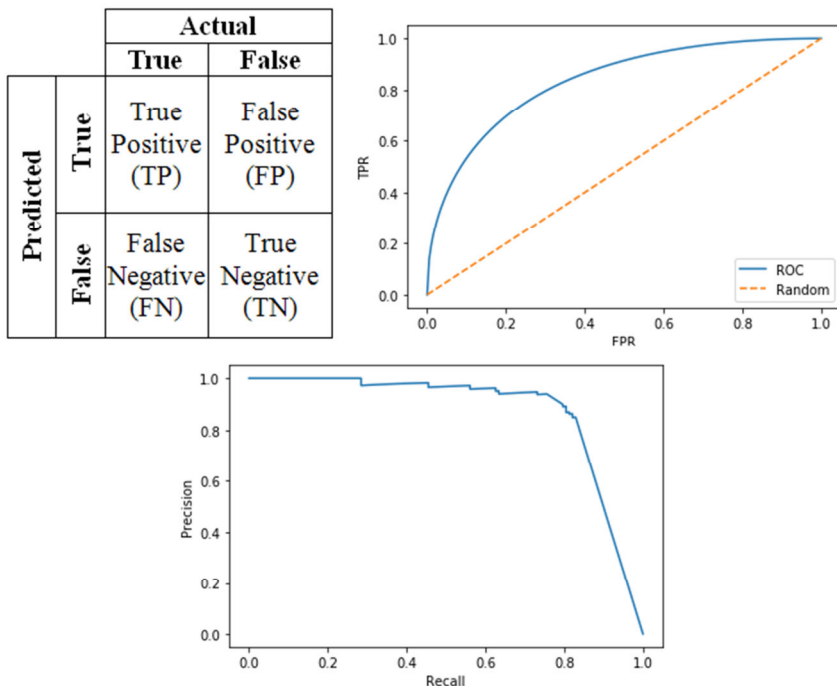


Figure 1. (a) Confusion matrix /left/; (b) ROC curve /center/ and (c) PRC curve /right/

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 which 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.

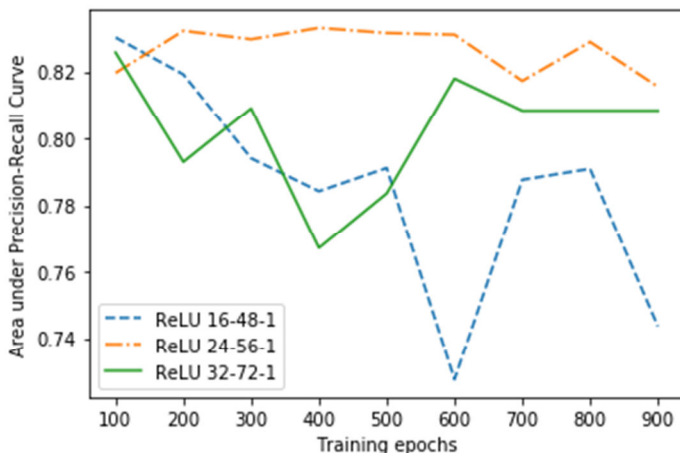


Figure 2. Comparison of DNNs with various layer widths

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.

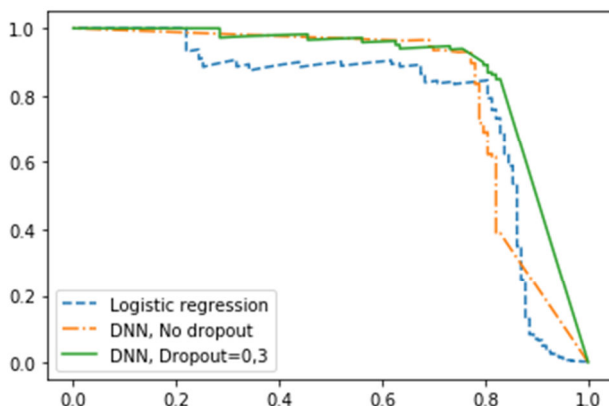


Figure 3. Precision-recall curves of best architectures with dropout and without dropout vs. logistic regression as baseline

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.

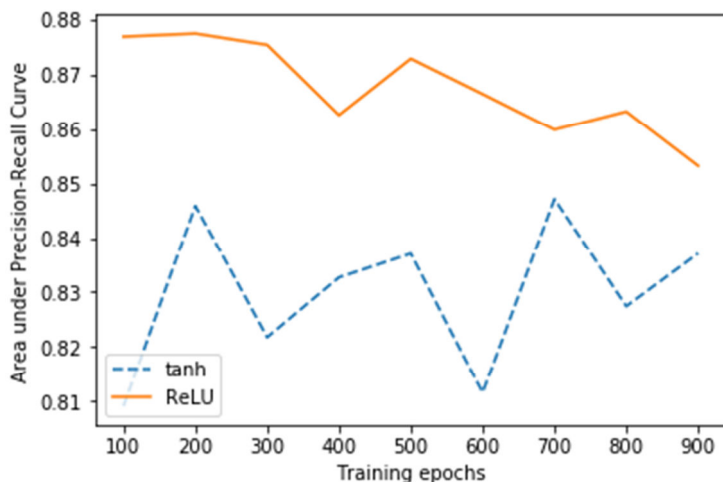


Figure 4. Hyperbolic tangent vs Rectified Linear Unit activations

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, which is consistent with the instability, observed in some GANs (Chintala & LeCun, 2016). 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.

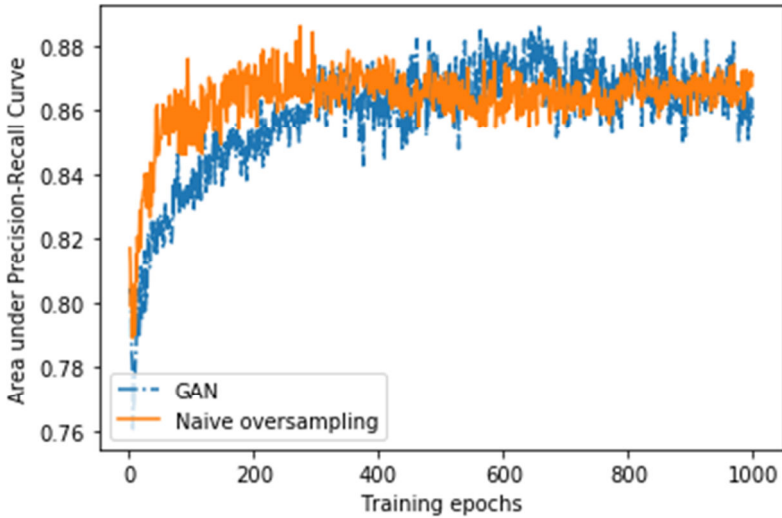


Figure 5. Naïve oversampling vs. GAN-inspired approach

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.

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 doesn't hide any 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-learning-through-adversarial-networks/>>
2. GOODFELLOW, I. J. et al., 2014. *Generative Adversarial Nets*. s.l.:arXiv:1406.2661v1.
3. PARUSHEVA, S. & ATANASOVA, T., 2016. *Card Fraud Prevention Capabilities With Intelligent Methods*. s. l., 16th International Multidisciplinary Scientific Geoconference SGEM.

PRIORITIES FOR DIGITAL TRANSFORMATION IN PROPERTY MANAGEMENT

Miglena Stoyanova¹

*¹ University of Economics – Varna/Department of Informatics,
Varna, Bulgaria, m_stoyanova@ue-varna.bg*

Abstract

The intensive development of information and communication technologies and the constantly increasing competition and challenges for business as a whole require digital transformation of business processes related to property management. The purpose of this research paper is to present the priorities for digital transformation in property management. The aspects of the sector influenced by the digital transformation are distinguished. The main method of analysis is content analysis of scientific publications and expert opinions of specialists from property management.

This research work is conducted as part of the project "Digitization of Economy in Big Data Environment" (DEBDE), project № BG05M2OP001-1.002-0002.

Keywords: *priorities, digital transformation, digitalization, property management.*

INTRODUCTION

Property management is a complex and dynamic process involving a wide range of activities that are often interrelated. All of them should be well coordinated, controlled and analyzed to make the process effective. The growth of property volumes and the constant demands for reducing the cost of their maintenance require the need to use information technology supporting this activity. Their main goal is to automate the majority of business processes, improve communication, management and monitoring, reduce costs through centralized management and control of property information, contracts, office supplies, rents, various types of costs and much more.

The current intensive development of information and communication technologies and the constantly increasing competition and challenges for business as a whole require digital transformation of business processes related to property management. The introduction of appropri-

ate digital technologies helps optimize resources, improves planning and customer service, increases the quality and productivity of the offered services, provides the opportunity for precise analysis and forecasting, facilitates the decision making process and enhances the competitiveness of organizations. Striving to attract more customers, reduce operating costs, increase property value and ensuring transparency and stability of business processes related to properties and assets are also among the main motives of business towards digitalization.

The purpose of this research paper is to present the priorities for digital transformation in property management, based on a summary of scientific publications, practical examples and expertise of specialists in the subject area. The findings show that companies from the property management area need to build and follow a realistic comprehensive digitalization strategy with a clear and expanded long-term vision. This requires defining a balanced transformation for all participants in the field – on the one hand, it must be well-designed, planned and effectively implemented by the company's leadership and, on the other hand, must allow for creativity, new ideas, faster market rates, flexibility and open partnerships. The study is useful for other researchers, who work in the field of digitalization, including professionals and practitioners, who are considering how to implement the digital transformation in property management.

1. ASPECTS OF THE SECTOR INFLUENCED BY DIGITAL TRANSFORMATION

Digital transformation affects all aspects of the sector - from construction to asset management, from marketing to sales. For this reason, it is recommended to use technology tools that can cover the various property management activities – starting with the correct description of the property, going through the contractual relations management, description and control of the operating systems, revenue and cost centers, support and reaching out to the ability to retrieve correct and up-to-date information. A distinction between digitalization of documents, processes and properties can be made in property management.

The digitalization of documents necessary for the property management is key to digital transformation. There are tools that can auto-

matically recognize and verify the type and content of the documents when they are entered. Some applications use artificial intelligence for this purpose, learning from the information provided. This can be very useful, for example in leasing. In addition, improvement of document management systems enables users to continuously modify documents and work on the same document simultaneously with others. The use of a digitized registration system makes it easier to search for data and allows access rights to be defined. The use of virtual spaces allows the storage of data on external, Internet-based cloud platforms, so the access to them is at anytime and anywhere. When the period of data retention expires, they are archived and the cloud memory is cleaned. In this regard, one of the challenges is to gain the trust of the people and organizations involved in the process and to find solutions that are in compliance with the data protection requirements applicable in an organization.

Digitalization is not limited to recording, processing, and archiving of data, but also involves **embedding organization processes into a coherent IT structure**. This means better interconnection of inner company processes in a network and the ability to involve partners of external services in the chain of processes. Existing ERP systems can be expanded with additional digital data processing capabilities. The number of start-up PropTech¹ companies is also increasing. They offer technology services mostly in the form of add-ons to existing ERP solutions. Specific standards in property management have to be adapted in existing ERP systems (Vasilev, 2013). Current developments suggest that in the future these companies will transform existing business models of organizations throughout the property lifecycle by using innovative information and communication technologies and delivering solutions that optimize and automate property management. BIM (Building Information Modeling) systems are increasingly used in the development of projects - from the digital modeling of buildings to the operational property management.

¹ PropTech (Property Technology) is a collective term used to define start-up companies offering technologically innovative products or new business models for property markets.

The digitalization of the properties themselves includes individualized, efficient design of buildings using different technologies. Currently, the market offers a wide range of sustainable solutions. In this regard, for example, the European Commission has adopted new directives for electricity management, which are mandatory for the large companies. They require intelligent reporting of the electricity and gas consumption in order to increase the energy awareness in EU Member States. Sensors and intelligent solutions allow measurement of information for the use of buildings or installation of tools for their optimization. Solutions of this type, related to the Internet of Things, are encountered by the term “Real Estate 4.0”. In the context of digitalization, properties are turning into smart properties, which in the future are expected to grow into a smart city.

2. PRIORITY AREAS FOR DIGITAL TRANSFORMATION IN PROPERTY MANAGEMENT

According to a PwC survey (PwC, 2017b), representing the views of experts from a number of global property management companies, digitalization is expected to provide a better basis for control and decision making in the sector. The clear vision of the role and benefits of digitalization across the business is essential in the short and medium term to be able to respond adequately to the needs of key stakeholders, including customers, tenants and service providers. The constant changes in their preferences cause the businesses to look for new ways to increase market share and their loyalty (Nacheva and Sulova, 2018). The development, implementation and integration of a digital strategy that allows for coordination of operational activities and customer experience through managed assets is a key point for the realization of the digital transformation in this area.

The most frequently mentioned areas for possible specific actions are the reduction of interfaces between applications and the acceleration of automation. As a result, respondents expect, on the one hand, to reduce the high costs of using different systems and media and, on the other – to improve access to service providers' systems, facilitate management and increase efficiency.

The companies surveyed also expect improvements in data management and quality (data standardization), which should enhance the

transparency of their property portfolios. In this regard, the greatest challenges are related to the availability of heterogeneous data and processes and a large number of incompatible systems, especially for international companies, whose portfolio is scattered around the world. Therefore, some companies are intensively working towards reconciliation of their data, processes and systems. As a result, data analysis is facilitated and precised, which is a prerequisite for making better and faster decisions.

Some companies define with high priority the transformation of the work environment (the office) as part of the digitalization process, which creates opportunities for workplace sharing, remote work, flexible working time, more appropriate and efficient use of the space. There is also potential for more effective supervision of the building technologies and/or the management of support activities (facility management).

More advanced in digitalization companies are planning to introduce sensor technologies to improve usability efficiency. This allows, for example, automatic shutdown of a building electricity, when it is not used. The complex relations in business processes in property management may be realized by integrating facility management software such as building management systems (BMS) and building automation systems (BAS), used to control installations (heating, ventilation, elevators, air conditioning, etc.) of buildings (Raychev, 2014), with other software. This communication requires specific organization of business communications and technological realization – by using web services (Vasilev, 2017).

Digitalization leads to fundamental changes both in and outside the IT environments of the companies, making it an important factor for the development and design of new business models and processes, for example introduction of new more flexible mechanisms for lending (Raychev, 2010) or maintenance, operation and repair of properties (Raychev, 2009). Besides cloud applications and big data, which are relatively widely used in the study area, there is an increasing interest in other new opportunities such as data mining. The main priority of business is to achieve greater flexibility and speed in the integration of innovative software solutions. These changes imply an increased need for investment in digital infrastructure.

3. PRIORITY TECHNOLOGIES FOR DIGITAL TRANSFORMATION IN PROPERTY MANAGEMENT

In recent years there has been an essential increase in technological innovations entering the market. In order to identify those that are significant for the property management and/or asset management sector, PwC (PwC, 2016) explores more than 150 technologies, evaluating them in terms of their feasibility and impact on business. The analysis identifies eight priority technologies that are supposed to have the greatest impact on the industry in the coming years. They are:

- **Artificial Intelligence** – It is expected to have the greatest impact on asset management, property management and maintenance activities. Some potential applications of artificial intelligence are related to: identity management, partner transactions, supply chain management, smart contracting, inspection/traceability, asset/property registration, asset management.

- **Augmented Reality** – It is expected to affect mostly asset management, property management, maintenance activities, and relationship management with tenants and customers. It is used for interactive demonstrations of catalogs, precise visualization of large buildings using advanced 3D models, provides new opportunities for advertising and saves time and resources.

- **Internet of Things (IoT)** – It is expected to have a significant impact on asset management, property management, maintenance activities, portfolios and customer relationships. Investors, for example, expect access to information from anywhere and at any time, in appropriate quality and quantity, as well as transparency for their investments. Different channels can be used for this purpose - websites, smartphones, tablets, digital watches.

- **Drones** – the greatest potential for this technology is observed in property management and maintenance activities, as well as in property evaluation. In the property sector, performance monitoring and preventive maintenance measures are essential to extend the lifecycle. This applies in particular to critical equipment, which can cause considerable damages in the event of a malfunction. Drones technology can provide images of hard-to-reach areas that are needed to quickly assess their

condition or to identify potential problems, preventing potential dangers to humans and materials.

- **3D printing** – the most common application is expected in property management and maintenance activities. In the near future this technology may change or even remove some business processes in the property management and construction sectors and have a significant impact on production processes. Many different types of 3D printing materials are already used, including concrete, glass, plastic and wood. The use of this technology allows the production of a wide range of goods with few resources and in a cost effective manner. The first ever 3D high-rise building is designed in Tokyo. It consists of modular housing, designed to be built on site, using a huge 3D printer (Haseef, 2015).

- **Virtual Reality** – it is used, for example, for simulations and visualizations in the design of buildings. Professionals are of the opinion that this technology will affect the areas of asset management, portfolio management and relationship management with tenants and customers (for example investors).

- **Robotics** – physical robots are primarily used in construction and management of maintenance activities. Virtual agents can be helpful in property management, for example in the front office, interacting directly with customers and employees, and significantly relieve work in many repetitive business processes (such as invoicing, performing calculations, verification of documents). The application of robots will give companies greater control over risk and compliance.

- **Blockchain Technology** – it is used for digital storage of documents, such as proof of ownership at some point in time, for retail or asset shares, for digital identity verification and custody services. Some countries use blockchain technologies for their land registers (for example Sweden). Other possible applications are smart contracts, especially by connecting to the Internet of Things (IoT). The experts expect the most significant impact of these technologies on management of: assets, portfolios, customer relationships, funds and investments. With their help, complex and inefficient processes, involving a large number of intermediaries, can be fully transformed and simplified (for example, direct interaction between supply and demand) and some cost factors

such as the use of clearing houses and other important third parties (notaries, government agencies, banks, etc.) may be eliminated (PwC, 2017a).

The presented eight technologies are expected to change the existing business models, especially strategies, customer engagement, operational needs and people. Combining them with already-in-use cloud technologies and new, user-friendly interfaces will greatly assist their faster adaptation.

CONCLUSION

In recent years, a number of property companies have begun to prioritize digital transformation as they understand the significant benefits that it brings in the current business environment. The opportunities offered by modern technologies, the rapid growth of various start-up companies, the increasingly demanding customer expectations derive the need to build and follow a realistic comprehensive digitalization strategy with a clear and expanded long-term vision. This requires defining a balanced transformation for the participants in the field – on the one hand, it must be well-designed, planned and effectively implemented by the company's leadership and, on the other hand, must allow for creativity, new ideas, faster market rates, flexibility and open partnerships.

REFERENCES

1. HASEEF, R. (2015) *Pod Vending Machine*. Available from: <https://www.haseefrafiei.com/pod-vending-machine> [Accessed 17/07/2019]
2. NACHEVA, R. and SULOVA, S. (2018) *Approach to exploring users' expectations of digital services' functionality*. *Ekonomiczne Problemy Usług*, Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecinskiego, 2 (131), pp. 137-145.
3. PwC (2017) *Blockchain and Smart Contract*. Available from: <https://www.pwc.de/en/newsletter/it-security-news-en/blockchain-and-smart-contracts.html> [Accessed 17/07/2019]
4. PwC (2017) *Professional. Global. Digital. Tomorrow's Corporate Real Estate Management*. Available from: <https://www.pwc.de/en/real-estate/trends-in-crem-2017.pdf> [Accessed 17/07/2019]

5. PwC (2016) *Tech breakthroughs megatrend: How to prepare for its impact* Available from: <https://www.pwc.com/gr/en/publications/assets/tech-breakthroughs-megatrend-how-to-prepare-for-its-impact.pdf> [Accessed 17/07/2019]
6. RAYCHEV, T. (2014) *Construction innovation (in Bulgarian)*. Varna, University publishing house "Science and economics", p. 123.
7. RAYCHEV, T. (2010) *Mechanism for differentiation of associations of homeowners in a condominium property regime in the case of crediting for remediation (in Bulgarian)*. Construction Entrepreneurship and Real Property. Proceedings of the 25th Scientific Conference with International Participation in November 2010, pp. 131-137.
8. RAYCHEV, T. (2009) *Sources of income from maintenance, operation and repair of real property in a condominium regime (in Bulgarian)*. Construction Entrepreneurship and Real Property. Proceedings of the Scientific Conference with International Participation in November 2009, pp. 124-130.
9. VASILEV, J. (2013) *Application of the HACCP standard in ERP systems*. Commodity science traditions and actuality. Proceedings from the XII Jubilee Scientific Conference with International Participation. Varna. Publishing house "Grafik", pp. 207-212.
10. VASILEV, J. (2017) *Electronic logistics in the context of globalization*. Varna, Monographic library "Tzani Kalyandzhiev", p. 193.

ORGANIZATIONAL VARIANTS OF IT DEPARTMENT AT THE UNIVERSITY

Mihail Radev¹

*¹ University of Economics - Varna/Informatics, Varna, Bulgaria,
radev@ue-varna.bg*

Abstract

The aim of this paper is to discuss variants for university IT department organization and mainly to explore and analyse two models for the management of IT services, systems, and infrastructure - outsourcing model through the use of an external IT service provider and in-house model. The two models are compared based on the author's expertise and literature review. The goal of the comparative analysis is to determine when outsourcing is useful for university and when not. The analysis can be used by universities to make decisions about managing and maintaining IT assets and services at the university.

Keywords: *Outsourcing model, In-house model, University IT department, IT management.*

INTRODUCTION

The main goal of this paper is to explore the organizational options for a modern IT department with a focus on those that allow a remote and outsourced organization of IT work. In relation to the set goal, the current organizational forms for an IT department are analysed, together with options for outsourcing all or part of the activity to external organizations. A comparative analysis of the advantages, costs and risks related to the different variants is performed. Finally, a recommendation is proposed for choice between own and outsourced IT department.

1. IT MANAGEMENT IN UNIVERSITIES – ORGANIZATIONAL ASPECTS

A fundamental paradigm of this research is the current up-to-date approach and our own belief that IT departments in universities, as well as in companies in other fields of human activity, must be considered and function as full-fledged business units. Through a set of policies, processes and procedures, these business units manage, implement, im-

prove and maintain customer-oriented IT services. The focus of modern IT departments is not on hardware, networks and systems, but on IT customer service in accordance with organizations' business goals. The management of IT services in this way is intended to serve the customers rather than treating IT as a department that manages technology (CIO Media, 2017).

The management and organization of such modern IT departments are in accordance with the rules of the modern management. IT department leaders are now guided by tasks that are common with those before the heads of organizations at each of their hierarchical levels – to reduce costs, to increase efficiency, to improve control, to increase productivity, to move on opportunity from capital to operating expenses. They measure and value activities, services and resources provided by their departments in a way similar to that of stand-alone business structures. The current view of the IT department as an expense is changing in the direction of considering the IT department as a revenue generator. The other departments in the organization must justify and plan their costs for IT support and IT resources and pay for the use of IT services. If the IT department doesn't have the ability to provide the needed resources, it could use external contractors, outsourcing partners, cloud infrastructures or hire additional employees. All this has to happen within his budget. The ability to implement different approaches for resources allocation is crucial to the IT department. This could help to the transformation toward an innovative, agile, proactive and adaptive organizational unit. At the same time the senior management should consider the IT department as a partner for achieving the strategic goals of the organization.

In many cases, the IT Department operates on the principle of a fire brigade that deals with sudden floods by plugging holes, rather than a business organizational unit that provides IT services to its customers. With such approach of a reactive activity organization it is impossible to maintain processes of IT services management, constant monitoring of the status of the services and infrastructure; documenting the actions of the specialists in the department and the services that are provided.

The main limitation of the described model for IT departments in universities to function as a business subdivision originates from the complete centralization of the budget at universities and the non-autonomy of the organizational units. This means that in order to be able to make such a transformation of the IT department it is necessary to decentralize the budget before. A certain degree of decentralization is also needed for the head of the IT department in order to be able to take operational decisions on his own or at least quickly.

In case the IT department is transformed into a full-fledged business organizational unit, it will work and follow the stages of the management process: planning, organizing, leading and control. (See fig. 1.) Planning involves defining goals, creating a strategy and developing a plan for co-ordinating activities; Organizing describes what tasks need to be done, by whom, how the tasks can be grouped, who is accountable to whom, where decisions are made; Leading includes motivating employees, choosing the most effective communication channels, resolving conflicts; Control is the process of monitoring performance, comparing with targets and correcting significant deviations (Robbins & DeCenzo, 2008).

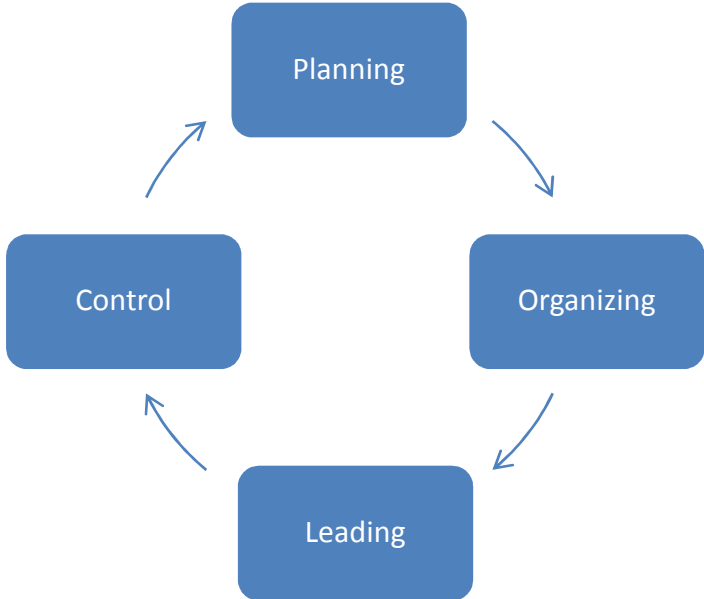


Figure 1. Stages of the management process

The university's IT department management plan should define the business objectives, requirements and principles of service management at the university, all costs and justification for them, benefits, options, possible problems, risks. Before planning, it is necessary to audit the IT infrastructure to assess where we are going, what is the current state.

The audit must include an audit of the network infrastructure. An expected result of this process is a detailed scheme of network connectivity at the university – physical network topology, active equipment, active equipment configuration, port description and address schema; audit of the wireless network to assess where additional antennas should be placed, where the signal is weak; assessment of the weaknesses in the network, assessment of the points that would stop the entire network. This whole "screenshot" of the current state can be performed by the IT department or an external company.

The proposed changes included and detailed in the management planning stage are in the following areas:

- focus on the needs and results of the clients – students, teachers, PhD students, administration;
- specialization of the team of end users support specialists, network administrators, system administrators; forming an annual IT budget;
- Based on the principle of a department offering IT services. The IT department acts as a service provider to the business – teachers, students, PhD and administration;
- Economic (business) measures of the Division's activities;
- Proactive management and monitoring of the IT infrastructure and the provided IT services and problems prevention.

The second management stage – organizing – includes description of tasks to be carried out, as well as a possible reorganization of the existing organisational structure and creation of a new one. The modern organizational forms of the company's activities are suitable for application in the IT departments of the universities. IT specialists can be divided into different teams, can form an internal hierarchy and implement self-government within the department. All modern organizational forms rely on horizontal communications in order to make managerial deci-

sions faster, offering and implementing innovations, which is appropriate for the specifics of the activity in the department.

A modern form of activity organization is the team-oriented organization (Serafimova, 2015). It can be applied within the organizational unit, assuming that the department employs specialists who can switch from team to team – for example, from the team maintaining the website in the team of software developers or from the team of network administrators in the system administrators team. Due to the multiple competencies of employees in IT departments, specialists can participate in several projects simultaneously.

Network structures of different organizations created to share production, territorial, market and/or non-market resources for the creation and realization of complementary products and services (Georgiev, 2007) may have an application in IT departments of universities, but consistent with the dominant hierarchical links in the main organization and the resulting need for some degree of centralization within the department. It is possible to use outsourcing of some or all IT services and/or cloud solutions for them which will lead to the creation of a network structure with university and other organizations. Usually there is a limited number of IT department employees who are capable enough to keep up with the ever changing technologies and demanding requirements. That's why the outsourcing of IT activities to external partners will allow the usage of their know-how, knowledge base and best practices for business process management which would eventually lead to cost reduction and more reliable IT services. The university's IT department can benefit from two opportunities offered from the network organizational structure – outsourcing and cloud services.

Outsourcing is the management of IT services, systems and infrastructure through the use of an external IT service provider. This can be done by organization's own infrastructure, with services and infrastructure being managed by an external organization.

Companies choose to outsource their operations to external organizations for the following main reasons (David, 2011):

- Economy of scale. Outsourcing is less expensive than managing an own IT department, because the outsourcing company serves multiple clients and its fixed costs are distributed among them. The specialization

of the outsourcing company in one or several activities increases its efficiency and allows their offer at lower prices;

- Focus on company's core business and key competencies and ensuring more people, time and capital to them. It enables the company to provide better products and services;
- Alignment with good practices, application of modern technologies, experts and import of know-how;
- Concentration on other internal activities in the value chain, critical to maintaining a competitive advantage.

This model, implemented in the organization's IT department, has specific advantages – reducing operating costs, focusing on strategic issues, increasing flexibility and improving the quality provided by IT services, facilitating access to technology and reduce the risk of using outdated technologies, reducing IT staff and technology costs (Gonzales, et al., 2014).

At the same time, outsourcing has risks. There is a high probability for forming a dependency on the supplier. The binding with an external partner means that the organization relies on the qualification of the supplier's specialists and procedures of the supplier's work. IT employees of the organization using external provider reduce or lose their IT qualification at all, security issues are also possible. There is a very serious danger of reputation risk – when hiring outsourcing company, the performance of the activity is transferred from the university to the company-provider, but not the responsibility to the clients. Therefore, in case of failure of the provider, the university should be able to effectively resolve the situation in a short period of time to avoid reputational damage.

The two options for using outsourcing are: 1. Delegation of all maintenance to an external company or 2. A mixed version. The disadvantages of the first solution are the following: the critical processes, those on which an organization's business is created and operated, should not be outsourced to other organizations.

Critical processes must be under the full control of the management of the organization. Likewise, the university can outsource the maintenance of the green areas, the canteens of students and teachers, etc. but managing and maintaining IT infrastructure – servers, information systems are very important processes and even if it is more cost-effective, it should not be outsourced to another company.

In the mixed version only the management of routine IT services is outsourced to external IT service provider and the more important and critical IT services and infrastructures are managed, monitored and supported by the corporative IT department.

CONCLUSION

For large infrastructures – with many servers, many users, it is more advantageous to the organization to maintain its own IT infrastructure and IT department, because this way it can control the infrastructure and data; data security is at a higher level; price is more profitable and IT knowledge is kept within the organization. The university also deals with science, research and projects. There may be patents, start-ups, spin-outs companies. This implies a higher degree of confidentiality in terms of competition and of external companies. The presence of laboratories in the university with IT equipment allows students, doctoral students and lecturers to engage in scientific developments to reach the market.

REFERENCES

1. CIO Media, 2017. *IT management for business needs*. [Online] Available at: http://cio.bg/8988_it_upravlennie_zh_nuzhdite_na_biznesa [Accessed 01 09 2019]. (in Bulgarian)
2. DAVID, F. R., 2011. *Strategic management: Concepts & cases*. 13th edition ed. s.l.:Prentice Hall.
3. GEORGIEV, R., 2007. *Organizational structures and corporate security*. [Online] Available at: [http://www.rgeorgiev.com/ организационните-структури-и-корпор/](http://www.rgeorgiev.com/организационните-структури-и-корпор/) [Accessed 02 09 2019]. (in Bulgarian)
4. GONZALES, R., GASCO, J. & LLOPIS, J., 2014. *Information systems outsourcing reasons and risks: a new assessment*. [Online] Available at: https://www.researchgate.net/publications/220672187_information_systems_outsourcing_reasons_and_risks_a_new_assessment [Accessed 02 09 2019].
5. ROBBINS, S. & DECENZO, D., 2008. *Fundamentals of management*. Sixth edition ed. s.l.:Pearson Prentice Hall.
6. SERAFIMOVA, D., 2015. *Management theory*. Varna: Steno (in Bulgarian).

IOT PROBLEMS AND DESIGN PATTERNS WHICH ARE APPROPRIATE TO SOLVE THEM

Mariya Armyanova¹

¹ *UE-Varna, Bulgaria, armianova@ue-varna.bg*

Abstract

Internet of Things provides the possibility to cooperation and manage the devices with the global network. But the development of IoT systems is fraught with a variety of complicated problems and a good solution is to use design patterns. IoT patterns are still not strictly defined and classified as the domain is under development in the moment. The purpose of the report is to present common problems and opportunities to overcome them through IoT patterns. The applied approach includes research and analysis of scientific publications, official websites and design patterns classification. The report results are useful for IoT system developers who want to know about common problems and how to be solved and where to find appropriate design patterns to their problem. The main conclusion of the report is that in order to be easily found when needed, the IoT patterns are appropriate to be classified by the groups of problems they solve. In the report, the approach for classification of IoT design patterns is pointed out and by the defined criteria they are classified.

Keywords: *design patterns, IoT, security, Design, Architecture.*

INTRODUCTION

Different organizations and research institutes realize the IoT, by using manifold technologies, standards, which are involved by different organizations and sectors. So it often leads to conflicts of interest. Therefore, when developing an IoT system, there are various complex problems to solve. Problems are compounded by increasing the count of devices connected to the global network. Many factors are driving the process from reducing the device cost to the new communications technology capabilities. The concept of tying different devices to the Internet is seen as Internet of Things (IoT). The IoT is a symbiosis between various leading technologies that aim to enable device recognition. The device number on the global network is expected to reach 26.66 billion in 2019. The trend is a steady increase in devices connected in IoT. Globally, their number is expected to reach 75.44 billion by 2025, which is threefold increase in 5 years (SRD, 2016). The trend is presented in fig. 1.

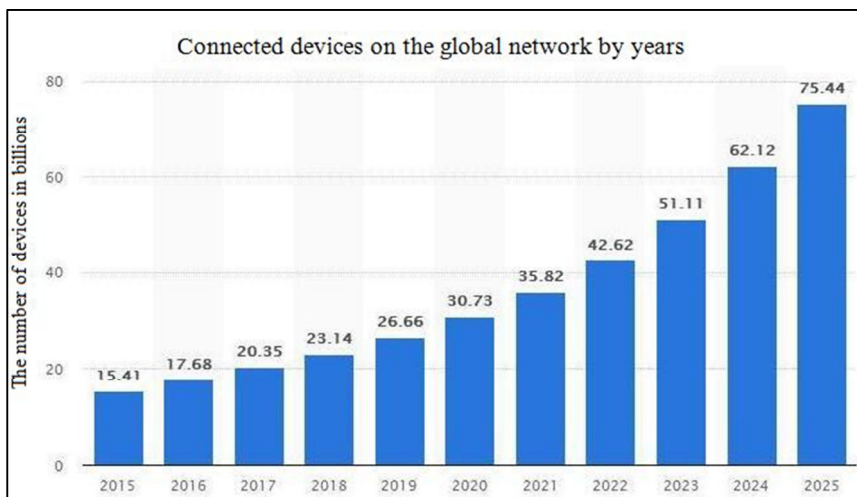


Figure 1. The trends in the device number for the period 2015-2025

The IoT is based on such technologies as machine learning, artificial intelligence, big data and intelligent analyzes. IoT platform is used in various domains from smart homes and smart cars to agriculture, healthcare and smart supply chain. Each domain and technology sets specific requirements, which are often contradictory. So this poses a variety of complicated problems to the development of systems, enabling the work of different devices into the IoT environment. For some of the problems, the development teams have already found an effective solution, which is formally described as a pattern. The patterns present a concept that enables developers to reuse the experience of problem solving. They can be seen as descriptions of common problems, their abstract solutions, and their consequences.

The individual manufacturers often develop devices only for a specific domain area, such as healthcare or smart cars, smart homes, so for such problems it is leading to an increase in the number of different solutions. The existence of too many IoT standards and technologies proposed by various device developers, as well as a lack of common understanding in the domain (Guth et al. 2016) again popularizes the use of design patterns as a possible solution to problems.

1. THE IOT AND ITS PROBLEMS

Internet of Things (IoT) is seen as a concept of linking different devices to the Internet so that they can be recognized by other devices. The term was introduced in 1999 by Kevin Ashton at his presentation.

The technical advancements and the communications developments also provide opportunities for the IoT development. Since the concept of IoT can be used in different domains, so the term refers the different systems, such as Smart Homes, Smart Office (Le Gal et al., 2001; Röcker, 2010), Smart Grids (Kopp et al., 2015) , or the Smart City concept (Nam and Pardo, 2011; Su et al., 2011), the Industrial Internet (IIC, 2015), and others. However, the main idea is to send the results of the sensors to the Internet to enable the analysis of all the accumulated data. The results of the analysis are used to the better device management in the domain.

According to Voas, the IoT involves sensing, computing, communication, and actuation. The IoT implements device unification into a ubiquitous layer of cyber-physical devices would then be able to measure itself and its environment; combine, process, and analyze these measurements to gain new insights; and use these insights to optimize itself and the environment (Voas, 2016).

Some of the difficulties in the IoT system implementation is due to the numerous requirements for the IoT platforms. The requirements relate to performance, reliability, scalability, reuse, modification and interoperability. The main IoT platform features that make implementation difficult are, for example, the lack of centralized management, continuously device connection or disconnection to the system, user anonymity, etc. Despite their implementation difficulties, they have many advantages for IoT systems. For example, decentralization provides scalability and reliability because it involves the use of all resources, eliminating intensive traffic to only one point, which reduces delays and guarantees the reliable work of the system even in the event of a resource failure. Of course, for developers this architecture requires solving of many problems, such as security-related issues, dynamic device connectivity, decentralized data collection and processing, and more.

The IoT systems complexity is also due to the distribution of tasks across many devices. The situation is also complicated by the physical remoteness and distribution of roles in the system development between many different often independent participants. The device is developed by a manufacturer, then is installed by another company, and is often maintained by a company that is different from the software development company.

The many of the IoT system problems are related to security. In a network environment, the device needs to communicate with many potential applications and systems. Therefore, it is difficult to guarantee its security, because of the difficulty of distinguishing malicious users who want to gain control of the device or access its network.

Another problem is that a potential target for malicious attacks is the device network. It often is attacked by sending unsolicited communication requests from an unreliable communication partner. As a result, security trust issues arise. Device owners are afraid to provide access to their devices or their data without full control, because the device communication with third parties or its access security is not always well protected. The devices should react differently depending on the request sender.

The problems are compounded by the growing number of IoT users' devices. Rights management, data sharing, and device control are complicated, too. They are often accessed through great number of gateways, and data collection and processing is through different cloud systems.

The integration of devices into a united system causes and other problems. Some devices do often not support the networks communication technology or protocol, because they use different standards. In order to save energy or due to other network features, some devices are not permanently plugged into the general network environment, making it difficult to access them.

The main challenges that IoT system developers face can be reduced to the following categories of problems:

1. The choice of the right architecture for IoT applications with to enable the acquisition and processing of huge amounts of sensor data simultaneously;
2. The combining of different devices, some of which do not support the communication protocol or the technology, into a common IoT system, or integrating multiple different protocols into an IoT platform;

3. The ensuring of the communication security and reliability with the devices, as well as the physical access to them;

4. The providing energy-saving capabilities and, at the same time, to ensuring that the device is switched on immediately when needed.

The security issues are becoming more important because of the increasing number of devices, which leads to more frequent hacking attacks as the number of potential targets increases. Moreover, the use of different technologies and mechanisms makes difficult to create effective protection for all of them.

The creating of an IoT system requires the provision of multiple aspects, which complicates the problems. The design patterns that summarize the experience of specialists offer the ability to solve some of the systems development problems.

2. THE IOT DESIGN PATTERNS

The IoT is a dynamically evolving domain, so new design patterns are constantly described to solve the problems that arise with the IoT systems development. Because of the many patterns there is confusion among developers regarding the use of the right pattern in the project situation. Therefore, the report offers an opportunity to classify the patterns to bring arrangement among set of IoT patterns. In 2018 (Bloom and etc., 2018) is proposed the grouping of patterns based on the architecture layers and categories of IIoT applications, which are only suitable for architectural patterns.

Washizaki (Washizaki et al., 2019) offers a patterns classification based on the level of abstraction. Here, the patterns are divided according to the architecture layer into: high-level or architectural patterns; mid-level or design patterns; and low-level or software-dependent patterns. Such a classification could direct the developer to an appropriate patterns for the architecture layer he is developing, but it is insufficient because most often the developer tries to solve a specific problem. The patterns are too many to find the right one. Therefore, it is necessary to add more classification criteria that are specific only to patterns in the domain area of IoT systems.

IoT patterns are a small part of all patterns used in the systems development. They are a subset derived from the classification of patterns

by the criterion Purpose in the general classification scheme (Armiyanova, 2018). Therefore, we propose the basic criteria for the classification of IoT patterns, the criterion Goal. The criterion Layer of the architecture is also appropriate, as it subdivides the patterns according to the development phases and, with such classification, the developer can easily find the appropriate pattern at the appropriate level.

The criterion Goal indicates the specific problem the pattern is addressing. According to Kardell (1997), the Goal is determined by the elements of the system that are influenced by the pattern: integration, functionality, communication, access, security, etc. This classification is applicable not only to design patterns, but also to architectural patterns that can also add goals. But the architectural patterns described to do moment are not so many, so they do not need to be classified.

Koster also presents IoT patterns according to their goals (Koster, 2014; Chandra and Mahindra, 2016). But the categories offered by the authors are many, as they classify all patterns that are suitable for use in IoT systems, and not specifically for IoT. Therefore, there are some categories that can be removed, such as information computing patterns. They are not specific for IoT systems and can be used to build cloud systems, too.

By the criterion Goal the patterns could be classified according to the main groups of problems that they addressed in the previous point. Therefore, design patterns can be divided into the following main groups, defined by the main problem categories: Integration patterns, Communication patterns, Security patterns, and Energy Saving patterns. By adding a group of Architectural patterns to the design patterns, a classification is obtained to allow easy detection of the patterns when needed (Fig. 2).

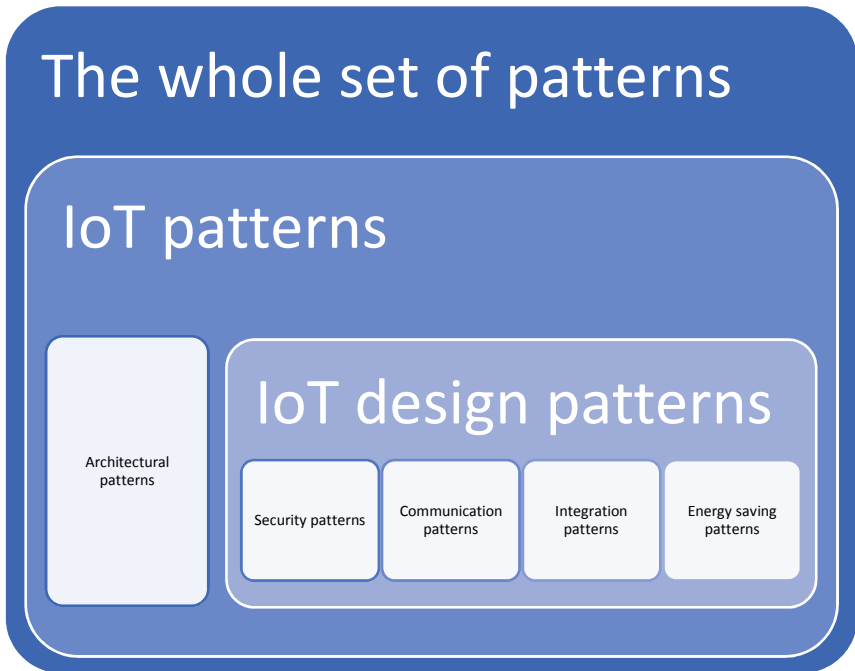


Figure 2. The classification scheme of the categories IoT patterns

The Security patterns are designed to prevent accidental insertion of code vulnerabilities and mitigate the effects of these vulnerabilities. They integrate knowledge of security and system structure, enabling the development and improvement of software. They allow to integrate security policy with software design and development. The Security patterns are numerous. Their levels range from architectural patterns involving system-wide design to execution-level patterns relating to the implementation of individual functions or methods in the system. Some of the security issues are common to all systems. Therefore, in the development of an IoT system, both specialized IoT patterns and patterns can be used that realize one or another aspect of security across all systems. For example, the Whitelisting Firewall patterns of (Villarreal et al., 2013) introduces a firewall to control communication and filter it only to authorized interlocutors without delaying it. Fernandez also offers simi-

lar patterns (Fernandez, 2013; Fernandez et al., 2014). He presents two firewall patterns to control network access to cloud resources. Hafiz cloud security patterns (Hafiz, 2013), Schumacher web application patterns (Schumacher et al., 2006), Dara symmetric and asymmetric encryption patterns (Dara, 2014), or the patterns of Romanovsky (Romanovsky, 2001) and Kienzle (Kienzle et al., 2002) are also suitable.

In addition to generic patterns, IoT-specific systems have been created. Our aim is to look at these patterns. Such are the Reinfurt patterns (Reinfurt et al., 2017). He offers Trusted Communication Partner, Outbound-Only Connection, Permission Control, Personal Zone Hub, Whitelist, and Blacklist design patterns. The Trusted Communication Partner pattern is intended to prevent the device or its network from being accessed by random communication partners by allowing the device to communicate with the device solely in a list of authorized pre-known partners. Another solution to prevent unauthorized access is offered in the Outbound-Only Connection pattern, which supposes only those incoming communications requests that respond to a device-initiated connection to be answered. The Permission Control pattern guarantees the security of device data by retaining on the backend server of the data access rights of the communication partners. The Personal Zone Hub involves the creation of a Hub that allows the management of the permissions, data sharing and control of all devices of a user. The Whitelist pattern involves creating a list of users that have access rights, with no one outside that list having access. The Blacklist pattern is the opposite. It involves creating a list of unreliable partners who are denied access rights.

The another major set of patterns are integration patterns that ensure way to combine devices located at different end nodes with different protocols in an IoT platform. Koster offers design patterns that can use different network and device technologies to solve problems with the physical infrastructure of IoT (Koster, 2014). However, its solutions are not described as typical patterns. Qanbari (Qanbari et al., 2016) also offers patterns for combining devices related to edge: Edge Provisioning Pattern, Edge Code Deployment Pattern, Edge Orchestration Pattern, Edge Diameter Of Things (DOT) Pattern. The Edge Provisioning Pattern provides control over a large number of hard-to-reach scattered end de-

vices, enabling them to be easily reconfigured, as well as to easily incorporate new ones. The Edge Code Deployment Pattern provides the ability to easily maintain device software through decentralized Git version control located on backend server. Edge Orchestration Pattern distributes functionality in such a way that it allows end nodes to control, configure and manage their own end devices, monitor their status, discover the services they need. The Edge DOT Pattern requires the creation of a Metering Server that unifies the way services are measured across endpoints, as different providers may use different usage patterns, such as event-based or time-based.

The next large group of patterns are communication patterns. Peter (Peter, 2016) offering such communication patterns: Request / Response, Event Subscription, Asynchronous Messaging, Reliable Messaging, Multicasting, Publish / Subscribe, Queues, Message Brokers, Federation, Discovery, and Delegation of Trust patterns are not described according to the strict definition of pattern documentation.

Reinfurt also provides patterns for solving communication problems (Reinfurt et al., 2019). Device Gateway, Device Shadow, Rules Engine, Device Wakeup Trigger, Remote Lock and Wipe, Delta Update, Remote Device Management, Visible Light Communication. The Device Gateway patterns is used when part of the devices does not support communication technology or protocol on the network. The pattern involves using an intermediary, such as a Gateway, to translate protocols from and to the device. The Device Shadow pattern is used to send requests to devices that are not permanently connected to the network. The pattern involves creating a virtual address to the device and usually interacts directly with the patterns in the other next group. The Rules Engine pattern allows users to customize the system through a set of simple rules. The Device Wakeup Trigger pattern involves sending a message over a communication channel to a device control unit that is not permanently connected to the network. The Remote Lock and Wipe pattern guarantees the security of the data stored on the device, allowing its data to be erased when the device is unlocked from the system. The Delta Update pattern helps reduce network traffic by sending only device data that has changed since the last time you sent it. The Remote Device

Management pattern allows local management of the device through a client installed on the device that interprets the server commands. The Visible Light Communication pattern involves sending communication signals to or from a remote device.

The last group of design patterns are energy saving patterns which is gaining popularity. Reinfurt (Reinfurt et al., 2017) also offers patterns for this category, dividing them into two types. The former describe the different types of devices by their energy requirements: Mains-Powered Device, Period Energy-Limited Device, Lifetime Energy-Limited Device, Energy-Harvesting Device. Therefore they are not interested in the software of the IoT system. The second type describes how to communicate with devices so that they save energy: Always-On Device and Normally-Sleeping Device. The Always-On Device pattern refers to cases where energy saving is inefficient. The other Normally-Sleeping Device pattern is used in cases where the device does not need to operate continuously and is implemented by deactivating all its energy elements. When the functionality of a such device is required, the control unit supplies power and awakens the device. Even if the control unit is not permanently connected to the network, it may be switched on at regular intervals to check for the need to wake up the device. The patterns that help communicate with similar devices from the reviewed communication patterns are Device Shadow, Device Wakeup Trigger.

The Architectural patterns determine how the elements of the system are organized and interacted. They introduce the basic characteristics and behaviour of the system, the principles and the multilayered structure into the system architecture. The IoT architectural patterns suggest building a single system in a heterogeneous environment. Therefore, components are introduced into the system when is necessary to allow the creation of different types of integration and communication of its components. These include Device-to-Device (D2D), Networking-to-Networking (N2N), Middleware-to-Middleware (MW2MW), Application & Services-to-Application & Services (AS2AS), Data & Semantics-to-Data & Semantics (DS2DS), and protection on different architecture levels that incorporates various mechanisms from authentication to authentication, use of authentication tags, or secure socket layer (SSL). The inclusion of one or the other architectural elements is also influenced by

the purpose of the system, since the systems are heterogeneous. For example, Smart Home and smart supply chain systems have their own specifics. But IEEE (Weyrich and Ebert, 2016) is trying to standardize the fuzzy concepts of definition, architecture and infrastructure. IEEE offers a three-tier architecture with Sensor layer, Networking and data communication layer and Application layer. Referring to this architecture, an IoT architecture pattern (Alreshidi and Ahmad, 2019, p.4) is defined, which implies a system with functions divided into three layers: User interfacing layer, which allows users to monitor and control devices remotely; Interconnection and logic layer, which ensures the coordination and collaboration of devices across the network; Computation and data storage layer, which manages the intelligent analysis and data storage.

CONCLUSION

The IoT systems improve application functionality, increase access to resources, save energy and reduce costs. But their development is attended by many problems. The report presents the IoT system main problems and the abilities for solving them by patterns. The aim is for developers to be easier to find the right patterns when using a top-down development approach. The developers are specialized in specific domains and issues, so the pattern classification by the criterion Aim, enables them to focus on known issue.

REFERENCES

1. ALRESHIDI, A. and AHMAD, A. (2019) *Architecting Software for the Internet of Thing Based Systems*. MDPI, Basel, Switzerland Available from: <https://www.mdpi.com/1999-5903/11/7/153/pdf> [Accessed 12/08/2019].
2. ARMIYANOVA, M. (2018) *Patterns Classification Criteria and Scheme*. Eastern Academic Journal, Bourgas, pp.70-89.
3. BLOOM, G., ALSULAMI, B., NWAFOR, E., AND BERTOLOTTI, I.G., (2018) *Design Patterns for the Industrial Internet of Things*. 14th IEEE International Workshop on Factory Communication Systems (WFCS). Available from: <https://ieeexplore.ieee.org/document/8402353> [Accessed 12/08/2019].

4. CHANDRA, G., MAHINDRA T. (2016) *Pattern language for IoT applications*. Conference on Pattern Languages of Programs (CPLoP). Available from: <https://hillside.net/plop/2016/papers/writinggroup/17.2.pdf> [Accessed 12/08/2019].
5. DARA, S. (2014) *Privacy Patterns in Public Clouds*. Proceedings of the Indian Conference on Pattern Languages of Programs (Guru-PLoP).
6. FERNANDEZ, E. (2013) *Security Patterns in Practice: Designing Secure Architectures Using Software Patterns*. Wiley.
7. FERNANDEZ, E. B. YOSHIOKA, N. and WASHIZAKI, H. (2014) *Patterns for Cloud Firewalls*. Proceedings of the Asian Conference on Pattern Languages of Programs (AsianPLoP).
8. GUTH, J., BREITENBÜCHER, U., FALKENTHAL, M., LEY-MANN, F. and REINFURT, L. (2016) *Comparison of IoT Platform Architectures: A Field Study based on a Reference Architecture*. In Proceedings of the International Conference on Cloudification of the Internet of Things (CIoT). IEEE. Available from: <https://www.iaas.uni-stuttgart.de/publications/INPROC-2016-48-Comparison-of-IoT-Platform-Architectures.pdf> [Accessed 12/08/2019].
9. HAFIZ, M. (2013) *Security Pattern Catalog*. Available from: <http://www.munawarhafiz.com/securitypatterncatalog/index.php> [Accessed 12/08/2019].
10. INDUSTRIAL INTERNET CONSORTIUM (IIC) (2015) Available from: <https://www.iiconsortium.org/pdf/IIC-Overview-11-24-15.pdf> [Accessed 12/08/2019].
11. KARDELL, M. (1997) *A Classification of Object-oriented Design Patterns*. Master's thesis, Department of Computing Science. Umeå University. Available from: <http://www8.cs.umu.se/~jubo/ExJobbs/MK/patterns.htm> [Accessed 12/08/2019]
12. KIENZLE, D., ELDER, M., TYREE, D. and EDWARDS-HEWITT, J. (2002) *Security Patterns Repository Version 1.0*. Available from: <http://www.scrypt.net/~celer/securitypatterns/repository.pdf> [Accessed 12/08/2019]

13. KOSTER, M., (2014) *Design Patterns for an Internet of Things*. Available from: <http://community.arm.com/groups/internet-of-things/blog/2014/05/27/design-patterns-for-an-internet-of-things> Available from: <http://iot-datamodels.blogspot.com/2014/05/design-patterns-for-internet-of-things.html> [Accessed 12/08/2019]
14. KOPP, O., Falkenthal, M., Hartmann, N., Leymann, F., Schwarz, H., Thomsen, J. (2015) *Towards a Cloud-based Platform Architecture for a Decentralized Market Agent*. INFORMATIK 2015, P-246, pp. 69–80. Gesellschaft für Informatik e.V. (GI), Bonn.
15. LE GAL, C., MARTIN, J., LUX, A., CROWLEY, J. (2001) *SmartOffice: Design of an Intelligent Environment*. IEEE Intelligent Systems 16, pp. 60–66.
16. NAM, T., PARDO, T. (2011) *Conceptualizing Smart City with Dimensions of Technology, People, and Institutions*. Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, pp. 282–291. ACM, NY.
17. PETER W. Communication Patterns for the Internet of Things, June 1, 2016. Available from: <https://software.intel.com/en-us/articles/communication-patterns-for-the-internet-of-things> [Accessed 12/08/2019].
18. QANBARI, S., PEZESHKI, S., RAISI, R., MAHDIZADEH, S., RAHIMZADEH, R., BEHINAEIN, N., MAHMOUDI, F., AY-OUBZADEH, S., FAZLALI, P., ROSHANI, K., YAGHINI, A., AMIRI, M., FARIVARMOHEB, A., ZAMANI, A., and DUST-DAR, S. (2016) *IoT Design Patterns: Computational Constructs to Design, Build and Engineer Edge Applications*. IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI), pp. 277-282.
19. REINFURT, L., BREITENBÜCHER, U., FALKENTHAL, M., LEYMAN, F. and RIEGG, A., (2019) *Internet of Things Patterns for Communication and Management*. Springer-Verlag. Available from: <https://www.iaas.uni-stuttgart.de/publications/ART-2019-08-Internet-of-Things-Patterns-for-Communication-and-Management.pdf> [Accessed 12/08/2019].

20. REINFURT, L., BREITENBÜCHER, U., FALKENTHAL, M., LEYMANN, F. and RIEGG, A., (2017) *Internet of Things Patterns for Devices* Available from: <https://www.iaas.uni-stuttgart.de/publications/INPROC-2017-15-Internet-of-Things-Patterns-for-Devices.pdf> [Accessed 12/08/2019].
21. REINFURT, L., BREITENBÜCHER, U., FALKENTHAL, M., FREMANTLE, P., and LEYMANN, F. (2017) *Internet of Things Security Patterns*. HILLSIDE Proc. of Conf. on Pattern Lang. of Prog. Available from: <https://www.hillside.net/plop/2017/papers/proceedings/papers/20-reinfurt.pdf> [Accessed 12/08/2019].
22. RÖCKER, C. (2010) *Services and Applications for Smart Office Environments - A Survey of State-of-the-Art Usage Scenarios*. International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering 4, pp. 51–67.
23. ROMANOSKY S., (2001) *Security Design Patterns*, Technical report, Available from: [http://www.cgisecurity.com/lib /security DesignPatterns.pdf](http://www.cgisecurity.com/lib/security%20DesignPatterns.pdf) [Accessed 12/08/2019]
24. SCHUMACHER, M., FERNANDEZ, E. B., HYBERTSON, D., BUSCHMANN, F., and SOMMERLAD, P., (2006) *Security Patterns: Integrating Security and Systems Engineering*. John Wiley & Sons.
25. SRD, Statista Research Department (2016) *Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions)*. Available from: <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide> [Accessed 2/08/2019].
26. SU, K., LI, J., FU, H. (2011) *Smart City and the Applications*. 2011 International Conference on Electronics, Communications and Control (ICECC), pp. 1028– 1031. IEEE, Piscataway, NJ.
27. VILLARREAL, N., FERNANDEZ, E., LARRONDO-PETRIE, M. and HASHIZUME, K. (2013) *A Pattern for Whitelisting Firewalls (WLF)*. PLoP 13. Available from: <http://www.lacpei.org/LACCEI2013-Cancun/ExtendedAbstracts/EA055.pdf> [Accessed 12/08/2019].
28. VOAS, J. (2016). *Networks of 'Things'*. NIST Special Publication 800-183. <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-183.pdf>

29. WASHIZAKI, H., YOSHIOKA, N., HAZEYAMA, A., KATO, T., KAIYA, H., OGATA, S., OKUBO, T. and FERNANDEZ, E. (2019) *Landscape of IoT Patterns*, SERP4IoT '19 Proceedings of the 1st International Workshop on Software Engineering Research & Practices for the Internet of Things, pp. 57-60.
30. WEYRICH, M., EBERT, C. (2016) *Reference Architectures for the Internet of Things*. IEEE Softw. 33, pp. 112–116.

CORPORATE BIG DATA AND HYBRID INTEGRATION PLATFORMS

Yavor Christov¹

*¹ University of National and World Economy, Sofia,
Bulgaria, ychristov@unwe.bg*

Abstract

The objective of this paper is to propose an approach to link corporate data from local to external entities, using hybrid integration platforms (HIPs). An evaluation on different integration approaches is made, following a review of HIPs. The study looks at cloud-based applications, enterprise's partner's application servers in data centers, and Internet of Things (IoT) devices. An observation of how HIPs helps reducing the costs, complexity and time needed for end-to-end integration process is made. Four phases of digital transformation are presented, following which, business can receive a number of advantages without compromising corporate data integrity and security. The presented methodology complements and further develops the proposed hybrid architecture approach to streamline digital transformation in different economic domains.

Keywords: *Hybrid Integration Platforms, Legacy IT Systems, Cloud Connectors, Big Data, API Economy.*

INTRODUCTION

There are huge risks involved in modernizing and changing internal IT systems where enterprises have invested heavily in the past. On the other hand, legacy IT systems cannot be simply thrown away as business processes cannot be stopped until new more agile IT solutions based on modern technologies like PaaS (Platform as a Service) and APIs (Application Programming Interfaces) are conceived, developed and deployed. For the majority of the integration market stakeholders, finding and implementing appropriate Hybrid Integration Platform, can be a challenging task. Such platform must provide an effective solution to cross the bridge between old and new IT systems. The current market demand for flexible and fast IT integration must take advantage of robust stable legacy IT systems. HIPs provide an efficient path forward to resolve integration problems within complex software environment. A possible approach for this is with the help of APIs (or cloud) connectors facilitating data transportation.

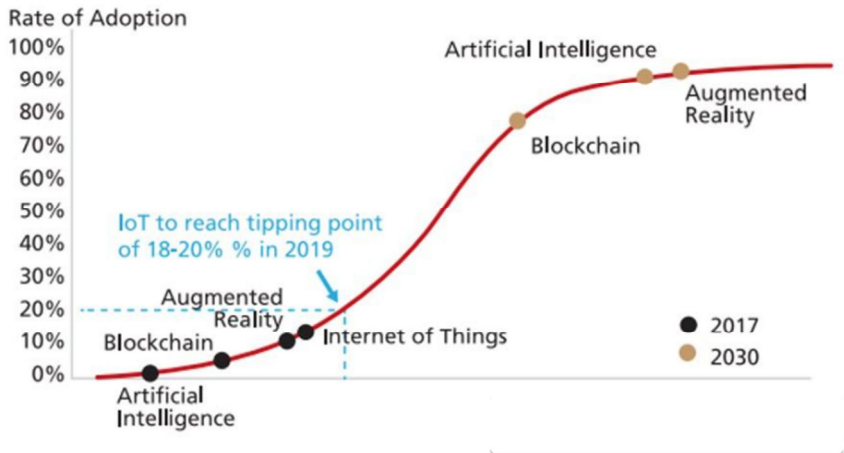
1. BUSINESS INTEGRATION DEVELOPMENT

The market on Hybrid Integration Platforms (HIPs) is still in development, and there are many players in the field, looking to capitalize on the prediction by Gartner (van der Meulen, 2018) that by 2022 at least 65% of large organizations will have implemented HIPs to power their digital transformation. There are different viewpoints and approaches on HIPs. One example is Ovum - an independent analyst and consultancy firm based in London with strengths in global coverage of Telecommunications, Media and Technology industries. It considers hybrid integration platform (Saurabh, 2019, p. 5) as *“a cohesive set of integration software (middleware) products enabling users to develop, secure and govern integration flows connecting diverse applications, systems, services and data stores, as well as enabling rapid API creation/composition and lifecycle management to meet the requirements of a range of hybrid integration use cases.”*. Axway Software (publicly held information technology company that provides software tools for enterprise software, Enterprise Application Integration, business activity monitoring, business analytics, mobile application development and web API management) views Hybrid Integration Platform (Camille, 2018) as *“on-premise solutions to seamlessly and securely integrate with cloud-based applications. This allows your IT team to fast-track innovation and enhance efficiency and lower the risk factors in accomplishing integration.”*

As one can see, different players in the field look in different way to HIPs, even that all they are focused on values for the companies that a fast and smooth integration can provide.

Historically, business driven integration projects used to be relatively easy to implement albeit complex endeavors. The reason for this is that IT systems used to belong to the same company and resided on-premises behind company network perimeter firewalls. So centralized decision-making process and management along with united efforts for all integration project activities, were carried out in the enterprises by designated internal teams. As shown in Figure 1 (Mittal et. al. 2018), today business needs are overwhelming IT teams with requests to move at ever-increasing pace and integrate constantly with new partners, new customers, new users and new devices.

IoT adoption to approach 100% over the next 10 years



IoT adoption gaining momentum

	2016	2017	2018	2030
IoT units installed base - total (m)	6,382	8,381	11,197	125,000
Consumer devices (m)	3,963	5,244	7,036	75,000
Consumer devices as a % of total devices	62%	63%	63%	60%
Connected devices per person	5	5	5	5
World population (m)	7,400	7,600	7,700	8,500
IoT adoption rate	11%	14%	18%	176%

Figure 1. IoT adoption fueling market demands for integration

Moreover, integration must happen across company borders and often includes on-premise-to-on-premise integration as well as cloud-to-cloud integration and, of course, mixed (or hybrid) integration. In other words, integration of different in build and nature IT systems that are distributed on-premise and in the cloud, be it a private or a public cloud. One possible classification of different integration approaches is given in Table 1 below:

Table 1

Different integration approaches

	On-premises integration	Native cloud integration	Hybrid integration
IT systems location	Majority of the IT systems are built and run on-premises	Majority of the IT systems are built and run in the cloud	There is a mix of IT systems built and run on-premises and in the cloud
Privacy and control	Very strong	Depends on external factors	Balanced and based on the business needs
Flexibility and opportunities	Very limited	Excellent	Very good

3. HYBRID INTEGRATION

On-premise (the more accurate term is on-premises) signifies that application software is installed and runs on servers located on company premises. This is a mature, well established and popular software deployment model since 1980s. Typically, at its core is the server-client software architecture.

Evolution of this architecture is distributed resource sharing, also known as cloud computing architecture. Cloud computing architecture is based on sharing computer system resources like storage, memory, network, processing power, etc. As a rule, those resources are located in big data centers and customers access them over the Internet. If the usage is limited to one enterprise cloud, it is called *private* or *enterprise* cloud. If cloud resources are shared across multiple entities, the term used is *public* cloud.

With the Internet of Things (IoT) wave, a number of integration and standardization problems emerged, even more, that the forecast for IoT growth is that it will continue in an exponential way. The pace at which new devices are added into the IoT ecosystem is phenomenal and unprecedented. The same applies for associated big data generation, transmission and consumption. There is, however, one great concern – security of this massive data streams, which is an issue still to be resolved.

Change is never an easy task. At the same time enterprises have to adapt swiftly, otherwise they can't stand to market competition to match the ever-increasing customer demands. It has happened before - the telecommunication giant Nokia's collapse due to its inability to act efficiently (David C., 2014), when smartphones entered the scene. The camera giant Kodak overlook the emerging digital photography and was forced to declare bankruptcy in 2012. Here comes the reason and objective to bring high attention to Hybrid Integration Platforms – they are aimed to prevent such scenarios. HIPs play a central role in the ongoing important and lifesaving digital transformation which enterprises are undertaking, abandoning the comfort of the well-known status quo, despite the fear of the changes ahead and the risks they bring. Successful implementation of this transition includes four basic transformation phases, outlined below:

- Analysis phase: The first phase must lay solid foundations for the whole transformation process. It must be executed carefully with an appropriate scope. All digital stakeholders affected by the integration process like enterprise legacy computer systems, business partners, end users and smart devices must be reviewed and considered.
- Planning phase: Following a successful completion of the analysis phase, the next logical step is careful planning of the hybrid integration. A detailed plan with respective ownership and success criteria is crucial.
- Implementation phase: During this phase the hybrid integration plan is executed. Some small adjustments to the plan are expected. If major changes of the plan are required, then it is better to return to the planning step to take into consideration gaps discovered during the implementation phase.
- Control and grow phase: The final phase is not the end of the transformation process. Continuous monitoring of all predefined parameters and key factors is ensuring that hybrid integration delivers intended business outcomes. Optimization and stabilization also play an important role during this period. Those are achieved by using positive loopbacks to optimize performance and ensure scalability and future growth.

Hybrid integration benefits include:

- Reduced costs – no need to change or replace old legacy IT systems
- Reduced implementation time – using API connectors

- Opportunities for new revenue streams – connecting with new partners

As shown in Figure 2 below, leveraging HIP unlocks valuable enterprise data and makes it available for business partners, end users, IoT devices, cloud apps, government and state agencies and many more. This truly is a compelling path forward into the emerging API Economy.

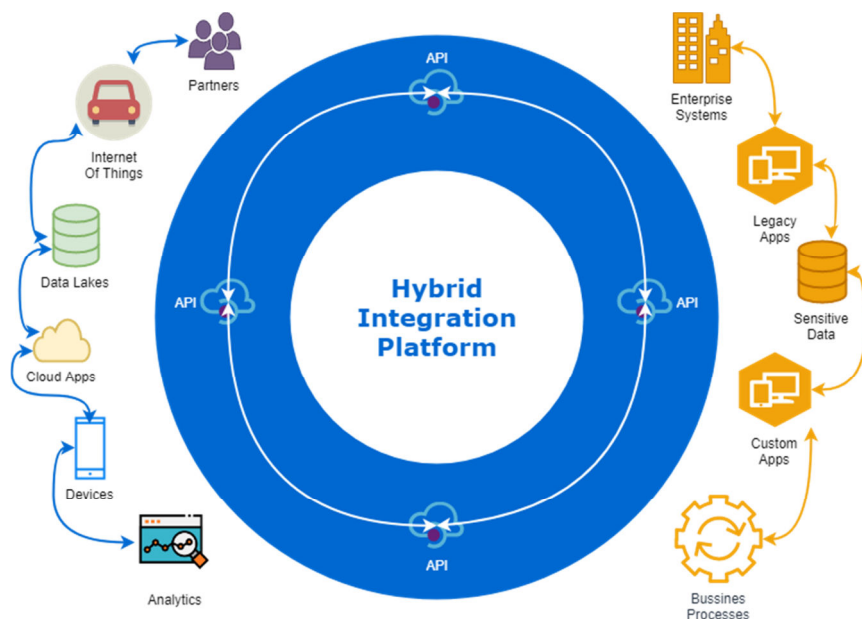


Figure 2. Hybrid Integration Platform Market

At UNWE, as part of a project for big data integration (Digitalization of Economy in a Big Data Environment), a work is carried on with Apache Hadoop cluster automation. One of the possible proposed solutions is to use Cloudera Manager RESTful API integration functionality that bears the following management options:

- Obtaining logs and monitoring the system – having the ability to automatically obtain the logs via API calls is crucial for system monitoring and automation. A test has been completed with comparison of Java and Python based API integrations. Python code is generally easier to understand and maintain.

- Starting and stopping services – ability to start and stop services without human interaction is very important for cluster scaling. Tests with Java and Python code are carried out. Both approaches are delivering comparable results as time needed to stop and start services.

- Polling cluster events – API provides also ability to poll specific cluster events monitoring. This is useful for triggering additional logic in Java and Python code when needed. For example, health check related events can be used for decision if new node should be started.

- Creating a disaster recovery replication schedule – this is the most complex scenario performed. API provides only basic functionality for replication and cannot be used instead of general replication procedures.

There is still a lot to be researched and learned before a complete analysis for API approach applicability for the project to be made.

CONCLUSION

The ability to implement continuous integration of complex software systems via API connectors and to automate the process is crucial in today's economy of hyper competitive markets. This paper provides an overview of what Hybrid Integration Platforms are and how HIPs can help to achieve this business goal and prevent enterprise scale crisis and decline. HIPs are not the only available solution, yet they present a solid framework that makes transition to continuous integration less painful and more efficient compared with other approaches.

ACKNOWLEDGEMENT

This study is partially supported by the Digitalization of Economy in a Big Data Environment Grant BG05M2OP001-1.002-0002-C02, funded by Operational Programme “Science and Education for Smart Growth” 2014-2020.

REFERENCES

1. CAMILLE, S. (2018) What is Hybrid Integration Platform? How to grow your company forward.
2. CORD, D. (2014) The Decline and Fall of Nokia. Schildts & Söderströms

3. MITTAL, S., WANG T. and KO C. (2018) Internet of Things the Pillar of Artificial Intelligence. DBS Asian Insights [Online] p. 64. Available from <https://www.dbs.com> [Accessed 8/17/2019]
4. SAURABH S. (2019), Hybrid Integration Platforms: Digital Business Calls for Integration Modernization and Greater Agility
5. Using the Cloudera Manager API for Cluster Automation [Online] Available from <https://www.cloudera.com> [Accessed 8/17/2019]
6. VAN DER MEULEN R., (2018) Use a Hybrid Integration Approach to Empower Digital Transformation

MAKING FINANCIAL DATA READABLE USING INLINE XBRL

Ivan Belev¹

¹ *UNWE, Sofia, Bulgaria, i.belev@unwe.bg*

Abstract

The paper aims to explore the features of Inline XBRL (Inline eXtensible Business Reporting Language) as a means to make financial and business reporting data human readable. The author researches and analyzes the official XBRL documentation and various sources to outline the differences between XBRL and Inline XBRL and give an overview of Inline XBRL and the technology behind it. The research covers the XBRL and Inline XBRL Standards and their comprising components as well as the way Inline XBRL is implemented. A simple Inline XBRL example is made that outlines the significance of this study – presenting the way Inline XBRL is implemented using a simple example.

Keywords: *XBRL, eXtensible Business Reporting Language, Inline XBRL, iXBRL.*

INTRODUCTION

The emergence and development of eXtensible Business Reporting Language (XBRL) revolutionized the way organizations exchange business and financial data. Before XBRL, companies were facing a difficult situation when regulatory organizations started to force short deadlines for reporting large volumes of financial data in specific data formats. Business and financial data could be reported and exchanged in a human readable format using different software tools and formats like Microsoft Word, Microsoft Excel, PDF or HTML reports. Regulators however were facing an even greater challenge – to collect, read, validate and analyse the data from all filing companies in a very short time and cope with all the different data formats.

XBRL offered a solution for both sides, making possible the digital exchange, validation and interpretation of financial data in a structured and unified data format. Like many other occasions, the solution to one problem led to another. The XBRL standard made it easy for machines to exchange and interpret the business data, but the format was almost

impossible for a human to understand without computer aid. The solution to that problem came with the creation of Inline XBRL standard, also referred to as iXBRL.

1. INLINE EXTENSIBLE BUSINESS REPORTING LANGUAGE (INLINE XBRL)

Inline XBRL, like XBRL, is an open standard managed by the XBRL International Consortium. Inline XBRL is also referred as iXBRL. iXBRL enables the structured XBRL data to be visualized in a human-readable presentation.

While XBRL uses eXtensible Markup Language (XML) to structure and tag the data, Inline XBRL uses the Hypertext Markup Language (HTML) standard to visualize the XBRL data by embedding additional tags into the HTML code so that the XBRL structure and data remains machine-readable but in the mean time is also human-readable in the form of an HTML report.

The first version of the Inline XBRL standard was developed in 2010 and published in 2011. In 2013, version 1.1 was published which is the last version of the specification at the time of this paper (September 2019).

The development of Inline XBRL originated from the requirements of the HM Revenue and Customs (HMRC) in the United Kingdom, requirements that later most of the regulatory organizations imposed. The reporting companies had to submit a human-readable form (usually in HTML or Microsoft Excel format) of the financial report along with the XBRL data. This required a tool for transformation of XBRL data into HTML or Excel formats and vice versa. In Insurance Supervision in Europe (carried out by the European Insurance and Occupational Pensions Authority - EIOPA) for example, such a tool was developed and provided only for the initial period of implementation of the XBRL reporting requirement – between 2015 and 2017. After the tool (called Tool for Undertaking – T4U) was discontinued, reporting companies and regulators had to provide their own implementation of such a tool. At the time of the report (September 2019) many companies are using their own implementation or third-party software to perform XBRL transformations.

With the development of Inline XBRL, some regulators (like the U.S. Securities and Exchange Commission - SEC for example) are pushing Inline XBRL to become mandatory. In the United Kingdom the iXBRL format is mandatory for filings under the HM Revenue and Customs (HMRC). The European Securities and Markets Authority (ESMA) also demands the adoption of a single data format for financial reporting, that is in the same time machine- and human-readable – called the European Single Electronic Format. On that ground all public companies in the European Union shall use Inline XBRL to report annual financial statements from the start of 2020. Other regulators in the United States and Europe have also planned mandatory iXBRL filings starting 2020 and 2021.

2. DIFFERENCE BETWEEN XBRL AND INLINE XBRL

An article about XBRL and Inline XBRL, published in the DataTracks blogs (DataTracks Article, 2012) summarizes the key differences between XBRL and iXBRL:

- The first and most notable difference is that while XBRL is only machine readable, Inline XBRL is machine and human readable at the same time;
- The file extension for XBRL files is either “.xml” or “. xbrl”. As for Inline XBRL – the file extension is “.html” or “.xhtml” because the XBRL tags are embedded in the HTML code;
- Similarly, the encoding in XBRL files is standard XML encoding while Inline XBRL files have XML and XHTML encoding;
- The output type of XBRL files is tabular in the form of XML code and the output type of iXBRL is a human readable HTML page;
- In terms of rendering the output, XBRL always requires an XBRL viewer to transform the XML code to a readable output as for Inline XBRL, it requires an XBRL viewer only if needed to render the XBRL layer of the data. Otherwise the Inline XBRL file can be opened via a web browser;
- XBRL files have limited formatting options while iXBRL files are very flexible in terms of formatting the way the result looks;
- Inline XBRL is more complex than XBRL in terms of the code itself, but the situation is the opposite if the focus is on the presentation of the data;

- The filing process for XBRL requires the filing company to prepare a separate human readable instance document as for iXBRL this is not required.

All those differences make Inline XBRL a very good option to replace XBRL reporting as it basically does the same job but adds additional key features.

3. ELEMENTS OF INLINE XBRL

According to the Inline XBRL Specification (Inline XBRL Specification 1.1), Inline XBRL is a standard for embedding XBRL fragments into an HTML document, ensuring that the XBRL tags are available for processing by XBRL parsing applications. The iXBRL specification defines all of the available components, elements and attributes as well as the constraints and validations for each of them. Some of the key elements, defined in the specification are:

- continuation element – defines data that should be treated as part of other elements – the footnote or the nonNumeric elements;
- exclude element – encapsulates data that should be excluded from the footnote or nonNumeric elements;
- footnote element – represents the link:footnote element from the XBRL Specification and references the XBRL facts;
- fraction element – defines an XBRL fact;
- header element – used to contain parts of the XBRL instance documents that are not displayed on the HTML page;
- hidden element – contains XBRL facts that should not be displayed on the HTML page;
- nonFraction element – defines an XBRL numeric item that cannot be represented by the Fraction element;
- nonNumeric element – describes an XBRL non-numeric item (usually a Text value, Boolean flag or date/time value);
- relationship element – used to define relationships between XBRL facts;
- resources elements – for specifying additional header resources;
- tuple element – represents the XBRL tuple, defined by the XBRL specification. Tuples as described by the XBRL specification represent complex multi-value facts and can include nested fact elements such as fraction, nonFraction, nonNumeric or other tuples.

In the Inline XBRL syntaxis all elements are written using the namespace prefix “ix:”.

In order for the Inline XBRL elements to be complete, the iXBRL Specification defines a set of attributes that can be assigned to different elements to enrich them with specific metadata. Although this report does not focus on the individual attributes, here is a list of available attributes as defined by the specification: arcrole, contextRef, continuationFrom, decimals, escape, footnoteRole, format, fromRefs, id, linkRole, name, precision, order, scale, sign, target, title, toRefs, tupleID, tupleRef, unitRef.

Another very important component that is used by the Inline XBRL standard is the Transformation Rules Registry that contains a set of transformation functions. The Transformation Rules Registry is specified by the Inline XBRL standard in a separate specification. The actual version of the specification (at the time of this article- September 2019) is Transformation Registry 3, published by the XBRL International Consortium in 2019. What the Transformation Rules Registry does is provide transformation functions that allow text strings in the Inline XBRL documents to be converted to the XBRL instance specific data types. The transformation register provide rules for handling the following datatype scenarios:

- Transformation to Boolean (true or false) values;
- Transformation to various date and time related values;
- Transformation to empty values;
- Transformation to various numeric values such as decimals, integers, etc.

Inline XBRL format is achieved by combining the Inline XBRL and the Transformation Rules Registry specifications. As a result, an XBRL instance can be reported in both machine and human readable format using a single HTML or xHTML file. The next section provides a simple example for Inline XBRL.

4. A SIMPLE INLINE XBRL EXAMPLE

This section provides a simple Inline XBRL example that aims to show how XBRL is embedded in HTML code. For this example, there are two XBRL facts that must be reported using Inline XBRL. For sim-

plicity of the example the facts do not include specific time period or entity information:

Table 1

Example data for Turnover and Cost, measured in BGN currency

Fact	Value
Turnover	7550
Cost	6225

Using plain HTML code this table can be rendered in a web browser with the following source code snippet:

```
<table>
  <tr><th>Fact</th><th>Value</th></tr>
  <tr>
    <td>Turnover</td>
    <td><b>7550</b></td>
  </tr>
  <tr>
    <td>Cost</td>
    <td><b>6225</b></td>
  </tr>
</table>
```

Figure 1. Plain HTML source code snippet for Turnover and Cost

Figure 2 represents the HTML output:

Fact	Value
Turnover	7550
Cost	6225

Figure 2. HTML output table

Using the Inline XBRL standard and syntax the same HTML output can be achieved with the code snippet, provided in Figure 3 with embedded XBRL data:

```

<table>
  <tr><th>Fact</th><th>Value</th></tr>
  <tr>
    <td>Turnover</td>
    <td><b><ix:nonFraction name="pt:TangibleFixedAssets"
contextRef="e2003" id="s2-1" precision="4" unitRef="BGN">
7550</ix:nonFraction></b></td>
  </tr>
  <tr>
    <td>Cost</td>
    <td><b><ix:nonFraction name="pt:IntangibleFixedAssets"
contextRef="e2003" precision="3" unitRef="BGN">
6225</ix:nonFraction></b></td>
  </tr>
</table>

```

Figure 3. XBRL enriched HTML source code snippet for Turnover and Cost

This code will produce the same HTML output table with no difference with the one, presented in Figure 2 when viewed in a web browser. The difference however is, that the example code in Figure 3 can be passed to an Inline XBRL processing engine that will extract the XBRL data from the HTML code. The resulting XBRL instance data is presented in Figure 4:

```

<pt:Turnover contextRef="e2003" id="s2-1" precision="4"
unitRef="BGN">7550</pt:Turnover>

<pt:Cost contextRef="e2003" precision="3"
unitRef="BGN">6225</pt:Cost>

```

Figure 4. XBRL instance facts, extracted from the HTML file

CONCLUSION

The adoption of eXtensive Business Reporting Language solves many problems related to the generation, exchange, verification and interpretation of financial and business data. This report tries to emphasise on the additional value that Inline XBRL can bring on top of XBRL and the problems that it solves. The simple Inline XBRL example shows

how iXBRL is embedded in HTML code to provide a single machine and human readable XBRL instance report.

REFERENCES

1. DataTracks Article (2012) What is iXBRL? Differences between XBRL and iXBRL. [Online] Available from: <https://www.datatracks.co.uk/ixbrl-blog/what-is-ixbrl/> [Accessed 10/09/2019].
2. European Insurance and Occupational Pensions Authority (EIOPA), XBRL Tool for Undertakings (T4U) [Online] Available from: <https://eiopa.europa.eu/regulation-supervision/insurance/tool-for-undertakings> [Accessed 10/09/2019].
3. SINHA, P. (2019), What is iXBRL? How is it different from XBRL? [Online] Available from: <https://iriscarbon.com/eu/what-is-ixbrl-how-is-it-different-from-xbrl/> [Accessed 10/09/2019].
4. U.S. Securities and Exchange Commission (SEC) Article (2018) Inline XBRL [Online] Available from: <https://www.sec.gov/structureddata/osd-inline-xbrl.html> [Accessed 10/09/2019].
5. XBRL International, Extensible Business Reporting Language (XBRL) 2.1 Specification, [Online] Available from: http://www.xbrl.org/specification/xbrl-recommendation-2003-12-31+corrected-errata-2008-07-02.htm#_4.11 [Accessed 01/09/2019].
6. XBRL International, Inline Extensible Business Reporting Language (Inline XBRL) 1.1 Specification, [Online] Available from: <http://www.xbrl.org/specification/inlinexbrl-part1/rec-2013-11-18/inlinexbrl-part1-rec-2013-11-18.html> [Accessed 10/09/2019].
7. XBRL International, Transformation Rules Registry 3 for Inline XBRL Specification, [Online] Available from: <https://www.xbrl.org/Specification/inlineXBRL-transformationRegistry/REC-2015-02-26+errata-2019-04-17/inlineXBRL-transformationRegistry-REC-2015-02-26+corrected-errata-2019-04-17.html> [Accessed 10/09/2019].

eXTENSIBLE BUSINESS REPORTING LANGUAGE (XBRL) – A TECHNICAL OVERVIEW

Ivan Belev¹

¹ UNWE, Sofia, Bulgaria, i.belev@unwe.bg

Abstract

The aim of this paper is to give a technical overview of eXtensible Business Reporting Language (XBRL) and to give simplified examples. It focuses on trying to explain what XBRL is and how it works. The author gives examples to explain the structure and the different elements of the XBRL standard. The scope of this research covers mainly the XBRL standard and its comprising elements. The author conveys research of various documents and sources and studies of the technical implementation of the XBRL standard. The significance of this report is that it explains the XBRL language and its elements using an easily understandable data sample.

Keywords: *XBRL, eXtensible Business Reporting Language, technical overview.*

INTRODUCTION

In today's modern digital world – data is everywhere. Organizations of various types and goals need to gather, prepare, validate, analyze, report and exchange different data daily or even more frequently. The exchanged data must be readable and understandable, not only by the human brain, but also readable and understandable by the computer system of the receiver of the information.

Throughout the years and nowadays as well, many data formats have been used and are being used for the exchange of business or financial data inside the organizations and between different organizations. Some examples are – printed paper reports, PDF reports, HTML reports (Hypertext Markup Language), spreadsheet reports (Microsoft Excel), structured file reports (CSV – comma separated values) and others as well. The data exchange is easily achievable in many ways and formats, but the problem is that the structure, format and datatype of the transferred data is not consistent and the receiving party needs to read and interpret the structure and format so that the data can be understood.

All this sets the stage for the invention of a more common way to describe and exchange financial and business data – through XBRL – eXtensive Business Reporting Language.

1. OVERVIEW OF EXTENSIBLE BUSINESS REPORTING LANGUAGE (XBRL)

The invention of XBRL is strongly connected with XML (Extensible Markup Language), since XBRL is an XML based language. The first version of the XML standard was published by the World Wide Web Consortium in 1998. The American accountant Charles Hoffman, like many others, was looking for a better way to create, share and distribute financial data. At that time all American companies were obliged to report financial data to SEC (U.S. Securities and Exchange Commission) in the form of paper-based financial statements. Later in 1999 the long-established Electronic Data Gathering, Analysis and Retrieval system (EDGAR) took a step forward and started accepting data reports in HTML, ASCII (American Standard Code for Information Interchange) and partially in PDF (Portable Document Format) format. In April 1998 Hoffman developed the first XML-based financial and audit statements. He then contacted other people and organizations, including Wayne Harding and later the developed prototypes made possible the creation of the first XBRL specification.

The current version of the XBRL specification is 2.1. It is considered a very stable version since it was last changed in 2013. XBRL is an open standard for creating and processing business reports digitally. The XBRL standard is managed by the XBRL International Consortium – a non-profit organization. XBRL is used in more than 50 countries around the world.

XBRL provides unique tags to be associated with business facts that are being reported. The structure and specific tags are described in a pre-defined public set of rules and definitions. This allows the preparation, validation, publication, exchange, consumption and analysis of the business data reports that are easily readable by both the sending and receiving party. The common rules and definitions are called a taxonomy and come in different packages and versions, depending on the context of the business data.

XBRL is used by:

- Regulators;
- Companies;
- Governments;
- Data Providers;
- Analysts and Investors;
- Accountants.

Some of the key features of XBRL are the predefined rules and definitions which are clear and public, the fact that the business rules can be tested and thus validated, the ability for the reporting entity to report the data in their own language and currency and more.

2. STRUCTURE AND ELEMENTS OF XBRL

The two most important components of the XBRL framework are the XBRL taxonomy and the XBRL instance.

The XBRL taxonomy defines the different concepts that are available to be reported using XBRL. It can be interpreted as a dictionary that is used to describe what business items can be reported in XBRL and the rules for reporting them. Different organizations can report all sorts of business data from different aspects – financial, accounting, insurance, risk, etc. For every different reporting need a separate taxonomy can be created that determines what can be reported for the particular use case. For example, insurance companies in Europe must report under the Solvency taxonomy (at the time of this report – Solvency II version 2.3.0). The taxonomy should be publicly available to all parties, that must report data according to it. Technically, the XBRL taxonomy is an XML Schema that includes definitions of all the different reporting terms, that can be reported using the taxonomy. Some examples for reporting terms are profit, assets, turnover, etc. The reporting terms are described as reporting concepts that also define additional attributes for the reporting of the concept, for example – data type, name, language, etc. It is important to note that the XBRL taxonomy does not include actual reporting data (facts). It only contains xml structured description of all the available terms that can be reported. Furthermore, a taxonomy can relate to other taxonomies that can extend or modify each other, forming a set that is called a DTS – Discoverable Taxonomy Set.

On the other hand, an XBRL instance contains the reported data (facts), based on one or more taxonomies (DTS). Every reported fact is identified using references to the taxonomy concepts. The XBRL instance should always reference the taxonomy schema that is being used. Each XBRL instance can include the following main elements:

- Contexts;
- Units;
- Items;
- Tuples.

The context element describes the reporting entity (company or organization), period and scenario. Each context element should contain an identification attribute (id) that is used so the context can be referenced by other elements, such as items. Periods are expressed by either a point in time (called instant), a time interval (start and end date) or the element forever (lack of specific time or period). The entity element identifies the reporting entity, described by the fact using an identifier. The scenario element is optional and is used to define business scenarios such as actual data, budget data or other types of scenarios, based on the taxonomy.

The unit elements are used to specify the measuring units for numeric items, for example – meters, miles or specific currency. Units are referenced by other elements using their id attribute.

The item elements and the tuple elements are used to represent the reported facts or values. When the facts are simple the item element is used. When the facts are compound or complex – the tuple element is used, containing other items and/or tuples. Items and tuples reference other XBRL elements such as contexts and units to determine additional details about the reported fact.

Other notable element of XBRL are the footnotes. When multiple facts share a common relationship that is somewhat irregularly structured, they can be referenced by a footnote element.

3. AN XBRL EXAMPLE

The author provides a simple example that showcases the concept of XBRL reporting. The example uses sample data for household monthly expenses for one fictional household family – the Whites family. First the data is shown in a table in a very basic representation:

Table 1

Example data for household monthly expenses – basic table representation

Month	Expense Type	Expense Value
July 2019	Food	700 BGN
	Transportation	350 BGN
	Utilities	200 BGN
	Entertainment	250 BGN
	Other	450 BGN
August 2019	Food	650 BGN
	Transportation	500 BGN
	Utilities	130 BGN
	Entertainment	1500 BGN
	Other	350 BGN

In order to represent the data in XBRL format, first it is shown in a more detailed table:

Table 2

Example data for household monthly expenses – detailed table representation

Household Family	Period	Expense Type	Expense Value	Expense Currency
Whites Family	2019-07-31	Food	700	BGN
Whites Family	2019-07-31	Transportation	350	BGN
Whites Family	2019-07-31	Utilities	200	BGN
Whites Family	2019-07-31	Entertainment	250	BGN
Whites Family	2019-07-31	Other	450	BGN
Whites Family	2019-08-31	Food	650	BGN
Whites Family	2019-08-31	Transportation	500	BGN
Whites Family	2019-08-31	Utilities	130	BGN
Whites Family	2019-08-31	Entertainment	1500	BGN
Whites Family	2019-08-31	Other	350	BGN

The detailed table includes a column for the household family name, separate rows for each period which is described using the last day of the month and separate column for the currency. This representation of the data is similar to what is shown in the XBRL example below.

The first key element that is required to represent the data in XBRL format is the taxonomy. As it is stated above, the taxonomy describes the different concepts that can be represented in the particular case. For the household family expenses example, the available concepts (facts) are the different types of household expenses – Food, Transportation, Utilities, Entertainment, Other. Each of them is represented using a monetary value in a specific currency that is assigned to a specific period. The taxonomy elements, used to describe a couple of the expense types are shown in the following code snippet:

```
<!--XBRL Taxonomy Elements-->
<element name="Food"
  type="xbrli:monetaryType"
  substitutionGroup="xbrli:item"
  xbrli:periodType="instant"/>

<element name="Transportation"
  type="xbrli:monetaryType"
  substitutionGroup="xbrli:item"
  xbrli:periodType="instant"/>
```

Figure 1. XBRL taxonomy elements for representing expense types

The example code shows only two of the expense types, but the representation of the other three is similar. Since every expense is a simple fact, it should be reported using XBRL item elements in the XBRL instance. The period type should be “instant” because every expense is assigned to a single point in time – the month period.

The actual expenses (facts) of the Whites household family, represented in Table 1 and Table 2 are reported using an XBRL instance. The XBRL instance references the taxonomy that was partially described in Figure 1. For clarification, three different parts of the XBRL instance are shown in separate figures. XBRL contexts are described first:

```

<!--XBRL Instance Contexts-->
<context id="Whites_July_2019">
  <entity>
    <identifier scheme="http://my.example">Whites Family</identifier>
  </entity>
  <period>
    <instant>2019-07-31</instant>
  </period>
</context>

<context id="Whites_August_2019">
  <entity>
    <identifier scheme="http://my.example">Whites Family</identifier>
  </entity>
  <period>
    <instant>2019-08-31</instant>
  </period>
</context>

```

Figure 2. XBRL taxonomy instance example – XBRL contexts

Two XBRL contexts are described in the XBRL instance example in Figure 2. Each context represents an instance of a reporting period that is also assigned to the entity – the Whites household family. XBRL entity identifiers should conform to a referenced XML scheme but in the provided example such scheme is non-existent.

The next figure (Figure 3) represents another required element of the XBRL instance – the XBRL units:

```

<!--XBRL Instance Units-->
<unit id="U-Monetary">
  <measure>iso4217:BGN</measure>
</unit>

```

Figure 3. XBRL taxonomy instance example – XBRL units

Only one XBRL unit is defined because the example includes household expenses in a single currency – BGN.

The last remaining part of the XBRL instance – the XBRL facts, represented by XBRL item elements is described in Figure 4:

```

<!--XBRL Instance Items (Facts)-->
<Expense:Food contextRef='Whites_July_2019' unitRef='U-Monetary' decimals='INF'>700</Expense:Food>
<Expense:Transportation contextRef='Whites_July_2019' unitRef='U-Monetary' decimals='INF'>350</Expense:Transportation>
<Expense:Utilities contextRef='Whites_July_2019' unitRef='U-Monetary' decimals='INF'>200</Expense:Utilities>
<Expense:Entertainment contextRef='Whites_July_2019' unitRef='U-Monetary' decimals='INF'>250</Expense:Entertainment>
<Expense:Other contextRef='Whites_July_2019' unitRef='U-Monetary' decimals='INF'>450</Expense:Other>

<Expense:Food contextRef='Whites_August_2019' unitRef='U-Monetary' decimals='INF'>650</Expense:Food>
<Expense:Transportation contextRef='Whites_August_2019' unitRef='U-Monetary' decimals='INF'>500</Expense:Transportation>
<Expense:Utilities contextRef='Whites_August_2019' unitRef='U-Monetary' decimals='INF'>130</Expense:Utilities>
<Expense:Entertainment contextRef='Whites_August_2019' unitRef='U-Monetary' decimals='INF'>1500</Expense:Entertainment>
<Expense:Other contextRef='Whites_August_2019' unitRef='U-Monetary' decimals='INF'>350</Expense:Other>

```

Figure 4. XBRL taxonomy instance example – XBRL items (facts)

The actual household expenses are reported using XBRL items. An item is reported for each combination of reporting period and expense type. Notably the item elements are named “Expense”, a tag that is not described in the taxonomy. This description is not required by the XBRL standard and the reporting instance can use any tag name. Each item references a taxonomy context (a period for the expense) and an instance unit of measure for the values. The XBRL instance items include another attribute – the decimals attribute with the value “INF”. The decimals attribute is used to determine the number of decimal places for the reported numeric value. In the provided example the value “INF” is used, which states that the number of decimal places for accuracy can vary and should be determined by the actual reported value.

All three parts of the XBRL instance together with the XBRL taxonomy provide a way to represent the data in a structured and unified way, that can be easily and unambiguously interpreted by anyone, who has access to the instance and the taxonomy package.

CONCLUSION

The eXtensive Business Reporting Language emerges with the need for a unified solution for describing complex financial and business data. XBRL evolves quickly to become a worldwide adopted standard used by large-scale organizations such as insurers, banks, governments and other organizations. XBRL makes possible the digital transformation of the organizations in terms of regulatory reporting and financial and business data exchange. This report shows a simple example of the way XBRL is used. However, the standard includes many other elements and features that make possible the representation of very complex data structures, as well as components for data validation and data visualization.

REFERENCES

1. HANNON, N. (2016) Accountant's Quick Guide to XBRL. [Online] Available from: <https://www.datatracks.com/blog/an-accountants-quick-guide-to-xbrl-part-i/> [Accessed 01/09/2019].
2. HOFFMAN, C., WATSON, L. (2009) XBRL For Dummies (1st Edition), Wiley: For Dummies
3. XBRL International, Extensible Business Reporting Language (XBRL) 2.1 Specification, [Online] Available from: http://www.xbrl.org/specification/xbrl-recommendation-2003-12-31+corrected-errata-2008-07-02.htm#_4.11 [Accessed 01/09/2019].

ENTREPRENEURSHIP IN THE EDUCATION OF UNIVERSITY STUDENTS

Anna Sobczak¹

*¹ Akademia im. Jakuba z Paradyża, Gorzów Wlkp., Poland,
a.sobczak06@gmail.com*

Abstract

Currently, academic entrepreneurship significantly contributes to the competitiveness of the national economy by connecting and strengthening cooperation between research centres and enterprises. The article discusses the results of research conducted on a sample of students regarding their entrepreneurial attitudes. This article presents the possible factors that affect entrepreneurship among respondents and starting their own business as well as the basic barriers to entrepreneurship. In a knowledge-based economy that faces many challenges, entrepreneurship can be an important determinant of economic, technological and social development.

Keywords: *entrepreneurship, academic entrepreneurship, entrepreneurship education, economic education, student entrepreneurship.*

INTRODUCTION

In the face of globalisation changes, the functioning of enterprises and organisations is becoming increasingly difficult. If we add geopolitics to this, the conditions in which organisations must now function are difficult. The pursuit of competitive advantage by individual states or groups of states is enormous, because only then one can affect the functioning of the global economy. Countries that gain an advantage become players who set trends and directions. Innovation, knowledge as well as entrepreneurship can guarantee the domestic economy an advantage over other economies. One of the first people to use the subject of entrepreneurship was J.B. Say, according to whom, the entrepreneur was a person transferring capital from an area with lower profitability to an area with a higher level of efficiency and profit, investing their resources and bearing the risk associated with the decisions made (Say, 1960, p. 550). Entrepreneurship is recognised as one of the key competences. Its definition is ambiguous and educating it may involve forming a variety of skills.

The concept of entrepreneurship may refer to human personality traits, the process of setting up a business and the ability to take advantage of opportunities in the environment and to take innovative actions (Borowiec-Gabryś, Kilar 2018). According to Piecuch, entrepreneurship is simply the art of dealing with various life situations, which can be learned and trained (Piecuch, 2010). Thus, education plays an important role in the development of entrepreneurship. According to J. Timmons and H. Stevenson, entrepreneurship is a process of creating or recognising opportunities and using them regardless of their resources. They emphasize in their works that an entrepreneur is the one who initiates and builds an organization not limiting himself to its description and analysis (Timmons, 1990, p. 5). According to the classic approach of P. Drucker, entrepreneurship should be understood as a specific approach to management consisting in "introducing innovations forming the foundations of future business", with "the best use of available resources" (Drucker, 1992, p. 36). In addition, "the entrepreneur is always looking for change, reacts to it and uses it as an opportunity" (Drucker, 1992, pp. 36-37). There is therefore a strong link between innovation and entrepreneurship. J. Schumpeter has already pointed to this relationship. In turn, according to J. Timmons, entrepreneurship is "a way of thinking, reasoning and acting, focused on searching for opportunities", as well as "a pursuit of opportunities without taking into account the limitations created by currently controlled resources" (Timmons, 1999, p. 27).

Entrepreneurship education in Poland already takes place at the high school level and is continued at the higher level. The introduction of entrepreneurship in secondary schools was intended to get young people familiar with business. Propagators of this direction wanted to familiarise young people with establishing and running a business. The very idea of this was right, but the problem arose with the implementation of such a noble goal. To begin with, the program of the subject, which is overloaded with content and has no reflection in practice, raises concerns. Another issue is that teachers place more emphasis on knowledge and less on skills. Teachers educating the subject mostly do not run or have never run a business. In addition, teachers of the subject - entrepreneurship - also teach other subjects,

which means that entrepreneurship is complimentary to the full-time job. Considering the above, young people who undertake university studies have acquired knowledge in the field of entrepreneurship, but in many cases they can not practically apply it.

In relation to the above, there is a need to shape entrepreneurial attitudes, in particular the skills to set up and develop a business. Entrepreneurship in Poland (but not only) is usually defined at the level of the small and medium enterprise sector. The emerging and existing small and medium-sized enterprises influence the development of many macroeconomic factors: GDP, international exchange or employment: the small and medium-sized enterprise sector employs about 70% of employees of all enterprises in Poland and many countries. Academic entrepreneurship is an important aspect that is highlighted by various organisations and institutions implementing programs that are designed to improve the competitiveness of the Polish economy. The growing emphasis on the development of the knowledge-based economy has led to the search for solutions that are to increase the level of cooperation between science and business units. Academic entrepreneurship is to be one of the development directions enabling this rapprochement. In the development of academic entrepreneurship, regardless of the meaning of the concept itself, students and their entrepreneurial attitudes play a large role. The main purpose of the article is to present the results of research on entrepreneurial attitudes of students, which covered individuals studying in non-economic and economic faculties. It was important to verify that the completed course could affect these attitudes.

1. ENTREPRENEURSHIP EDUCATION IN STUDENT'S OPINION

Teaching entrepreneurship as one of the most important skills students should be equipped with should be constantly monitored by universities and evaluated. Skills that help students open their own business should be one of the main goals of the university, because it depends on how entrepreneurial and creative the management staff of future enterprises will be. And what follows, what companies will build the economy of the country.

In 2016, the study programs are analysed in four Visegrad countries: the Czech Republic, Poland, Slovakia and Hungary. A total of

188 study programs were reviewed (Egerova, 2016). Study programs in economic and other faculties (technical, artistic, humanities) were examined. The results of the analysis show the university's approach to shaping entrepreneurial attitudes, in particular the preparation to set up and run own business.

From the data in Table 1, it results that about 40% of fields of study have a subject related to entrepreneurship in their study programs. They prepare students to open their own business and, in addition, they have the word entrepreneurship in their names. In Poland, the share of programs with at least one subject directly related to entrepreneurship is the lowest and amounts to 31%. Besides, the studied fields of study also provide subjects that can be associated with entrepreneurship because they relate to issues related to the functioning of the enterprise. These are, for example, subjects such as management, marketing, accounting, finance, business plan, negotiations. Subjects from this group are several times more common. In total, subjects directly and indirectly related to entrepreneurship were included in the study programs of 93% of the studied fields of study. (Nowaczyk, Sobczak, 2019)

Table 1

Fields of study with at least 1 subject directly related to entrepreneurship and without subjects directly and indirectly related to entrepreneurship

Country	Number of analysed fields of study	% of fields of study with subject/s directly related to entrepreneurship	% of fields of study without subjects related to entrepreneurship
Czech	59	39	9
Hungary	36	44	6
Poland	42	31	7
Slovakia	51	43	6
Total	188	39	7

Source: based on *Entrepreneurship education. Opportunities and Challenges for Universities in Visegrad Countries*, ed. by D.Egerova, NAVA, Pilzno 2016, tab. 4.3 i 4.5, s. 58-59.

In Poland, the studied fields of study had a total of 13 subjects related to entrepreneurship education and other 162, which are useful in teaching how businesses operate. Subjects directly related to entrepreneurship, having the word entrepreneurship in their name, included, among others, knowledge of legal regulations related to conducting business activities, sources of supporting entrepreneurship, registering business activities, creating a business plan and economic analysis (Nowiński, Nowaczyk, Sobczak, Tomczyk, Fabiś 2016).

In addition to subjects educating entrepreneurial competence, entrepreneurship faculties and specialties are run at Polish universities. At some universities there is the field of Entrepreneurship, e.g. at the University of Silesia in Katowice. An example of a specialty in this area is the Entrepreneurship specialty in a small and medium company in Master's degree management studies at the Faculty of Management of the Poznań University of Economics (Nowiński, Nowaczyk, Sobczak, Tomczyk, Fabiś 2016).

As a result of focus studies conducted periodically (from 2016) among students of economic and non-economic faculties (the same number of students take part in each faculty each year) the impact of teaching entrepreneurship to students on developing entrepreneurial attitudes and supporting entrepreneurial initiatives taken at the regional, national or global level by students was diagnosed. The collected research results have allowed to indicate the changes that universities should take in order to better prepare young people for entrepreneurship.

Students of economic faculties emphasised that entrepreneurship education is necessary, especially in economic studies, because it is a group of students who will in the future deal with broadly understood business. The formula of teaching the subject has not changed over the years and students of "practical entrepreneurship" are still not being taught. In contrast, students of non-economic faculties mostly would like to have entrepreneurship classes, although there was some who found it unnecessary. In all editions of the study, students who already have companies indicated that entrepreneurship classes should be compulsory, because only in this way one is able to encourage new people to take their own individual initiatives and motivate to greater

economic activity. In the opinion of this group of students, entrepreneurship education should be conducted at every level of education, and in terms of economic studies, it should be compulsory and subject to the special care of the university and the state, because its level reflects the students' decision to open their own business. Students stressed that entrepreneurship should be a course implemented in every field of study, because the market economy requires representatives of each field to be able to think of their field as a business. This new look at entrepreneurship shows that young people with business experience recognise the essence of entrepreneurship and its creation among representatives of all fields of study.

People who are currently running their own businesses claimed that by attending entrepreneurship classes and basing only on them they would not make the decision to start their own business. In their opinion, the classes only provide information and do not allow one to practice the whole process. The information obtained during the classes was, in their opinion, too theoretical and did not indicate how, step by step, having an idea to start a business. Some respondents also indicated that the number of hours on the subject was insufficient for such exercises to be carried out. In addition, young people stated that the knowledge needed to run their own business is dispersed on several or even a dozen or so subjects, but no one combines all the information (Nowiński, Nowaczyk, Sobczak, Tomczyk, Fabiś 2016).

Despite the growing economic development, fewer students plan to set up a company, and the number of undecided people and those who do not want to run their own business has increased. Easier access to a large labour market can hinder the risk of running a business. Having said that, in smaller cities, own business is often the only solution not to be unemployed.

During the research, students emphasised that a lot of attention was paid to showing the topic of starting their own company in all education. The knowledge they receive during classes, although it is often supported by examples, turns out for students to be insufficient to know what steps should be taken to open a business. Students indicated that after completing the course in entrepreneurship they were rather afraid to

start their own business, as the classes also pointed to the difficulties and problems they might encounter on their way. Often, these factors and fear of the unknown prevented them from opening a business. Another factor that discouraged the establishment was the lack of any support from the institution or the state in this respect. The respondents often indicated that they did not know which offices to go to, how to fill in the documents, where and whom to ask for advice and what legal form to choose. Another problem is planning activities, analysing the competition's offer, product or service marketing activities, or all matters related to the personnel and financial part. Students were afraid that in case of questions or problems they would not know where to look for help. Their fears were, moreover, justified, because the examples from life that their fellows gave, who were already running their business, confirmed this. Everyone - students of economics and non-economics faced problems and lack of support from institutions and organisations.

The research has shown that students feel entrepreneurial. Positive responses were given by 65% of female students and 70% of last year' students. A 2016 survey found that 55% of female students and 64% of male students are or rather feel entrepreneurial. Therefore, although the participants in the current study felt more entrepreneurial, this did not translate into the intentions of starting their own business. Most students, 89%, consider themselves entrepreneurial, who are mainly characterised by risk-taking skills, ingenuity as well as creativity and initiative. Approximately 47% of the surveyed population did not work during the studies related to the chosen field of study. This proves that the main problem is the unwillingness to build your career path after starting studies in a given field and the lack of ideas for starting and type of business.

Students running their companies indicated that when setting up a company they used the help of others - external companies or people who had already followed this path and could be a signpost and help for them. Most pointed to family help in this area. Students stated that entrepreneurship education should be run by universities and lecturers were indicated as those who should guide, motivate students and strengthen them in the pursuit of establishing their own businesses.

According to the respondents, it is very important to know how to run own business and the skills to run it. Students emphasised that during classes they received knowledge about starting a business and a lot of details related to it, but most indicated that there were no practical classes related to running a company. The surveyed students of economic studies pointed to a number of program content related to finance, cooperation with people or the law, which in their opinion were relevant from the point of view of doing business. Non-economic students complained about the lack of knowledge in this area. Students realise that to run a business and develop it, knowledge and management skills are needed. Non-economic students indicated that they had ideas for their own business, but the finance-related part must be left to specialists in this area. All students of non-economic faculties claimed that by completing the economic faculty it was easier to set up their own business. In turn, students of economic faculties claimed that only students in the field of management were prepared to conduct business, because in their studies they had had the widest range of subjects.

At the same time, all students emphasised that success on the market depends not only on knowledge and the idea for a business, but on creativity, planning skills, time management, teamwork skills, coping with stress and risk taking. Students of all faculties indicated that they rarely practised these skills during their classes. Students of all faculties highly assess the arranged meetings with interesting people during their studies, which show the business path of these people. The problems, challenges or even resistance and difficulties they had to face. As they emphasised, it is more valuable to meet a man who talks about running a business, showing its dark sides and ways to overcome these crises than telling only about the positives of doing business. For students, student internships, company visits, meetings with successful people or people who run their companies and want to share their practices and experience play a very important role. Such a practical aspect of education is perceived by students as an opportunity to listen, watch and learn from others who have practical knowledge and teach how to deal with specific situations.

A group of 74% of students after graduation are not ready or planning to open their company. Among the arguments against setting up own enterprise the lack of adequate financial resources was most often indicated - students who want to take out a loan to open a business accounted for 4% of all respondents. Another factor is fear of the unknown and "easiness" - it is better to work for someone and not have worries, do the work and have time off. The respondents also pointed out that family and friends (who usually work full-time) advise against opening own business claiming that it is very difficult and there is no guarantee of success. Students admitted that they were too young to run their own businesses, because they wanted to have time to relax and socialise.

The attitudes presented here point to an irrational search for explanations and a reluctance to undertake a new initiative. Looking for negatives can be analysed as an explanation for some kind of laziness or an easy way. Thinking about themselves on the labour market, students see themselves as employees rather than as employers. This situation confirms the general mood of reluctance to start your own business. In running their own business, young people are attracted by money and the possibility of flexible working time as values. However, they realise that making enough and satisfying amounts of money can be stretched over time.

For people who already run a business, the "flexible working time" factor meant 24/7 work. This group of respondents eagerly spoke and discussed arguments for and against opening their own business, willingly referred to their own problems and shared their experience. Most often, they were encouraged to run their own business, the opportunity to decide about themselves, build something lasting for their family, flexible working hours and finances. It is optimistic to say people who currently help their relatives in running a business - they want to start their own business in the future. Currently, they want to gain experience working for someone else or save money for their own business.

Some students whose relatives or friends conduct business in the future count on their help in opening their own business. Everyone

emphasised that if they meet everyday business people - satisfied and motivated, this attitude also motivates them. Fulfilled entrepreneurs infect others with their enthusiasm and encourage them to pursue their own plans and dreams of business. Students indicated that they talk to and observe people running their own businesses and derive strength and enthusiasm for starting their own business from these experiences.

It seems that combining all the suggestions would give the best results in terms of entrepreneurship learning. If the respondents had the opportunity to combine theoretical knowledge with practice and activities supported by meetings with entrepreneurs, they would be more willing to take the initiative to open their own business. Unfortunately, in reality, it is difficult to encourage good managers or business owners to find time to share their experiences with students. Such people complain about the lack of time and thus it is difficult for them to arrange time for this type of activity.

Opinions on ideas for improving entrepreneurship education at universities and awakening entrepreneurial spirit among students were divided. Some pointed to an increase in the number of exercises performed independently and making decisions so as to discuss later results in class. Others indicated study visits to companies and institutions as well as meetings with entrepreneurs who could talk about running their own business along with the opportunity to ask questions (Nowaczyk, Sobczak 2016).

The respondents who do not run their own businesses and asked if they had enough information to run the company replied that they did not. They usually search for information on opening a business on the Internet. Few (1%) indicated that when setting up their business they would contact their lecturers from the university.

CONCLUSIONS

A very small percentage of young people undertake to open their business. Young people are afraid to open a business because it is more convenient to work for others. For young people, this is one of the most important arguments when making decisions. Those who have no problem with new technologies and entering new industries are more willing to take up the challenge and open their business. As research

shows, entrepreneurship and the ability to start a business are influenced by the nature of the fields of study, shaped competences during studies, teaching methods used, as well as additional activities enhancing the development of entrepreneurship, but the most important is the attitude to independence. Entrepreneurship education in higher education should be adapted to the level, profile and field of education. In economic faculties, entrepreneurship education should focus on building competences related to setting up a business, while in non-economic faculties entrepreneurship competences should include skills to create and use technological or artistic ideas to a greater extent.

REFERENCES

1. BOROWIEC-GABRYŚ, M., KILAR, W., RACHWAŁ, T. (2018), *Entrepreneurship as a competence of the future* [issue:] Competences of the future, S.M.Kwiatkowski (ed.), Warsaw: FRSE publisher, p. 74.
2. DRUCKER, P. (1992). *Innovation and entrepreneurship. The practice and principles*. Warsaw: PWE.
3. *Entrepreneurship education. Opportunities and Challenges for Universities in Visegrad Countries*, ed. by D. EGEROVA (2016), Pilzno: NAVA, p. 56-57.
4. NOWACZYK, G., SOBCZAK, A., *Role of entrepreneurship education in higher education institutions in the process of business development in Poland*, [issue:] Entrepreneurship: Antecedents and Effects, ed. by M.Kosała, M.Urbaniak, A. Żur, „International Entrepreneurship” no. 2/2016, p. 43-52.
5. NOWACZYK, G., SOBCZAK, A., *Entrepreneurship in educating students, adults and seniors*, [issue:] Developing entrepreneurial competencies, ed. by T.Rachwał, FRSE, Warsaw 2019.
6. NOWIŃSKI, W., NOWACZYK, G., SOBCZAK, A., TOMCZYK, Ł., FABIŚ, A., (2016), *Innovative entrepreneurship education, National report*, Poznań/Oświęcim/Kraków, Poznań School of Banking, State University of Applied Sciences in Oświęcim, Pedagogical University of Cracow, p. 21-27, 31-38.
7. PIECUCH, T. (2010). *Entrepreneurship. Theoretical fundamentals*. Warsaw: C.H. Beck publisher.

8. SAY J-B., (1960), *A treatise on political economy*. Warsaw: PWN.
9. SCHUMPETER J., (1960). *Theory of economic development*. Warsaw: PWN.
10. TIMMONS J. 1990. *New Venture Creation*. Boston: Irvin.
11. TIMMONS, J. (1999). *New Venture Creation. Entrepreneurship for the 21st Century*. 5th Edition: Irwin/McGraw-Hill.

E-COMMERCE IN THE PROCESS OF DIGITAL TRANSFORMATION: PREPAREDNESS, REQUIREMENTS AND TECHNOLOGICAL TENDENCIES

Mihaela Markova¹

*¹ Dimitar A. Tsenov, Academy of Economy, Svishtov,
mihaela.markova@gmail.com*

Abstract

Business digitalization is a widely discussed topic nowadays that is relevant to all aspects of economic and administrative operations. The rapid development of information and communication technology has a considerable impact on the automatization of several business activities, especially trade.

E-commerce is facing constant changes whilst developing, which ensures higher flexibility when attracting new clients. Choosing the right technology to develop an online platform could significantly contribute to simplifying the trade process, which on the other hand positively influences client experience and boosts sales numbers. The purpose of this paper is to analyze the state of e-commerce in the process of digital transformation. An online survey is conducted among some Bulgarian companies. Some of the conclusions of the survey are addressed to businessmen who own online shops.

Keywords: *digitalization, digital transformation, e-commerce, innovation, social networks.*

INTRODUCTION

The usage of digital transformation as a term has been spreading across social practice. Digitalisation is making its way into numerous areas, driven by the rapid development of information and communication technology and the overall process modernisation. The most widely spread innovations such as Cloud computing, Internet of things, Big Data, Business Intelligence, Social Networks define its application in each business sphere. Digitalisation is a step towards a radical change in the operating cycle of an organization – management, production, marketing and planning.

1. DIGITAL TRANSFORMATION OF BUSINESS

With the rapid development of information and communication technology, a new concept emerges, namely digital economy. Digital

economy is based on electronic data exchange. Its focus lies not only on high tech innovations, but also on structural and process changes within a company. A research (Zimmerman, 2000) shows that managing a business, being led by digitalization, could be analysed as a process connected with using the net and its services to support communication with both clients and rival companies, marketing and commercial channels.

Digitalisation improves the world and this can be most clearly observed in the field of data processing. In traditional business, data is usually hard to collect and even harder to store and use. The introduction of information systems has made data management significantly easier. Nowadays data can be processed independent of time and distance. Cloud technologies are a typical example, being an innovative tool for data storage and data management. Another example is the fact that a lot of companies support servers with information in foreign countries, which provides them with the opportunity to cut cost in terms of expensive hardware and additional employees.

Information and communication technologies change the way innovations are being introduced to business processes. Normally an innovation requires resources for testing and implementation. Nowadays electronic means have made this process a lot easier, irrespective of whether the innovation concerns an information process or the creation of a new product. Different technologies contribute to a more successful introduction to the market of all sorts of innovative products.

To continue, digitalization also creates new client communication channels and ways to introduce different products and ideas. A company can have a solid market position obtained through high quality and fair prices in many countries, independent of its location. This assures high customer satisfaction, which has proven to be one of the best marketing strategies. In addition, social media has become a powerful tool when it comes to brand advertising.

Technology changes customer value creation. Any change in value and quality – aspects that customers consider the most – can be noticed by rival companies and used to their advantage. Such dynamic changes in the market create the need to closely follow customer behavior and interests. In particular, the following aspects should be studied before

introducing a particular product: customer requirements, so that they can be met exactly, and the trade process, so that profitability increases.

2. E-COMMERCE – PREPAREDNESS, REQUIREMENTS, TECHNOLOGICAL TRENDS

E-commerce is a crucial part of digitalisation. It can exist either independently, or as a result of the interplay between traditional and digital business. Technically it could be argued that e-commerce lies the foundation of digitalisation, since it is based on the automatisisation of different processes. Its emergence and successful implementation lead to changes in every aspect of business – retail chains, exclusive stores, airlines. E-commerce is also a typical example for transformation – from creating an online shop to selling and delivering to the client.

An empirical survey (Markova M., 2019) that concentrates on the advantages of the right choice of technology for an online shop explains that it is crucial whether the business is prepared to become a part of the digital world. It is based on small, middle and large entities with different business activity. The participants have been chosen randomly based on a mail list. The applied method is online survey that has been e-mailed to 584 companies, of which 11% have responded. The survey has been conducted within 4 months in 2019. It contained 50 closed questions on the process of adaptation of business to the new market reality and what consequences digitalisation has for business. Since completion of all questions was mandatory, all answers were valid. However, the sample lacks external validity – it is only informative about the interviewed firms.

The main thesis is that transformation of business in organisations supposes the creation of e-commerce applications that need to be as effective as possible and correspond to the resource availability. The main hypotheses are based on the proportion of firms that have a website, what it is used for, how many of the websites are also online shops and what technologies they use. The results show that 80% have an already developed website, 17% are currently developing one and the rest do not support a website at all.

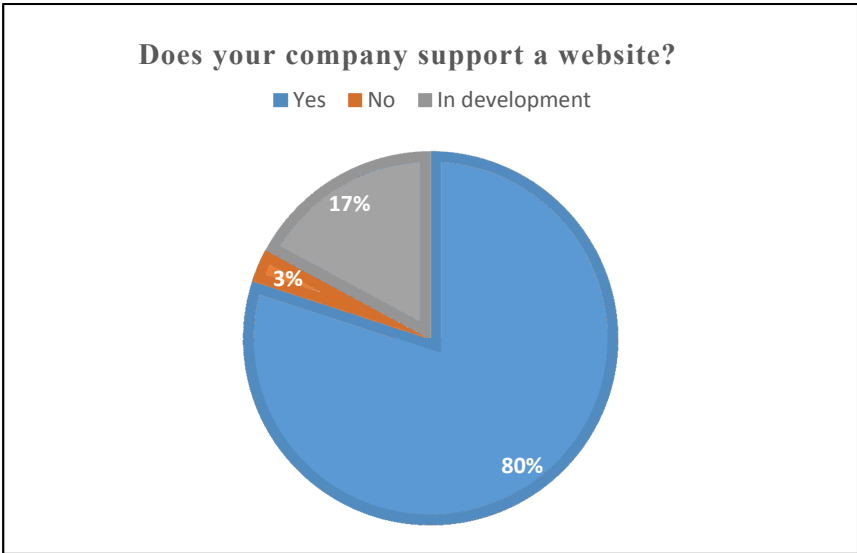


Figure 1. Companies which support a website

It is noticeable that only a very small part of the investigated firms does not support a website, which could be explained with the fact that having a website increases profitability and this is a key element of business development. Moreover, it makes a company visible to the world independent of time and location. However, simply having a website is not enough, what it is used for and its strategic contribution are far more important:

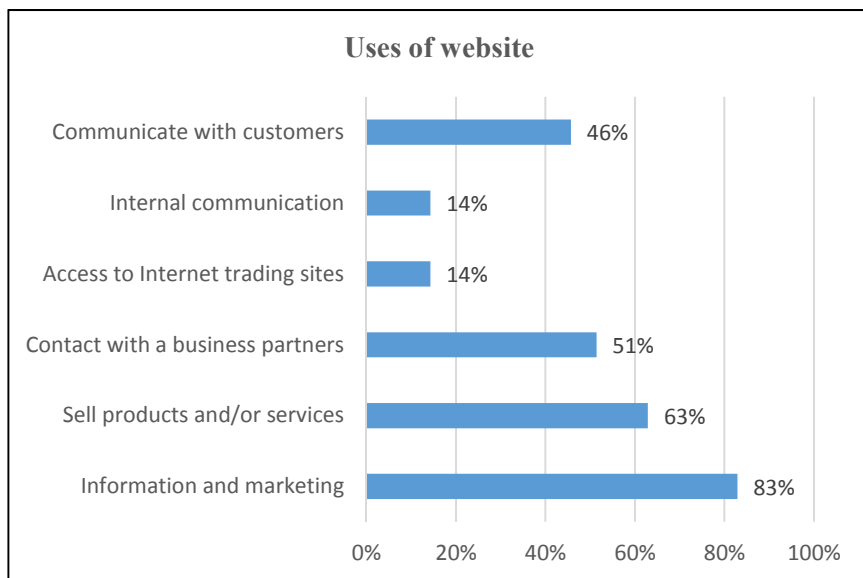


Figure 2. Type of website uses

The biggest part of the investigated firms use their website to provide information about their business (83%) and to sell products and/or services (63%). These results show that more and more companies digitalise their sales and marketing processes. The most up-to-date website technologies allow adding different functionalities that increase website attendance. Such tool is e.g. the SEO optimization which ensures that websites are shown first when using search engines, generates fast links to social networks and shortcuts to additional resources. A lot of firms also communicate with business partners through their website (51%) or their customers (46%), probably through different CRM systems that can be integrated into the website.

The choice of particular technologies when creating an online shop can be based on different criteria - for example, the presence of an IT department that has the capacity to create and support such platforms. A substantial part of the firms, however, mostly count on a few specialists that support the firm hardware and install and manage applications and software, but do not have the professional background to conduct a pro-

ject regarding digital transformation of trade processes. In this case many firms outsource the development of web-based and mobile applications.

There are different technological tools when it comes to creating a suitable application to sell products. The type of online shop is defined through the chosen platform for its creation. There are several platforms that can be used:

- based on a particular web design technology;
 - based on a platform with remote hosting (e.g. Amazon, Shopify)
 - based on an e-commerce platform (OpenCart, Magento, Prestashop, Gombashop, WooCommerce, SpreeCommerce, ZenCart)
 - integrated into social networks (Facebook)
 - based on native web programming languages – e.g. HTML, PHP, CSS, JavaScript, Java.
- A study in the field shows the following results regarding the preparedness of business for digital transformation among small, medium and large entities according to the presence of an online shop:

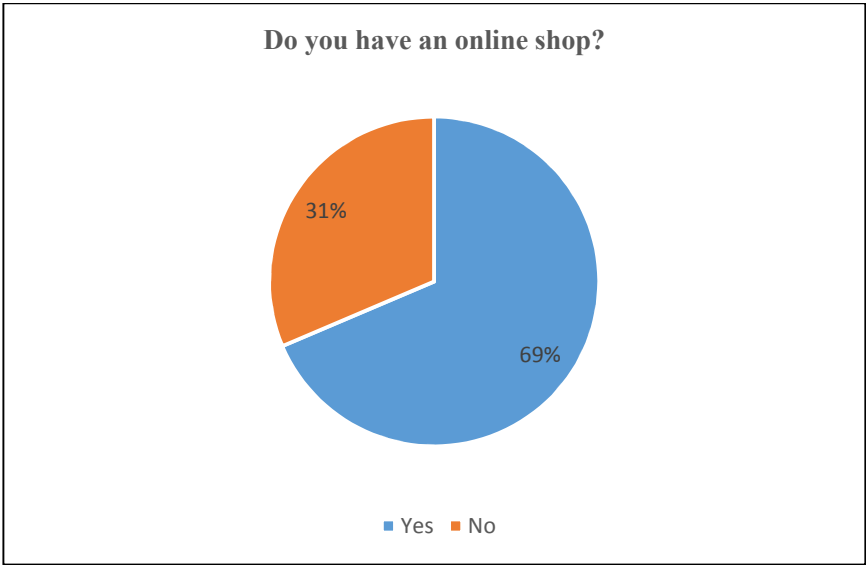


Figure 3. Presence of online shop

The diagram shows that a significant number of companies (69%) support a website. A smaller part (31%), on the contrary, do not support a website. Those are firms that do not develop a business activity that requires a website or such that are still not prepared to digitalise theirs. A part of the study analyses the platforms used when creating a website:

Table 1

Platforms for the development of an online shop

Platform for the development of an online shop	Percentage of firms using this platform
Prestashop	3%
Gombashop	3%
Opencart	11%
WooCommerce	11%
Joomshopping	3%
Magento	3%
DotNet Duke	3%
Own platform	6%

Online shop support is easier when platforms that provide basic modules are used. The usage of platforms is advantageous because a small initial investment is required to create and support an online shop, there are also extra modules and a specific hosting of the firm, where the owner can easily introduce changes to the modules. Such platforms make creating an online shop faster and provide different templates and the opportunity to change and adapt the online shop to the identity of the particular brand so that the shop corresponds to market expectations.

There are also a lot of different Add-ons that make the online shop functional. One of the most significant advantages is that specific theoretical background in web programming is required. Creating an online shop this way is easy and fast and could be taken on by the IT department of the firm or outsourced to an extern company. The study shows that 69% of the investigated firms outsource the hosting of a website, 46% - its development and 14% - its subsequent support. This means that not all firms engage the same teams that developed the website to also carry out its sup-

port afterwards – they would rather bring in their own specialists. 17% of the investigated firms outsource network infrastructure management and 14% do not outsource such services at all:

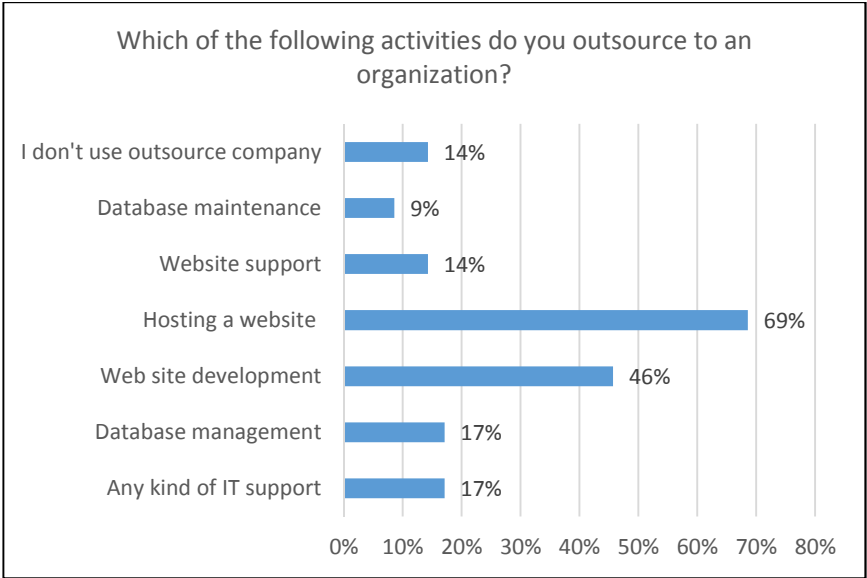


Figure 4. Most used outsource services

E-commerce changes the market behavior of rival companies and creates a lot of methods for meeting customers’ demands. It is more and more widely accepted that a company needs more than simply an online shop, e.g. a page on social networks. This is called Omni channel communication – reaching the customer through different offline and online channels, including the combination of an online and a “real” shop.

A study regarding the attitude of firms towards business digitalisation shows that 74% think that the online shops should be complemented by a mobile application. The rest (26%) feel that a Mobile first design of the online shop is enough. A Mobile first design assures that web pages are shown with their original interface on all mobile devices, irrespective of the particular model or operating system:



Figure 5. Needs of mobile application

CONCLUSION

Digitalisation provides business with new opportunities for better organization and operations management. The search for easier ways for customers to sell and buy goods is one of the main reasons for creating different forms of online shopping. With technologies developing, online shopping has become a crucial element of business digitalisation – from the first electronic transaction until today each and every process of business activity has been digitalised.

The study reaches the conclusion that business develops towards digitalisation. E-commerce is becoming more and more mobile device-oriented and being on the Internet is a must - otherwise a significant part of the customers could be lost. As a consequence, the omni-channel approach has been established, which gives business the opportunity to happen anywhere - through a website and on social media.

Another result of the study shows that carrying out e-commerce processes successfully and effectively depends on the businessmen's ability to apply the right strategy regarding their presence on channels that are popular among their customers.

REFERENCES

1. MARKOVA, M., SHISHMANOV, K., TASHKOVA, M., (2019) Current trends in the creation of e-commerce application, Publishing house “Academy of economics”, Svishtov (work in press)
2. ZIMMERMAN, H., (2000) Understanding the Digital Economy: Challenges for new business models, AMCIS 2000

BASIC SOFT COMPUTING METHODS IN USER PROFILE MODELING

Petya Petrova¹

*¹ University of Economics – Varna/Department of Informatics,
Bulgaria, petya.p.petrova@ue-varna.bg*

Abstract

User profile modeling is a challenge because of the high degree of subjectivity and uncertainty of human behavior. The traditional methods used to create user profile models do not have the necessary flexibility to capture the inherent uncertainty. The purpose of this research is to present adequate methods for modeling a user profile in their role as a learner. The soft computing methods - neural networks, fuzzy logic, fuzzy clustering, neuro-fuzzy approaches and genetic algorithms – applied individually or in combination with other machine learning methods could be used for this purpose, due to the appropriate specificity of each one of them. The results of a research on the suitability of the basic soft computing methods for modeling a learner's profile are presented in this paper.

Keywords: *Soft computing; User profile modeling; Learner profile model; Student profile model.*

INTRODUCTION

The main challenge in creating a profile model is due to the relevance and precision of the information received from users. The factors that influence the creation of an adequate consumer model are data precision, application of the right methods, the level of data noise and, above all, the degree of subjectivity and uncertainty of human behavior. A user model is created through a user modeling process in which the unobservable user information is derived from the observable information from the same user; for example, using interactions with the system (Zukerman, Albrecht and Nicholson, 2001; Frias-Martinez *et al.*, 2005).

The opportunities for creating user profile models in the current academic work are specifically selected to provide an alternative to creating a personalized student profile model. Further in the current document the term "learner profile model" will be associated with "user

profile model". They are classified on the basis of how human behavior is represented as a model and what its purpose is. This largely depends on the type of task the model will be used for. To choose the right model, we stick to the following five types of tasks: (1) *Forecasting*. The ability to anticipate the users need using their past behavior; (2) *Recommendation*. The ability to offer interesting facts to the user based on additional information that is not deduced from their previous behavior (Frias-Martinez *et al.*, 2005); (3) *Classification*. The process of building a model in which data is grouped or classified into one of many predefined classes (Potey and Sinha, 2015); (4) *Clustering*. Adaptation with focus on a group of users. Groups with specific characteristics are identified and clusters are formed on this basis. This is achieved through clustering techniques so as to cluster user models since the group is viewed as a cluster of similar user models. (Nguyen, 2014); (5) *Filtering*. Selection of data that meets user preferences, derived from the whole available user information. Each of the five tasks can be solved using the knowledge contained in the user profile model. For example, consumer data can be *filtered* based on the knowledge of their purposes and interests and *classified* based on predefined classes into stereotypes. Stereotypes can be *grouped into clusters* similar to the model. From the resulting data a *forecast* can be made about the preferences of the user. Finally, *recommendation* can be made that is a result of the knowledge gathered and the information analyzed.

Creating a user profile model can be achieved in two ways: 1) *classic*, where information is provided directly by the user, through a survey, and 2) *automatic*, in which information is collected using an algorithm embedded in the system for creating a user-model which is an implicit process for the user. The second way is a more modern solution and does not involve users answering questions. Machine learning methods can be applied for automatic user profile modeling, using techniques for recognizing patterns in the user behavior, recording them, and integrating them into a profile. The results which the machine learning methods provide are a summary of what has been learned and a description of the knowledge gained. These facts can be used to analyze the original data and make forecasts. From this point of view, data mining and other machine learning methods make it possible to create

user profile models automatically. Data Mining (DM) scientists provide an overview of how traditional data mining techniques can be applied for creating a user model (and in particular – a learner's one) and the overall architecture of such systems (Eirinaki and Vazirgiannis, 2003; Atanasova et al., 2019). Romero et al. (2013) propose a model for predicting student achievement using DM techniques such as grouping rules and associative rules. DM is also widely used to model a student profile whose primary purpose is to characterize the student through their emotions, achievements, skills, learning preferences, and the fulfillment of individual educational requirements by adapting the teaching experience to match them better. (Lemmerich, Ifland and Puppe, 2011; Nandeshwar, Menzies and Nelson, 2011). Another area of DM application is student assessment, which allows students' skills to be differentiated at a subtle level (Lopez et al., 2012). This also facilitates student feedback and support (Leong, Lee and Mak, 2012). More generally, DM can be applied for modeling the student profile regarding their emotions in the context, engagement, meta-knowledge, and collaboration tasks (Baker, 2014; Alexandrova and Parusheva, 2017).

However, the traditional methods such as machine learning and DM cannot solve the problem of the uncertainty in the user behavior. Soft computing methods family are designed to overcome this drawback. They are suitable for solving problems and tasks in conditions of inaccurate, unclear or noisy information; information for which accurate data is not available or is based on partial truth (Gupta, Sinha and Zadeh, 2000). The idea of soft computing originates in 1991 when a scientist working in the field tried to create a new kind of artificial intelligence. In March the same year the Berkley Initiative on Soft Computing (BISC) was created at the Berkeley Industrial Liason Program conference. According to BISC, software computing consists of fuzzy logic, neural network, evolutionary computing, etc. These methods are very suitable for solving personalization tasks. Soft computing methods could be used to perform research on the basis of qualitative metrics, so that indeterminable data (such as perceptions, sensations, assimilation, success rate) can be compared with real metrics resulting in good decisions in a given situation. Below we will look at the application of precisely such methods that use soft-computing means to model fuzzy uncertainties

and, in particular, to look for personalization solutions when creating user profile models.

The purpose of this research is to present adequate methods for modeling a user profile in their role as a learner. Basic soft computing methods are considered due to their appropriate specificity. They can be used individually or in combination with other machine learning methods to meet the set goal.

1. SOFT COMPUTING METHODS FOR USER PROFILE MODELING

1.1. Fuzzy Logic (FL)

Fuzzy logic has the ability to capture unstructured or ambiguous information and to model it under conditions of uncertainty (Klir and Yuan, 1995). Unlike traditional methods, FL stems from fuzzy set theory which deals with reasoning that is more approximate than precisely derived from classical predicate logic (Yan, Power and Ryan, 1994). It is this specificity that enables FL to capture the uncertainty inherent in real data. Not only the students' personal interests and behavior should be taken into account when building a student's profile, but also the way teachers work, in order to maintain the balance between them and to achieve a higher level of results based on the student's achievements. In this regard, Yadav R. et al. (2011) propose a fuzzy expert system (NFES) that distinguishes between a high IQ student and a low IQ student. Its main purpose is to modify learning according to each student's pace. NFES has the capacity to make decisions, monitor the growth of each student and determine the next step in their learning (Pal, 2014). The Student Learning Assessment Model was created with the help of FL to evaluate students' academic achievements by gathering information and providing results for comparison with statistical value (Singh Yadav and Singh, 2011). Researchers also focus on evaluating teacher's performance because teachers are at the heart of student achievement (Moran, 2015). Through the use of FL, educators can assess a teacher more efficiently than the current procedures. A number of researchers have used fuzzy time series to predict and forecast student enrollments in higher education institutions (Singh, 2015; Jilani, Burney and Ardil,

2019). FL is also used for filtering. Fuzzy-filtering scientists offer a way to match filtered data to the consumer interests and preferences as a method of user personalization (Vrettos and Stafylopatis, 2001). Although FL is not a machine learning technique, due to its proven ability to capture uncertainty in humans and adapt it to the task of personalization, it is permissible to use collaborations between soft computing and machine learning methods when creating user models. Some examples of these combinations are Fuzzy Clusters (see section 1.4.), Fuzzy Association Rules, and Fuzzy Bayesian Networks. Through the Bayesian Networks, a model has been proposed to discover the different styles of student learning, as each student has their own learning behavior. The results are more accurate and stable over the time compared with the traditional methods. This model helps to model student's profiles with high accuracy (García *et al.*, 2007). It is also effective to capture the teacher's gaps in assessment in order to avoid making the same mistakes and to identify practices that will be useful in the future (Xenos, 2004).

Although the FL method is well suited for user profile modeling, it faces some challenges in real-world applications. The main one is that it does not have a data learning mechanism. This means that the knowledge about the field of application must be explicitly given by the designer. Neuro-Fuzzy Methods which will be discussed in another section (see section 1.5.), are emerging as an approach to alleviate this challenging situation.

1.2. Neural Networks (NN)

The neural network is an information processing approach that is inspired by the way the biological nervous system and the brain process information (Fausett, 1994). A key element of this paradigm is how information is processed. It consists of a large number of highly interconnected processing elements (neurons) working in unison to solve specific tasks (Haykin, 1998). Soft computing methods, in particular neural networks, can help to predict the content of elements in complex nonlinear cases (Tarasov *et al.*, 2018). NN can make sense from complex and/or inaccurate data or from models that are too complex to be explored by other computational methods. Another benefit is that they do not need initial knowledge of the problem being addressed. These

characteristics make the NN a method for modeling human behavior and a suitable tool for creating user profile models. NN have been widely used to create user models, mainly for classification and recommendations, so that users with similar characteristics can be grouped together and create profiles to made. Bidel, Lemoine, and Piat (2003) provide some examples of classifying user models with NN, and Sas, Reilly, and O'Hare (2003) recommend using the NN method for predicting consumer behavior in a virtual environment. Due to the ability of NN to capture any type of knowledge, they are also used for filtering and forecasting tasks. The neural network can be used to provide forecasts that present new situations and answer questions. Beck, Jia, Sison, and Mostow (2003) model student behavior for an intelligent learning system.

Although the number of successful applications created with NN is significant, this method still has limitations in creating user profile models. The main obstacles are the training time it takes to create a model (in some cases, it can take hours or days) and the set of required data. Training time can cause inconvenience in dynamic model preparation. Although there are techniques that can re-qualify NN, when creating custom models it is usually re-qualified from scratch, ie. when there is more information, a new user or a new document, this is added to the system. Another disadvantage of NN is the limited interpretability of their solutions. While other techniques, to varying degrees, can be interpreted and manually modified, the representation of knowledge in the NN is not easy to interpret. As a result, their application is avoided when user-friendly models are needed.

1.3. Genetic and evolutionary algorithms (GA and EA)

Genetic and evolutionary algorithms are algorithms that include search and iterative optimization of the solution. They are based on the mechanism of natural selection in order to find the best possible algorithm for solving a task. GA can be seen as an evolutionary process and helps the exchange of the newly-discovered knowledge in the form of genes between the individuals, as well as a mutation, thereby helping to recover lost data or explore unexplored regions in the desired space of solutions. They start with a set of potential solutions - population-forming

chromosomes. They use the solutions of one population to form a new one that is closer to the optimal solution of the problem under consideration (Potey and Sinha, 2015). They can be used to build effective classification systems, algorithms for finding the associative rules, and other similar tasks related to data collection. Their reliable search technique has given them a central position in data mining and machine learning. The approaches have been used primarily for recommendations that can capture consumer goals and preferences in the form of rules. Examples of this approach are given by Romero, Ventura, and de Bra (2003) for student modeling, to capture user preferences for improving web searches. They have also been applied for filtering and classification as well as creating models which predict student's achievements. In the student's profile modeling, the primary role is played by the assessment of the individual student. This process can be analyzed through GA and helps to identify the most important factor affecting student's performance - their total achievement (Xing *et al.*, 2015). Teacher assessments also include student's evaluations that have been recorded throughout their academic careers. This process deals separately with evaluations obtained on different grounds - theoretical, practical and mathematical, which leads to more accurate results (Miranda Lakshmi, Martin and Prasanna Venkatesan, 2013). Genetic and evolutionary algorithms are more efficient than machine learning and data mining methods. This is because GA perform global searches and thus handle attribute interactions more successfully, whereas traditional methods use local search.

The method considered is suitable for searching in large and complex problem spaces but this creates a precondition for limitations both in its computing power and in relation to its potential for dynamic modeling and interoperable complexity.

1.4. Fuzzy Clusters (FC)

Fuzzy clustering has been intended as a method by which unstructured and unclassified information is structured into clusters based on any kinds of similarities. The clustering algorithm finds a set of concepts that cover all possible examples. One of the most widely used fuzzy clustering algorithms is the Fuzzy K-Means algorithm (Bezdek,

1981). It has been used to evaluate students' performance in relation to the entire academic career of teachers. The effectiveness of student's learning is based on the teaching methods of the teacher. This model helps to refine some of the limitations of the traditional model. Therefore, the K-Means clustering algorithm is also a suitable tool for analyzing and monitoring changes in student's work (Oyelade, Oladipupo and Obagbuwa, 2010). There are two types of clusters for creating user models: usage clusters and surfing clusters (Internet behavior). The users grouping has a tendency to create clusters of users that exhibit similar surfing patterns. In user modeling, FC are primarily used for recommendations and classification tasks. This is useful when personalized content is needed, such as demographic data or consumer interests.

The main problem faced by cluster techniques is how to define the concept of distance. In order to achieve this, thorough knowledge of the researched subject area is required. When the method is applied for modeling a user's profile, this task gets even more difficult because of the nature of the data: user preferences and interests, interactions and user behavior, success and progress when it concerns students, etc. On the other hand, this data may not be available digitally which is another limitation. The previously developed user profile models based on FC do not completely have the fuzzy character of the method. This is because the goal is more flexible and adaptable systems to be created and this means finding more meaningful ways to collaborate between different features and customization techniques, and grouping them into clusters.

1.5. Neuro-Fuzzy Method (NFM)

The neuro-fuzzy method is a combination between NN and FL. It uses NN for training and FL for generating fuzzy conclusions. It applies a process of fine-tuning of the membership rules and features of the inputs and outputs derived from the NN which are used in user profile modeling. This approach overcomes the disadvantages of NN and FL. NFM automate the process of transferring expertise or knowledge based on fuzzy rules (Jang, Sun and Mizutani, 1997). These methods are particularly suitable for applications where user interaction is desirable when designing or interpreting the model. Basically, FL-based systems

include an automatic learning process provided by the NN. Along with fuzzy logic methods, NN have also been used for evaluation purposes. An Adaptive Neuro-Fuzzy Interference System (ANFIS) was developed to improve the speed, reliability and flexibility that were lacking in previous systems (Taylan and Karagözoğlu, 2009). This combination of NN and FL can give accurate numerical results to predict student's performance. By evaluating student's performance, it is easier to distinguish between the ones that are at risk. On the other hand, providing students with correct counseling and training will increase the graduation rate (Herzog, 2006).

The combination of NN with FL forms an appropriate method for creating user profile models. Drigas *et al.* (2004) propose an application of the recommendation task by offering job positions to unemployed people, using (as a basis) the data from the user profile and comparing it with the enterprise profile. Another group of researchers use the approach to solve the classification and recommendation problem by planning the web content of a course according to the student's level of knowledge (Magoulas, Papanikolaou and Grigoriadou, 2001). Although NFM are designed to combine the positives aspects of NN and FL, they still maintain some of their limitations. These mainly concern the time required for training and application of the dynamic modeling.

2. SUMMARY AND SELECTION OF METHOD FOR USER PROFILE CREATION

What has been presented so far demonstrates that each method has its own specific opportunities for creating a user-model depending on what it will be used for. The main advantage of soft computing methods is that they have the ability to capture connections and relationships from implicit data and model them for the purposes of the task at hand. Based on existing real data, soft computing methods are able to capture user plans, goals, and preferences. The knowledge gathered can be applied to model their future behavior and build a user profile model based on their general characteristics under the format of data structures. The models created will be able to use the ability of soft computing to mix different behaviors and capture decision-making processes in order to implement a personalized system that is more flexible and reasonable in terms of user interests.

The different methods offer different opportunities and can be used for different purposes, such as: FL - to derive goals and plans through a mechanism for imitation of the human decisions; NN- to present general characteristics resulting from the derived information and to create a stereotype of the user on this basis; FC - for grouping users into more than one stereotype at a time; NFM - to capture expert knowledge and tune it in order to make assumptions for the user. Soft-computing methods can be applied to build a profile model individually or in combination with traditional machine learning methods.

In Table 1, we summarize the characteristics of the methods presented in six dimensions.

Table 1

**Characteristics of the basic soft computing methods
for creating a user model**

	Computer complexity	Dynamic modeling	Data quantity	Uncertainty	Noise	Interpretability
<i>Fuzzy Logic</i>	Medium	Yes	Missing	Yes	Yes	Low
<i>Neural Networks</i>	High	Yes	Large	Yes	Yes	High
<i>Genetic/Evolutionary Algorithms</i>	High	No	Missing	No	Yes	High
<i>Fuzzy Clusters</i>	Medium to high	No	Medium to large	Yes	Yes	High
<i>Neuro-Fuzzy Methods</i>	Medium to high	Yes	Medium to large	Yes	Yes	Medium to high

The first three dimensions cover the major issues that are involved in machine learning for creating a user model according to Webb, Pazzani and Billsus (2001). Namely: 1) Degree of computer complexity of the applicable method; 2) Dynamic modeling - whether the method allows the user profile model to be adjusted or changed during the working process; 3) Quality of the training data – it reflects the amount of data needed to create a reliable profile model. The next three dimensions are characteristic for soft computing methods: 4) Uncertainty - whether the methods can handle and process the uncertainty of human behavior when modeling a user profile; 5) Noise -

The ability to process noisy data, ie. whether the methods can handle and process the data noise; 6) Interpretability - the level of clarity of the results generated by the method, i.e. the extent to which the captured knowledge is easily interpretable by humans. Machine learning and DM researchers consider interpretability as a common problem and critical point in creating user models (Kim and Street, 2004).

The conclusion that can be drawn from Table 1 is that FL is a dynamic and flexible method but also moderately complex and difficult to interpret if used alone. NN have many advantages - they are easy to understand, process noisy data and data generated under uncertainty but require a lot of processing time. In contrast, NFM and FC hybrid approaches have a medium to high degree of computational complexity, have the ability to handle medium to large arrays of noisy data and uncertainty and the end result is easily interpretable. GA and EA prove to be the most inflexible and complex methods.

Table 2 presents the classification of soft computing methods in terms of possible interpretation requirements, addressing the five main types of tasks introduced at the beginning of this paper.

Table 2

Soft computing methods and requirements for interpretability of various tasks

Task	Required	Not required
Forecasting	Neuro-Fuzzy Methods; Genetic Algorithms.	Neural Networks
Recommendation	Neuro-Fuzzy Methods; Fuzzy Logic.	Neural Networks; Evolutionary Algorithms; Fuzzy Clusters.
Classification	Neuro-Fuzzy Methods	Neural Networks; Fuzzy Clusters.
Clustering	-	Fuzzy Clusters; Neural Networks.
Filtering	Fuzzy Logic; Genetic Algorithms.	Neural Networks; Evolutionary Algorithms.

Two possible interpretability values are considered: *Required* or *Not required*. They show whether the nature of the results obtained from

the solution of the problem through the methods of soft computing are understandable (does not require explanation) or not (require explanation). They also show in which cases and under what methods additional actions and specialized knowledge in the field are required.

Both tables can be used in combination to facilitate the choice of method when creating a profile model. The process of choosing may start with an overview of Table 2 which indicates the clarity of the methods depending on the purpose of the task being solved and the results achieved. Then Table 1 should be used to make the final decision about which method or combination of methods will facilitate the creation of the model.

CONCLUSIONS

With each passing day consumers, incl. and students, expect more intelligent and personalized services. The key point in providing such services is to capture user expectations, attitudes and interests and, on their basis, to create a personal profile for each user which can be automatically upgraded and can plan consumer behavior in the future. Adequate creation of a user profile model depends on the implementation of adequate methods. Machine learning and DM methods are suitable for automatically creating and identifying profiles and interests. But their main disadvantage is that they lack the ability to capture and process the inherent uncertainty of human behavior. To address this aspect, soft computing methods are proposed as an adequate tool for creating personalized user models.

This review shows that one of the major problems with designing a user profile model is the lack of any type of standardization. In order to improve this situation, the current document proposes a set of guidelines that will form the framework for the design of user profiles based on soft computing approach.

Soft computing methods in combination with machine learning methods would achieve effective collaboration in modeling the natural complexity of human behavior and have great potential to create a user profile model.

REFERENCES

1. ALEXANDROVA, Y. and PARUSHEVA, S. (2017) Social Media in Higher Education from Students' Perspective. doi: 10.5593/sgemsocial2017/14/S04.092.
2. ATANASOVA, T. et al. (2019) Monographic library " Knowledge and business ", book 4 Intelligent Analysis of Students Dataset Publishing house " Knowledge and business " Varna.
3. BAKER, R. (2014) 'Educational Data Mining: An Advance for Intelligent Systems in Education', *Intelligent Systems, IEEE*, 29, pp. 78–82. doi: 10.1109/MIS.2014.42.
4. BECK, J. E. et al. (2003) 'Predicting student help-request behavior in an intelligent tutor for reading', *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2702, pp. 303–312.
5. BEZDEK, J. (1981) Pattern Recognition With Fuzzy Objective Function Algorithms, *Pattern Recognition with Fuzzy Objective Function Algorithms*. doi: 10.1007/978-1-4757-0450-1.
6. BIDEL, S. et al. (2003) 'Statistical machine learning for tracking hypermedia user behavior', *2nd Workshop Machine Learning, Information Retrieval, and User Modeling*, 9th Int. Conf. User Modeling, Pittsburgh, PA.
7. DRIGAS, A. et al. (2004) 'An expert system for job matching of the unemployed', *Expert Syst. Appl.*, 26, pp. 217–224. doi: 10.1016/S0957-4174(03)00136-2.
8. EIRINAKI, M. and VAZIRGIANNIS, M. (2003) 'Web mining for Web personalization', *ACM Trans. Internet Techn.*, 3, pp. 1–27. doi: 10.1145/643477.643478.
9. FAUSETT, L. (ed.) (1994) *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
10. FRIAS-MARTINEZ, E. et al. (2005) 'Modeling human behavior in user-adaptive systems: Recent advances using soft computing techniques', *Expert Systems with Applications*, 29(2), pp. 320–329. doi: 10.1016/j.eswa.2005.04.005.
11. GARCHA, P. et al. (2007) 'Evaluating Bayesian networks' precision for detecting students' learning styles', *Computers & Education*, 49, pp. 794–808. doi: 10.1016/j.compedu.2005.11.017.

12. GUPTA, M. M., SINHA, N. K. and ZADEH, L. A. (2000) *Soft Computing and Intelligent Systems: Theory and Applications*. Elsevier Science (Academic Press Series in Engineering). Available at: <https://books.google.bg/books?id=7-pQAAAAMAAJ>.
13. HAYKIN, S. (1998) *Neural Networks: A Comprehensive Foundation*. 2nd edn. Upper Saddle River, NJ, USA: Prentice Hall PTR.
14. HERZOG, S. (2006) 'Estimating student retention and degree completion time: Decision trees and neural networks vis a vis regression', *New Directions for Institutional Research*, 2006, pp. 17–33. doi: 10.1002/ir.185.
15. JANG, J.-S., SUN, C.-T. and MIZUTANI, E. (1997) 'Neuro-Fuzzy and Soft Computing-A Computational Approach to Learning and Machine Intelligence [Book Review]', *Automatic Control, IEEE Transactions on*, 42, pp. 1482–1484. doi: 10.1109/TAC.1997.633847.
16. JILANI, T., BURNEY, S. M. A. and ARDIL, C. (2019) 'Fuzzy Metric Approach for Fuzzy Time Series Forecasting based on Frequency Density Based Partitioning'.
17. KIM, Y. and STREET, N. (2004) 'An Intelligent System for Customer Targeting: A Data Mining Approach', *Decision Support Systems*, 37, pp. 215–228. doi: 10.1016/S0167-9236(03)00008-3.
18. KLIR, G. J. and YUAN, B. (1995) *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Upper Saddle River: Prentice Hall, 13.
19. LEMMERICH, F., IFLAND, M. and PUPPE, F. (2011) Identifying Influence Factors on Students Success by Subgroup Discovery., EDM 2011 - Proceedings of the 4th International Conference on Educational Data Mining.
20. LEONG, C. K., Lee, Y. H. and MAK, W. K. (2012) 'Mining sentiments in SMS texts for teaching evaluation', *Expert Syst. Appl.*, 39, pp. 2584–2589. doi: 10.1016/j.eswa.2011.08.113.
21. LOPEZ, M. I. et al. (2012) 'Classification via clustering for predicting final marks based on student participation in forums', *Proc. of 5th Int. Conf. on Educational Datamining*, pp. 148–151.
22. MAGOULAS, G., PAPANIKOLAOU, K. and GRIGORIADOU, M. (2001) 'Neuro-fuzzy Synergism for Planning the Content in a Web-based Course', *Informatica*, 25, pp. 39–48.

23. MIRANDA LAKSHMI, T., MARTIN, A. and PRASANNA VENKATESAN, V. (2013) 'An Analysis of Students Performance Using Genetic Algorithm', *Journal of Computer Sciences and Applications*, 1(4), pp. 75–79. doi: 10.12691/jcsa-1-4-3.
24. MORAN, A. J. (2015) A fuzzy logic approach to teacher performance measured by principal evaluations. Columbus State University. Available at: <https://books.google.bg/books?id=3SeWswEA CAAJ>.
25. NANDESHWAR, A., MENZIES, T. and NELSON, A. (2011) 'Learning patterns of university student retention', *Expert Syst. Appl.*, 38, pp. 14984–14996. doi: 10.1016/j.eswa.2011.05.048.
26. NGUYEN, L. (2014) 'User Model Clustering', *Journal of Data Analysis and Information Processing*, 02(02), pp. 41–48. doi: 10.4236/jdaip.2014.22006.
27. OYELADE, J., OLADIPUPO, O. and OBAGBUWA, I. (2010) 'Application of k Means Clustering algorithm for prediction of Students Academic Performance', *International Journal of Computer Science and Information Security*, 7.
28. PAL, S. (2014) A Study of Academic Performance Evaluation Using Fuzzy Logic Techniques. doi: 10.1109/IndiaCom.2014.6828010.
29. POTEY, M. and SINHA, P. (2015) 'Review and Analysis of Machine Learning and Soft Computing Approaches for User Modeling', *International journal of Web & Semantic Technology*, 6, pp. 39–55. doi: 10.5121/ijwest.2015.6104.
30. ROMERO, C. et al. (2003) 'Discovering Prediction Rules in AHA! Courses', in *Proceedings of the 9th International Conference on User Modeling*. Berlin, Heidelberg: Springer-Verlag (UM'03), pp. 25–34. Available at: <http://dl.acm.org/citation.cfm?id=1759957.1759965>.
31. ROMERO, C. et al. (2013) 'Predicting Students' Final Performance from Participation in On-line Discussion Forums', *Computer&Education*, 68, pp. 458–472. doi: 10.1016/j.compedu.2013.06.009.
32. SAS, C., REILLY, R. and O'HARE, G. M. P. (2003) 'A connectionist model of spatial knowledge acquisition in a virtual environment', *MLIRUM'03 2nd Workshop on Machine Learning, Information Retrieval and User Modelling at 9th International Conference on User Modelling*. Available at: <http://comp.eprints.lancs.ac.uk/2108/>.

33. SINGH, P. (2015) 'A brief review of modeling approaches based on fuzzy time series', *International Journal of Machine Learning and Cybernetics*, 8. doi: 10.1007/s13042-015-0332-y.
34. SINGH YADAV, R. and SINGH, V. (2011) 'Modeling Academic Performance Evaluation Using Soft Computing Techniques: A Fuzzy Logic Approach', *International Journal on Computer Science and Engineering*, 3.
35. TARASOV, D. et al. (2018) Artificial neural networks selection for soil chemical elements distribution prediction, *AIP Conference Proceedings*. doi: 10.1063/1.5044054.
36. TAYLAN, O. and KARAGIYOZLU, B. (2009) 'An adaptive neuro-fuzzy model for prediction of student's academic performance', *Computers & Industrial Engineering - COMPUT IND ENG*, 57, pp. 732–741. doi: 10.1016/j.cie.2009.01.019.
37. VRETTOS, S. and STAFYLOPATIS, A. (2001) 'A fuzzy rule-based agent for web retrieval-filtering', *Lecture Notes in Artificial Intelligence*, pp. 448–453.
38. WEBB, G. I., PAZZANI, M. J. and BILLSUS, D. (2001) 'Machine Learning for User Modeling', *User Modeling and User-Adapted Interaction*, 11(1), pp. 19–29. doi: 10.1023/A:1011117102175.
39. XENOS, M. (2004) 'Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks', *Computers & Education*, 43, pp. 345–359. doi: 10.1016/j.compedu.2003.09.005.
40. XING, W. et al. (2015) 'Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory', *Computers in Human Behavior*, 47, pp. 168–181. doi: 10.1016/j.chb.2014.09.034.
41. YAN, J., POWER, J. and RYAN (1994) *Using fuzzy logic towards intelligent systems*. New York, N.Y. Prentice Hall. Available at: <http://library.lincoln.ac.uk/items/30925>.
42. ZUKERMAN, I., ALBRECHT, D. and NICHOLSON, A. (2001) 'Predicting Users' Requests on the WWW', *UM*. doi: 10.1007/978-3-7091-2490-1_27.

METHODS FOR STORING AUTHENTICATION DATA - A HISTORICAL REVIEW

Petar Dimitrov¹

*¹ University of Economics – Varna/Department of Informatics,
Bulgaria, p.d.dimitrov@ue-varna.bg*

Abstract

Storing user credentials is a process which opens up a challenge in the information technologies. Researchers have always tried to maintain balance between usability and security. This paper aims to identify the different evolutionary steps through which user credentials have gone by. The methodology of this paper consists of content analysis of various scientific publications as well as hands-on experience by researchers in the field of cryptology and cryptanalysis. Several methods for storing authentication data are examined in chronological manner, beginning with first generation computers and latest trends in cryptanalysis. Various flaws of the examined methods are pointed out and solutions in the form of new methods are adopted by the industry. In this historical review it is concluded that with the continuous technological advances any algorithm for storing authentication data might eventually prove insecure.

Keywords: *authentication, password, encryption, hashing, cryptographic algorithms.*

INTRODUCTION

Managing access to increasingly large number of information sources, both in personal life and in professional one, opens the possibility to multiple security threats. Offering easy to use access to large sets of electronic data in various online platforms, desktop application and even mobile services (Penchev 2016), require their developers to be extra vigilant for providing secure and at the same time easy to use environment. In historical aspect, one of the first instances of securing access is the password – it is well documented even during the time of the Roman Empire, but in the context of information technologies “passwords” are being used after the development of the first computers, working in time share mode during the 1960s. when multiple users access same resources in a system, which uses passwords to identify them. Nowadays there are of course other methods of solving this task. In this publication,

we are examining some problems while managing authentication credentials.

1. METHOD OF "PLAIN TEXT" STORING OF USERS CREDENTIALS

In the middle of 1960s in the Massachusetts Institute of Technology (MIT) the CTTS (Compatible Time-Sharing System) project implements for the first time solution to delivering access to multiple files for multiple users. The method for inputting a personal password is chosen among other methods and this proves to be applicable in creating the first main frame computers (McMillan, 2012). In the before mentioned system user identification data alongside with their passwords are stored in a file, named UACCNT.SECRET in plain (non-encrypted) format. There are multiple events when users were able to see other peoples' credentials in plain text.

Even though it is highly insecure, the plain text method of storing user credentials can still be seen in modern information resources. For example in May 2019 Google published a blog post which states that some of their G Suite users' passwords are stored in plain text (Frey, 2019). This happened because a new functionality has been implemented in 2005 that allows G Suite (also known as Google Apps back then) administrators to create passwords for their users and those users are kept on Google servers until users login for the first time, when they are forced to change their temporary password. Although this particular case is not connected to any data leak, undoubtedly it is an example how storing passwords in plain text is considered a "bad practice".

In January 2019 Pedro Kanahuati, Facebook Vice President Security and Privacy, states that internal company system stores more than 600 million Facebook and Instagram users' credentials in plain text (Kanahuati, 2019). According to Kanahuati the stored passwords did not leak and the reason for this is a not well-setup error reporting system.

From those and other examples it can be concluded that even nowadays storing passwords in plaintext is still used, even if it directly contradicts the industry best practices in security.

2. IMPLEMENTATION OF CRYPTOGRAPHIC ALGORITHMS

After analyzing the weak points of the “plain text” method of storing user credentials, researchers during the 1960s decide to apply cryptography as a way to increase security in passwords. Although cryptography is a science, developed for millennia, it marks new heights during 20th century with the progress of mechanical and electro-mechanical machines. The two world wars also strongly affect cryptography development as the countries involved all create ciphers for communication. In 1972 the National Bureau of Standards with conjunction with National Security Agency of the USA began a process for creating a national standard for encrypting unclassified information to be used in all government structures (Tuchman, 1997). In 1974 IBM responds with an algorithm, called Lucifer, which is one of the first implementation of symmetric encryption – an algorithm using a specific key for both encrypting data and decrypting it. IBM’s algorithm, later known as DES (Data Encryption Standard), uses 56-bit encryption key, which consists of 2^{56} possible keys. Even by the time it was implemented, some scientists were concerned that the key is too easy to guess (Gilmore, 2015). Martin Hellman and Whitfield Diffie from the Stanford University calculated that 72 quadrillion combinations can be evaluated in a single day by a machine, which costs 20 million USD – an amount that is quite possible for various government institutions. This was the reason RSA Security organized the first of many “challenges” on 28 January 1997 – to prove DES algorithm is not suitable for storing sensible data (for more information see: RSA Laboratories Secret-Key Challenge <<http://www.rsa.com/rsalabs/97challenge/>>).

One of the first attempts to decrypt the testing message of RSA Security is conducted by a group of scientists, led by Rocke Verser, called DESCHALL Project (short of DES Challenge). Verser envisions a model where multiple computers, connected in a network, are used (Fig.1). He posts this model on Usenet in late March 1997. The server coordinating the multiple computers was based on 486 architecture and had 56MB of RAM. Not long after applications was developed for various architectures in order to be used on the volunteers’ personal computers. Applica-

tions for the modern at the time x64 architectures are also developed.

With the help of this application, a computer running with Pentium processor with 200MHz base clock speed, a million keys per second were able to be processed.

Table 1

**Performance and specifications of computers used
in DESCHALL Project**

Processor	Architecture	Base clock speed	Keys per second
Pentium	486	200 MHz	1 million
PowerPC 604e	x64	250 MHz	1.5 million
UltraSPARC	x64	167 MHz	2.4 million
Alpha	x64	500 MHz	5.3 million

The number of computers using this software quickly grows and towards the end of the project, the number hit 14 thousand unique computers per 24 hours. With this amount of computing power, the speed went up to over 7 billion keys per second. The actual key was found 96 days after the start and by that time 25% of the possible keys were processed. The owner of the PC, which found the actual key was awarded with \$4000 and the rest of the reward was distributed amongst the project founders.

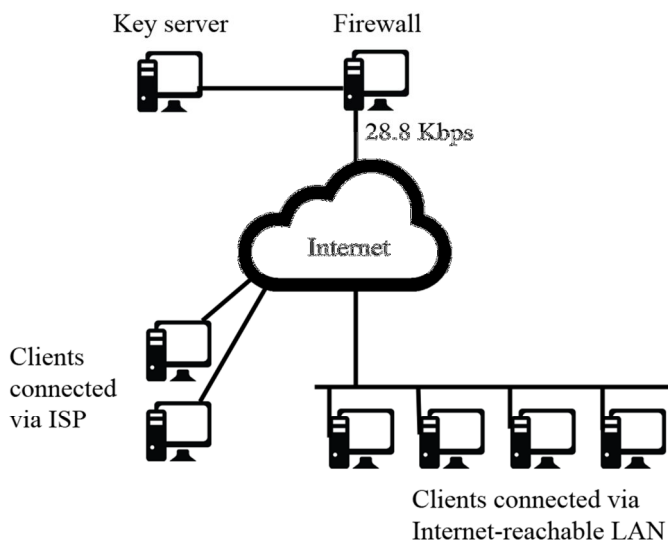


Figure 1. DESCHALL project architecture

The project authors drew several conclusions (Curtin and Dolske, 1998):

1. Small key cryptosystems do not provide adequate security against any but the most trivial of attacks.

2. Whereas previous attacks against “live targets” required the attacker to be relatively well-funded, the kind of power necessary to attack real targets is becoming available to those who are not well-funded, but dedicated enough to make an investment of their time.

3. The potential for performing very large computations without the use of expensive, dedicated hardware, or supercomputers can be seen. Over 7.2 quintillion (10^{18}) instructions were executed. Succinctly, massive Internet computing power is here.

The DESCHALL Project proves that with enough computing power DES algorithm can be reversed. After the successful decrypting in June 1998, the time for decrypting next challenges decrease (Table 2).

**Challenges, organized by RSA Security and the period
for their decryption**

Challenge	Date	Winner	Period
DES Challenge I	January 1997	DESCHALL Project	96 days
DES Challenge II-1	January 1998	distributed.net	39 days
DES Challenge II-2	July 1998	Deep Crack	56 hours
DES Challenge II-3	January 1999	distributed.net, Deep Crack	22 hours

In front of the US congress the director of the FBI at the time Louis Freech states the following: "That is not going to make a difference in a kidnapping case. It is not going to make a difference in a national security case. We don't have the technology or the brute force capability to get to this information." (Computergram International, 1998) After the period for decryption dropped below 24 hours, it became clear that DES did not have the necessary qualities for protecting sensitive information.

3. METHOD OF HASHING FOR STORING PASSWORDS

The term "hashing" is found for the first time in the scientific literature during 1960s, although it has been popular even before that (Knuth, 2000). Researchers seek a way to hide sensible data (most often passwords) through one way function, whose task is to process data with various length and output specific in length and format string. In 1976 Whitfield Diffie and Martin Hellman identify the needs for such particular function in the basis of digital signature (Diffie and Hellman, 1976). Although hashing functions have their use in indexing and searching records in different software applications, the focus of this paper will be on their use in the field of cryptanalysis.

Usually the input data of hashing functions are larger than the output data and this characteristic leads to the occurrence of so called "collisions". This is a result of the so called "Dirichlet principle" which states that if n items are put into m containers, with $n > m$, then at least one container must contain more than one item. There are also theoretical assumptions that there can be a hashing function without collisions, which outputs unique strings for whichever string is input (Fig.2).

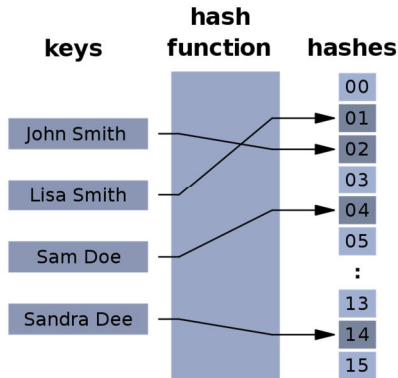


Figure 2. Ideal hashing function for 4 names

An ideal hashing function can be described through the following criteria:

- The function is deterministic in its nature – in order for a function to be deterministic, it must output same data when same data is input;
- The function must be able to calculate hash string fast for data with various length;
- It must not be possible to generate input data by a given hash;
- It must not be possible to find two different input data with same hash (collision resistance);
- Small change in the input data must generate hash which must not be comparable with the hash, calculated before the change – the so called avalanche effect is first used as a term by Horst Feistel (Feistel, 1973), but it was conceptually developed by Claude Shannon in his classified paper “Mathematical Theory of Cryptography” in 1945.

Every hashing function consists of two main components – compressing function and construction. The compressing function converts a given string of data to a shorter one, and the construction describes the way compressing function is called in the hashing process up until the end result.

The hashing functions can be categorized in three different classes – block-cipher based functions, modular arithmetic based functions and

specialized hashing functions. The first attempts for creating such functions were based on block-ciphers like DES crypt algorithm. One of the big problems with this approach is not being able to produce output with larger length than the block length – 64 bits in the case of DES. This inevitably leads to generating collisions.

Other researchers focus on modular arithmetic to create models of hashing functions, based on theoretical assumptions like factorization and the discrete logarithms. An example function is that by Mihir Bellare, who proposes the hashing function to be based on the discrete logarithm problem with prime numbers (Bellare et.al., 1994). Another example function is VHS (Very Smooth Hash), created by Scott Contini, Arjen Lenstra and Ron Steinfeld in 2005 (Contini et.al., 2006). It is based on a hard mathematical problem and that makes finding collisions as hard as solving the actual problem. One year later Markku-Juhani O. Saarinen finds issues with VSH and in his paper proves the function is missing some major properties.

The limitations of using block-ciphers lead to development of specialized hashing functions and they are the first to be bound not to specific hardware, but optimized for use in user applications. One of the most common examples is the MD5 algorithm, proposed by Ronald Rivest in 1991, who previously was working on MD2, MD4 and others.

The MD5 algorithm (short for Message Digest) is published as an improvement over MD4 and its speed is 10 times faster than DES, when used in software environment. One of the key characteristics of MD5 is that it is accessible without special licensing and is easy to be exported in comparison to other hardware-based algorithms. Those among other characteristics make MD5 to be extremely popular even today, although a year after its release, exploits for creating collisions is found (Boer and Bosselaers, 1994).

4. METHOD OF SALTED PASSWORD HASHING

As previously stated, the hashing function is deterministic in its nature and as such it always produces same output given the same input. While this is true, if a breach of security and unauthorized access to the hashed passwords occur, different users with same passwords would have same hashes.

The term "salt" is used to describe 12-bit string, generated by the system clock of the system (Thompson and Morris, 1979). This randomly generated string is concatenated to the user password and then the concatenated result is sent to the hashing function. By doing so users with matching passwords have different hashes and salt strings. A typical example of using salt is the `/etc/passwd` file (Fig.3), which is used in early versions of Unix-based systems up until 1980s. The file contains username, user group, the hashed password and salt in unencrypted format. Every user inside the system can access this file, because there are many system applications that use it, thus making the one way hashing the only line of defense to user passwords. In those early versions of Unix-based systems the salt was limited to 12 bit length (4096 possible combinations) and because computing resources were improving, it became clear that this method required additional protection.

```
# cat /etc/passwd
root:x:0:0:ROOT account:/root:/bin/bash
bin:x:1:1:bin:/bin:/sbin/nologin
daemon:x:2:2:daemon:/sbin:/sbin/nologin
adm:x:3:4:adm:/var/adm:/sbin/nologin
lp:x:4:7:lp:/var/spool/lpd:/sbin/nologin
sync:x:5:0:sync:/sbin:/bin/sync
shutdown:x:6:0:shutdown:/sbin:/sbin/shutdown
halt:x:7:0:halt:/sbin:/sbin/halt
mail:x:8:12:mail:/var/spool/mail:/sbin/nologin
uucp:x:10:14:uucp:/var/spool/uucp:/sbin/nologin
operator:x:11:0:operator:/root:/sbin/nologin
myuser:x:513:520:Test User:/home/myuser:/bin/bash
----- output truncated -----
```

Figure 3. Structure of the `/etc/passwd` file

In every Unix-based distribution from 1980s there is an additional `/etc/shadow` file (Fig.4) which is accessible only to the root user of the system, thus securing storing hashes and salts without giving the right to any user of the system to access them.

```
# cat /etc/shadow
root:$1$UFnkhP.mzcMyajdD9OEY1P80:17413:0:99999:7:::
bin:*:15069:0:99999:7:::
daemon:*:15069:0:99999:7:::
adm:*:15069:0:99999:7:::
testuser:$1$FrWa$ZCMQ5zpEG61e/wI45N8Zw.:17413:0:33:7:::
```

Figure 4. Structure of the /etc/shadow file

Using salt in hashing user passwords does not make the password impossible to compute. Furthermore if a hacker can access a specific user's hash and salt, the only thing stopping the hacker from exposing the password is the actual algorithm of concatenation in the hashing function itself, e.g. whether the salt is added in the beginning or in the end of the password. The main reason the practice of adding salts to hashed passwords is because it adds entropy to the user passwords.

5. METHOD OF "KEY STRETCHING"

Another popular method of storing sensible data like user passwords is the use of “key stretching” algorithms. The first documented occurrence of the term is in Robert Morris and Ken Thompson's paper in 1978. They apply 25 iterations of DES algorithm to 12-bit salt strings. The password back then was limited to 8 characters which make it comparably easy to decrypt via modern technology, but algorithms like PBKDF2 (Password-Based Key Derivation Function 2), using SHA-2, larger salt strings and more iterations are more effective in securing password storage.

One could assume that hashing specific password for second time increases the security of the password, but that statement is false. Probability theory and the birthday problem states that in a class of 30 students, probability one of them to have the same birthday as another student from the same class is about 70%. This approach led to developing Birthday attacks in cryptanalysis. If we take MD5 as an example, and if it is assumed that the percentage of random collision occurrence is 0.001% for specific number of records, after processing the concatenated hash and salt with MD5 once again, the chance of collision will increase twice. In order to mitigate this problem, one should concatenate the salt in plain text again to the hash before applying MD5. In reality the colli-

sion chance is way higher than 0.001%. It is calculated that for 2^{64} number of unique hashes there is 50% chance for finding a duplicate hash (collision).

6. MODERN USES OF HASHING ALGORITHMS

One of the more recently organized challenges is the The Password Hashing Competition (PHC) which was initiated by Jean-Philippe Aumasson in fall 2012. Its aim was to find an unifying standard to fulfil the needs of modern applications and to best protect against attackers. There were 24 candidates, each with different approach, but all with the same goal. The challenge lasted for 3 years and in 2015 a winner was announced – Argon2, designed by Alex Biryukov, Daniel Dinu, and Dmitry Khovratovich from University of Luxembourg. The algorithm is compatible with various programming languages and is both flexible and secure enough to withstand even most sophisticated attacks.

CONCLUSION

Securing authentication data has proven to be a very difficult task as various algorithms are constantly racing with the state of technologies. There have been multiple challenges, organized not only to prove a point but to help create best possible scenario for mitigating risks. Due to the constant pace of technology advances, it could be argued that it might not be possible to invent a perfectly secure software algorithm. Even though algorithms like Argon2 are quickly accepted from the community to be an industry standard, the question is not which one is the best algorithm for storing authentication data, but how long would it take for one to become obsolete. One solution for creating software environment, which is both secure, and at the same time usable, could come in the shape of a hardware 2-Factor Authentication tokens.

REFERENCES

1. BELLARE, M., GOLDREICH, O. and GOLDWASSER, S. (1994) Incremental cryptography: the case of hashing and signing. *Advances in Cryptology, Proceedings Crypto'94*. LNCS 839, Y. Desmedt, Ed., Springer-Verlag, pp. 216-233.

2. BOER, B. and BOSSELAERS, A. (1994) Collisions for the compression function of MD5. *Advances in Cryptology, Proceedings Eurocrypt'93*. LNCS 765, T. Helleseht, Ed., Springer-Verlag, pp. 293-304.
3. CONTINI, S., LENSTRA, A.K. and STEINFELD, R. (2006) VSH, an efficient and provable collisionresistant hash function. *Advances in Cryptology, Proceedings Eurocrypt'06*. LNCS 4004, S. Vaudenay, Ed., Springer-Verlag, pp. 165-182.
4. CURTIN, M. and DOLSKE, J. (1998) A Brute Force Search of DES Keyspace ;login:, May 1998, Volume 23, Number 3 [Online] Available from: <http://www.interhack.net/pubs/des-key-crack> [Accessed 08/08/2019].
5. DIFFIE, W. and HELLMAN, M.E. (1976) New directions in cryptography. *IEEE Trans. on Information Theory*. Vol. IT-22, No. 6, pp. 644-654.
6. FEISTEL, H. (1973) Cryptography and Computer Privacy. *Scientific American*. 228 (5): 15–23. doi:10.1038/scientificamerican0573-15.
7. FREY, S. (2019) Notifying administrators about unhashed password storage. *Inside Google cloud blog* [Online] Available from: <https://cloud.google.com/blog/products/g-suite/notifying-administrators-about-unhashed-password-storage> [Accessed 08/08/2019].
8. GILMORE, J. (2015) DES (Data Encryption Standard) Review at Stanford University. A meeting between Stanford computer science researchers, NBS, and NSA, to discuss the proposed Data Encryption Standard (DES), 1976. [Online] Available from: <http://www.toad.com/des-stanford-meeting.html> [Accessed 08/08/2019].
9. KNUTH, D. (2000) *Sorting and searching*. Boston: Addison-Wesley. pp. 547–548.
10. MCMILLAN, R. (2012) The world's first computer password? It was useless too. *Wired Magazine* [Online] Available from: <https://www.wired.com/2012/01/computer-password/> [Accessed 08/08/2019].

11. PENCHEV, B. (2016) 'Effectiveness of a Conceptual Model for Increased Mobile Banking Security', *Serdica Journal of Computing*, 10(1), pp. 49–62.
12. THOMPSON, K. and MORRIS, R. (1979) Password Security: A Case History. *Communications of the ACM*, Vol. 22, No. 11., pp. 594-597.
13. TUCHMAN, W. (1997) A brief history of the data encryption standard. *Internet besieged: countering cyberspace scofflaws*. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA. pp. 275–280.
14. SAARINEN, M.-J.O. (2006) Security of VSH in the real world. *Progress in Cryptology - Indocrypt 2006*, LNCS 4329, R. Barua and T. Lange, Eds., Springer-Verlag, pp. 95-103.

DEVELOPMENT STAGES OF STARTING SOFTWARE COMPANY, PROBLEMS AND APPROACHES FOR SOFTWARE DEVELOPMENT

Svetoslav Ivanov¹

*¹ University of Economics / Informatics, Varna, Bulgaria,
svetoslav_ivanov@ue-varna.bg*

Abstract

The article aims to review the problems and stages of development in starting software company. The subject is starting software company that is defined with limitations, goal. Problems and stages of product and company development are reviewed from different perspectives – entrepreneurs, entrepreneurial process, product development, software as a product, company and process management. The main goal is to develop products or services to prove their expediency, feasibility and market potential. Specific models for stages and development approaches are presented to manage the entrepreneurial process and reach market success and growth. Product and business model development are based on lean development and agile software development methodologies.

Keywords: *software development, stages, problems, entrepreneurship, starting company.*

INTRODUCTION

Information technologies (IT) development and the effect on the business and society drive their implementation in different areas of economic activities. The role of IT and information systems move from supporting role and business management to key competitive advancements and vitally important activities. Many industries change their processes through the implementation of digitalization which is based on IT technologies. This shift in the IT role is achieved through innovation in information systems, technological base, business management, and entrepreneurship. In this article, we will review two key components important for this process – the software product¹ and the innovation

¹ In the article we will refer as a product to products and services without going into details and differences.

process by creating new software products in starting software companies. The topic is important, because:

- The process of digitalization changes other existing industries
- The intensive rise of the software products industry.
- High level of uncertainty and companies' failures (Blank, 2013).
- The success of SME is key for economic growth (EU commission, 2019)

Therefore, the subject of the article is to review the problems and stages of creating new software products by the entrepreneur in a starting software company. The article will define the terms in this subject, the goals of a starting software company, review different problems and stages of development according to different perspectives and methodologies.

1. STARTING SOFTWARE COMPANY

The key objective of this article is the starting software company. By the means of the name, this term belongs to the terms starting technology company, technology start-up, starting enterprise, software company. The legal base in the European Union and Bulgaria uses the term enterprise and separates 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 (European Commission, 2019). Another approach is to measure the annual revenues or the stage/amount of investments. Or the simple criterion "before becoming a big company". They are equivalent to the definition for size above.

The second approach for defining starting or start-up company is the age. The definitions in the literature are different for the age – up to three years, up to 5 or 10 years. The technological development and shorter cycle of innovation insist on period of fewer than 5 years. A similar criterion is “have little or no operating history” (Blank, 2013).

The third approach defining a 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 (Ries, 2011). The definition for a

start-up from this point is “a company or project initiated by an entrepreneur to seek, effectively develop, and validate a scalable business model” (Katila, Chen and Piezunka, 2013; Ries, 2011). 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) (Blank, 2013). It distinct from Product Development Model in an established company and is the most important criterion for distinction between a starting and an established company.

The fourth approach is to give an accent, especially for the term start-up company, depending on whether the company wants to become big or more precise “to develop and validate a scalable business model” (Blank, 2013). In this article, we will skip the starting companies which plan or behave to stay small – freelancers, small family or service businesses. Therefore, we will add this definition “search for growth or scalable model”.

Including the word software in the term shows that the company is going to develop technological product software. In this article, we will use a technology company in the means to produce a technological product. In the literature, there are sufficient definitions for the term software, it’s content and development, and specific features we review below.

In the literature can be found different forms of entrepreneurial activity. Many of them are activities that copy other company activity and products, reuse a ready product, business model, brand at a high level like a franchise, business acquisition, etc. 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.

Finally, a starting company is initiated by an entrepreneur. We will use the term entrepreneur for a single entrepreneur or a team of entrepreneurs who initiate and lead a starting software company. There is researches (Davidsson, 1989; Storey, 1994) 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.

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 these companies. The criteria could be goals, model, strategy, processes, product, team. These could be (and not limited to):

- The entrepreneur – one or a team, type of team and combination of skills.
- The company mission, values and selected innovation strategy (Drucker, 1999).
- Type of product – product, service, platform, target industry and problem.
- Customer (problem) or technology (find market application) driven (Blank, 2013).
- The type of business model they use (Maurya, 2012; Croll and Yoskovitz, 2014).

2. PROBLEMS

Development of new software product and creating a new business in a starting company is a complex activity. It includes many problems to solve, decisions to take and actions that must be done. We will review here many problems described in the literature as entrepreneurial problems, management problems, software development problems.

First, the entrepreneur should establish their team and vision of the company. We stated the entrepreneur and his personal goals, motivation, orientation to market, capability for growth and management as one key factor. An entrepreneur should have appropriate personal features and character, as well as proper knowledge, skills, and training in the area. The Entrepreneur and the team should have capability in company management, market environment, strategy formation and execution (Storey, 1994; Blank, 2013) and of course, to work as a team. If they do not have the skill should take either training, take into the team of entrepreneurs a co-founder who has the skills and knowledge, use mentors or hire employees.

Second, a starting software company is an entrepreneurial process – to develop new software product, prove vitality and success on the market, establish and grow company. The selection and execution of the process is an important decision. It helps to keep focus and be organized. The key questions are well defined in the literature (Stevenson, et.al, 1985; Drucker, 1999; Maurya, 2012; Blank, 2013; Ries, 2011), and the main ones are:

- What business to do, why, what are the type, goals and the area of business.
- Discovery of opportunity, problem, solution, evaluation, and selection of opportunity.
- Develop a plan – goals, strategy, business plan, define product, market.
- Determine the necessary resources – financial, material, personnel, intellectual, etc.
- Test the market, acquire customers, refine the product for customers.
- Manage with competition and competitive forces.
- Enterprise management – strategy, marketing, production, personnel, resources.
- Growth management – improve business model, increase markets, channels, etc.

Third, the theory and practice show that it is most effective for the entrepreneur to follow a process and use practices and tools that help to solve the above questions efficiently. Therefore, the stages of development and processes from various perspectives are presented in the next point.

Fourth, the entrepreneur should develop a new product that faces market opportunities and uncertainty. There are various conditions that the entrepreneur should take into consideration and problems to solve (Ries, 2011; Blank, 2013; Alvarez, 2014; Kahn, 2012):

- Limited resources – both financial, people, material, information.
- Limited time of opportunities – competition also searches for opportunities.
- High uncertainty – competition, technology change, market forces, and opportunities.

- Understanding the market and customer needs.
- Find a solution and make it a product - customer problem to solve, a solution to form as a product or market application for a technological product.
- Market validation – the product is successful when the market accepts it.
- Make partnerships – team, with mentors, investors, vendors, customers, other partners
- Fast learning – gather experience for market fast and effectively and implement it.
- Development of a validated business model.

All these above require a specific approach and methodology, methods and tools to manage with them. Taking any decision about them requires the entrepreneur to know the state of the company's resources and environment. Therefore, specific tools for analysis should be used to discover a possible solution. Managing the processes and the business, validating the business model and other tasks require taking appropriate decisions. This could be done with the use of appropriate metrics. Many authors (Croll and Yoskovitz, 2014, p.45-62; Ries, 2011; Blank, 2013) state that the metrics should be actionable (lead directly to take decisions), specific for the business model and the stage, and even exists one leading metrics.

Sixth, entrepreneurship requires the use of resources. The entrepreneur should make this available or accessible. The company needs financial resources, needs proper people for company tasks, material, informational, intellectual property. Some authors bring the idea of intellectual capital which is the result of entrepreneur's knowledge, skills, the created knowledge in company and product development, available marketing connections, structural capital for company building, etc. An entrepreneur should take care of all these to be available in a possible way – taken from the team, borrowed, invested, purchased, acquired, developed.

Seventh, the entrepreneur should take care of company management. This means to manage people, processes, resources, relationships with external partners – investors, vendors, customers, employees, con-

tractors, participation in ecosystems, market and virtual networks, etc. Also, at some point, when the company is established, the product is at good state and sells in a niche market, the company must take actions to scalable and do growth. This leads to sequential problems like investments, management changes, personnel changes, coordination problems, etc. Because growth is important, next point reviews this perspective.

Eight, the starting software company does 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 (Larman, 2001; Poppendieck, 2003; Schwaber and Beedle, 2001; Panayotova, Dimitrov, Petrov, Bychkov, 2016; Polkowski, Vasilev, Ghamdi, 2016):

- Fast development in IT and software technologies require short releases.
- Complexity of the product and various vendors of platforms, components, tools, etc.
- Quality is a complex term – implement features, security, scalability, performance, user experience, robustness, integration with 3rd software or systems and other.
- High-quality requirements or increasing quality requirements over time.
- Serious investment/resources before first working or marketable copy.
- Multi-disciplinary – requires the participation of various specialist.
- The 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 of constant learning.
- Requirements - not fully clear initially, need discovery, may change over time.
- Must work on various devices, platforms, browsers, servers, etc.
- Development of technologies requires regular upgrades
- Regular upgrades because of component changes, bugs, security threats, legal, etc.

- The product requires regular 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 – initially, the user or customer is not clear, requirements may change during the search of market niche fit, as well as other changes. 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 the requirement.

3. STAGES OF DEVELOPMENT

Development of new software product and creating a new business in a starting company is a complex activity. It includes many decisions and actions that must be taken and are described in the literature as an entrepreneurial process. One of the most popular definitions for the content of this process is from Stevenson including four phases – definition and evaluation of opportunities, development of the entrepreneurial plan, determining the necessary resources and enterprise management (Roberts, Stevenson, et.al. 2006). Each phase contains different steps. While using such a process is helpful, we should review more specific processes for the definition of starting software company.

First, such an approach and process should count for new product development. Therefore, we could refer to processes under the name new product development (NPD). It is defined as “a thorough understanding of customers' needs and wants, the competitive situation, and the nature of the market is an essential component of new product success” (Kahn, 2012). Their project structure includes four phases:

- Requirements building – what product should do to meet market/business need.
- Product design – how the product will meet the requirements.
- Product implementation – construction and testing of the product.
- Commercialization – production and market launch.

Based on this structure, there are various models developed. From the mentioned in the literature we should mainly note these models - BAH model (base model), Stage-Gate model (the well-known Waterfall model in software development (Kahn, 2012; Larman, 2001) appropriate for stable environments), Lean Startup - a model inherited from Lean Development (Ries, 2011), Exploratory Product Development by Dothar and Morrisay.

Next, we should review the typical limitations of a starting software company and its product. First, there are entrepreneurial limitations for a starting business - limited resources, limited time of opportunities, high uncertainty and high-rate of failure (Blank, 2013; Ries, 2011). Next, it should meet the principles of successful entrepreneurship – clear vision, appropriate strategy, flexible design, focus on product development, market validation, business model development, and validation, etc (Blank, 2013). Third, the specific conditions for software products - fast development in IT technologies, changes in software technologies, various vendors for platforms, tools and components, aggressive behaviour from large companies to develop new versions and new features on short cycles, high quality requirements and high price of software development specialists, need of short learning cycle (Poppendieck, 2005; Ries, 2011). Therefore, we need to review methodologies that help the company manage uncertainty, do market and model validation, follow fast learning curve to discover the product requirements and do short cycles of development.

The Theory and Practice, especially in product development and software development propose such methodologies. They should be flexible, manage uncertainty through adaptation, feedback and fast learning. Also, to manage appropriately quality, time and resources for product development. These methods are known as agile and follow the general NPD stages – Requirements, Design, Implementation, Deployment. Each stage has iterations with actions like planning, execution, release, getting feedback and updating from feedback. Each iteration ends with a given level of product development done. During the iteration, the product and plan could be adapted to the market requirements and discovery process. The control of time is through fixed time per iteration. The adaptation is

via prioritization in planning and feedback at each iteration with regular feedback from the client (Larman, 2001). Appropriate methodologies are Lean Software Development (Poppendieck, 2003), Scrum (Schwaber and Beedle, 2001), Extreme Programming (Larman, 2001). In each approach the flow of work is determined by the client with accent to be effective (maximum value) and velocity. There are tools for feedback, prioritization, following empirical results and learning.

We should note that the above methodologies have specific characteristics – they are dedicated to software product development with a regularly available client and feedback. In a starting software company, there is no client or the client and requirements are not clear or refined, so they must be discovered. Also, the business model is not clear and it should also be built and validated. Stage Requirements is most open to risks.

As we mentioned above the company does an entrepreneurial process that has the features of discovery and market validation for the product, business model and growth/scalability. Therefore, we should use methods into the first phase “Requirements” to determine the product requirements and as feedback in later iterations during the design, implementation and commercialization. Such methods are developed as:

- Fuzzy Front-End of NPD – formulate the concept, decide investment (Koen, 2001).
- Design Thinking – discover a problem and design solution for it (Dorst, 2012).
- Customer Development – develop a product and initial customer base (Blank, 2013).
- Lean Start-up - based on Customer Development and Lean Development principles brings a hypothesis-testing based model (Ries, 2011; Maurya, 2012).
- Lean Canvas – a tool for business modeling that adapts Osterwalder’s Business Model Canvas for starting companies (Maurya, 2012).

The above methods can be integrated with Agile Development as shown in Figure 1. It combines Design Thinking to generate the idea for problem and solution, Lean Startup for building and validating the solution and business model for product-market fit, and Agile Development

to implement it. As shown, Lean Startup consists of cycles “build, test and learn” for the solution and product-market fit similar to Agile Development cycles. Both, Lean Startup and Agile Development process could go in parallel and have information flow, because of the first supplies requirements from learning activity and the second, the product to test. Another combination, called Lean Launchpad, is to combine Business Model Canvas (Maurya, 2012), Customer Development and Agile Development (Blank, 2013).

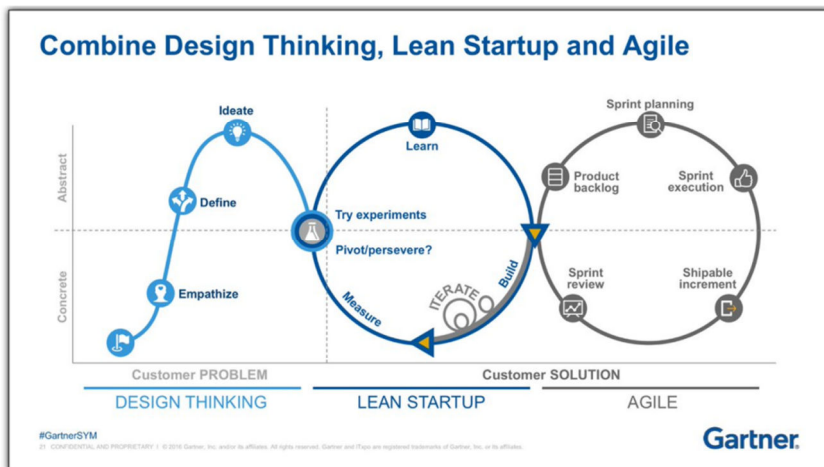


Figure 1. Integration of Design Thinking, Lean Startup and Agile. (Source: Gartner, Inc.)

Another perspective we could review is the integration of building a product, team and business in a simplified model. Such a simplified model with minimum needed activities is in Figure 2. The first phase is Formation with tasks to establish team, define problem and solution as idea and concept. The second phase is Validation to build a Minimum Viable Product (MVP is defined by Ries, 2011) that proves market vitality and potential. The third phase is Growth to the focus on the company organization to establish and grow.

When a starting company is developed through stages the entrepreneur should count for the business results to make appropriate decisions.

A model connected to the customer base and business results is the one presented by Croll and Yoskovitz (2013, p.153-157) when they describe the analytics used to measure the success of the business. The five stages are close to the stages in the above models and are listed in Table 1.



Figure 2. Startup Development phases (Source Startup Commons, available from <https://www.startupcommons.org/startup-development-phases.html>)

Table 1

Stages list according to Croll and Yoskovitz (2013), Maurya (2013), Ries (2011).

№	Stage	Croll and Yoskovitz explanation	Lean Startup terms
1	Empathy	Identify a real problem and solution	Customer Development
2	Stickness	Early customers tests for market fit	MVP building
3	Virality	Early customers acquisition and tests	Organic growth, Sticky engine
4	Revenue	Optimize the revenue	Monetization, Price Engine
5	Scale	Expansion - markets and channels	Inorganic Growth

We should mention the question of financing starting companies. The interest in financing is because a starting company has a window of time for the opportunities, there is competition and financial agents search for investment opportunities. In most cases, the investment stages are connected to stages of the business as shown in Figure 2. Stage Formation is funded by the entrepreneur's resources, friends, family. Stage Validation (called Seed Capital) from entrepreneurs, business angels, crowdfunding, etc. At stage Growth, the interest is from and to venture capitals, funds, etc in different levels.

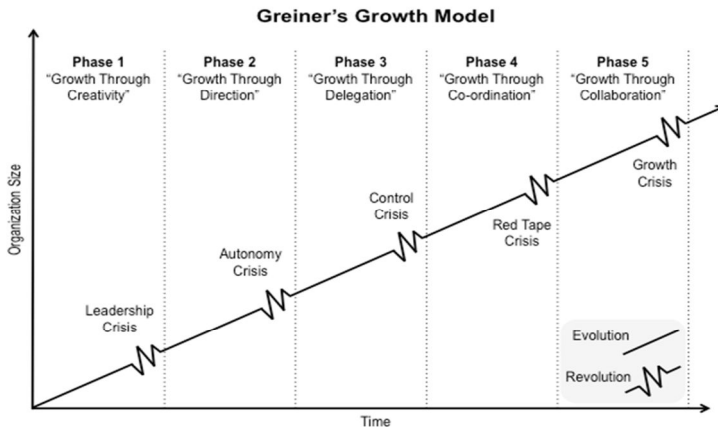


Figure 3. The Five Phases of Growth. (Source Grainer 1998)

Finally, we should view stages of development from management and organizational view. One such model is from Larry Greiner with 5 stages of growth and revolutionary crisis's between stages as shown in Figure 3. "The resolution of each revolutionary period determines whether or not the organization will move forward into its next stage of growth" (Greiner, 1998). The model is useful with recommendations for organization, management practices and actions in each phase to reach the goal for growth and scalability. Otherwise, the company may stay at a given stage, stop the growth or even fail if it is in a too early phase.

CONCLUSION

The current article defined the term starting software company and the main company goal. It could be successful with proper new product development that fits the market niche with a proper business model to reach company vitality and growth. The problems and stages are discussed from different perspectives of the company definition. The stages are reviewed with focus on Lean Product Development and Agile Software Development.

REFERENCES

1. BLANK, S. (2013) *The Four Steps to the Epiphany*. K&S Ranch Publishing.
2. CROLL, A., YOSKOVITZ, B. (2013) *Lean Analytics*. Sebastopol, CA: O'Reilly Media.
3. DAVIDSSON, P. (1989) *Continued entrepreneurship and small firm growth*. Stockholm: The Economic Research Institute.
4. DORST, K. (2012) *Frame Innovation: Create new thinking by design*. Cambridge, MA: MIT Press.
5. DRUCKER, P. (1999) *Innovation and Entrepreneurship, Practice and Principles*. New York: Harper & Row.
6. EUROPEAN COMMISSION (2019) *What is an SME?* Available from: https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en [Accessed: 2/08/2019].
7. GRAINER, L. (1998) *Evolution and Revolution as Organizations Grow*. Harvard Business Review. May-June 1998.
8. KAHN, K. (2012). *The PDMA handbook of new product development*. 3 Ed. Hoboken, New Jersey: John Wiley & Sons Inc.
9. KOEN; et al. (2001). *Providing clarity and a common language to the 'fuzzy front end'*. Research Technology Management. 44 (2): p.46–55.
10. MAURYA, A. (2012) *Running lean: iterate from plan A to a plan that works*. The lean series (2nd ed.). Sebastopol, CA: O'Reilly.
11. LARMAN, C., (2003) *Agile & iterative development: A manager's guide*. Pearson Education.

12. PANAYOTOVA, G., DIMITROV, G., PETROV, P., BYCHKOV, O. (2016) *Modeling and data processing of information systems*. 3rd International conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, Poland, IEEE, pp.154-158.
13. POLKOWSKI, Z., VASILEV, J., GHAMDI, R. (2016) *ICT green ecosystem*. Studies and Proceedings of Polish Association for Knowledge Management, 82, 2016, pp.61-73.
14. POPPENDIECK, M., POPPENDIECK, T. (2003) *Lean Software Development: An Agile Toolkit*. Addison-Wesley Professional.
15. RIES, E. (2011) *The Lean Startup*. Crown Publishing Group.
16. ROBERTS, M., STEVENSON, H., SAHLMAN, et al. (2006) *New Business Ventures and the Entrepreneur*. 6th Ed. New York: McGraw-Hill/Irwin.
17. SCHWABER, K., BEEDLE, M. (2001) *Agile Software Development with Scrum*. Upper Saddle River, NJ: Prentice Hall.
18. STOREY, D.J., (1994) *Understanding the small business sector*. London: Routledge.

THE ROLE OF THE SMART CITY CONCEPT IN THE PROCESS OF URBAN TRANSFORMATION

Antonio Hadzhikolev¹

*¹ University of Economics – Varna, Bulgaria,
antonio.hadzhikolev@ue-varna.bg*

Abstract

According to UN statistics, by 2050 two-thirds of the global population will live in cities. The increased urbanization ratio will cause different problems as well pose new challenges before the government of cities. In order to successfully solve these problems, an urban transformation is needed. One of the possible solutions is oriented towards the application of modern ICT in the city's wide infrastructure. The aim of this paper is to present the smart city concept and its role in future urban transformation and development. The main benefits of implementing such technology, according to the author of the paper, can be found in two areas: 1) improving the quality of life, and 2) helping the decision-making processes in order to make data-driven decisions. These benefits can not only lead to but also ensure the sustainable development of cities in the future.

Keywords: *Internet of Things, smart cities, ICT application, urban transformation.*

INTRODUCTION

In recent years several are the main trends in the information and communication technologies (ICT) sector: mobility, wearable devices, big data and the concept “Internet of things”. We think that the last one will have a key role in the future development of the ICT sector and the world as a whole. One possible application of the concept is the so-called smart cities. The **aim** of this paper is to present the smart city concept and its role in future urban transformation and development. This work is **limited** only to a theoretical study of the problem.

1. URBAN TRANSFORMATION IN THE 21TH CENTURY

With the increasing urbanization ratio, the global urban infrastructure is facing a different kind of problems. The statistics show that for the period 1960 – 2017 the global population living in cities has in-

creased by more than 20 percent (Roser and Ortiz-Ospina, 2019), or in absolute numbers – from one billion in 1960 to 4,18 billion in 2017. According to United Nations statistics, by 2050 around 60 percent of the global population will live in cities (Meredith, 2018). This will pose new challenges before cities' administrations, as well as worsen the already existing **problems** that can be classified in the following areas:

- Budget-related problems;
- Overpopulation;
- Increased usage of resources;
- Traffic and mobility-related problems;
- Poor infrastructure;
- Air pollution;
- Problems related to education and healthcare;
- Increased number of crimes;
- The need to create new jobs, and others.

In order to solve these problems, **an urban transformation is needed**. Urban transformations are one of the specific characteristics of the cities (Clerici and Mironowicz, 2009, pp. 23-32). Since the middle nineteenth century, a fourth one is underway. Urban transformations can be associated with two main aspects: 1) increasing the size and effects of cities on the world's economy, and 2) transformation of cities' internal structures. The first one is related to the need of facing the increased number of citizens, while the second one relates to adapting the existing cities' infrastructure towards the necessary changes. The traditional expansion of cities is not capable anymore of solving the problems and challenges caused by globalization and migration.

One of the plausible solutions for urban transformation is oriented towards the **integration of the latest innovations in the ICT field** in the global city infrastructure.

2. THE INTERNET OF THINGS – FUNDAMENTAL BASE OF SMART CITIES

The current phase of the evolution of the Internet is mainly related to the increased number of devices connected to the global network. According to Cisco, by 2020 there will be about 50 billion connected

devices (Newman, 2017), the main contributor for that being the objects of our daily life. In the current decay we are also facing: an increase in the volume and the variety of data that is generated on a daily basis which leads to the so-called “big data” problem; the evolution of the artificial intelligence, and its related technologies; mass use of cloud technologies; trend of miniaturization of devices, and others. They all lay in the base of **the “Internet of things” (IoT) concept**. The first person who used the term IoT is Kevin Ashton in a presentation about the benefits of using RFID before the management of Procter & Gamble (Ashton, 2009). He states that this technology would allow the devices to be interconnected without the interaction of people. According to Cisco, the real beginning of the IoT concept are the years 2008 and 2009 when the number of devices, connected to the global network, exceeded the number of world’s population at that time. The once called “Internet of people” is now the “Internet of things”.

Based on analysis of different definitions (Dorsemaine et al, 2015, pp. 72-77; Atzori, Iera and Morabito, 2017, pp. 122-140, and others), we can summarize that IoT is an evolving network based on interconnected information and communication technologies that allow for different physical objects from the analog world to be connected to the network using sensors and RFID by which they interact with the surrounding environment and communicate with each other, exchanging data in the Internet, without the need of human interaction. The common thing between the definitions of different authors lays in the idea that the first version of the Internet is used for sharing data generated by humans, while the second one – for data generated by objects. Although the Internet has evolved a lot from its prototype – ARPANET, created at the end of the sixties, in its core it hasn’t changed much. This allows us to conclude that the “Internet of things” can indeed be viewed as the next phase of the evolution of the Internet.

The application of the IoT concept can be found in many areas, including smart homes, connected industry, smart energy, smart cars, smart buildings, and others. The IoT concept is fundamental to the idea of smart cities.

3. SMART CITIES – DEFINITION AND KEY ASPECTS

At the beginning of the nineties of the twentieth-century discussions began related to different possibilities of facing the needs of the growing urban population and improving the quality of life in cities. In different publications over the years, the concept can be found under different names: web, virtual, digital, environmental, and smart city. At first, the concept was limited only to the integration of ICT in the urban environment, but later its meaning became even broader placing the citizens in the center of the smart city ecosystem.

In 2005, thanks to its experience, the American technological company Cisco received funding from the Clinton Foundation in order to conduct a research on the topic of how cities can be made more sustainable (Information Age, 2012). Over the next five years, the company worked on the “Connected Urban Development” project along with the administration of different cities, among them San Francisco, Amsterdam, and Seoul. The idea was to test different pilot initiatives in order to prove the concept’s potential. In parallel, IBM also worked on the problem. In 2008 the company launched its initiative called “Smarter Planet”, the aim being study of the possibilities of application of modern IT technologies in order to solve some of the global problems. The next year another program, called “Smart Cities”, started that was focused mainly on the problems in the urban environment. Pioneer in this project was the Brazilian capital Rio de Janeiro where IBM created a center for emergency situations. It allowed the authorities to review data collected from a broad network of installed sensors that were able to help the decision-making process in different crises.

In 2007 the first European group dedicated to the smart cities' problems was created, and in 2011 – the first American. In 2014 several standardizations by ISO and ITU were made. According to Arup, the global market of services for smart cities will be worth 400 billion per year (Fernández, 2013).

Different definitions for the smart city concept exist from its invention at the end of the twentieth century. Some of them are listed in **table 1**.

Table 1

Some definitions of the smart city concept

№	Organization	Definition
1	Cisco	“A smart city uses digital technology to connect, protect, and enhance the lives of citizens. IoT sensors, video cameras, social media, and other inputs act as a nervous system, providing the city operator and citizens with constant feedback so they can make informed decisions” (Cisco)
2	Telefonica	“Innovative city that uses technology (advanced infrastructures, platforms and services) to provide more efficient urban services, improving the quality of life of citizens, and developing a new relationship between the government and local companies, meeting present and future needs of the city in economic, social and environmental terms, guaranteeing sustainability” (León, 2016)
3	ISO	“A new concept and a new model, which applies the new generation of information technologies, such as the internet of things, cloud computing, big data and space/geographical information integration, to facilitate the planning, construction, management and smart services of cities” (ISO, 2014)
4	ITU	“A smart sustainable city is an innovative city that uses information and communication technologies (ICTs) and other means to improve quality of life, efficiency of urban operation and services, and competitiveness, while ensuring that it meets the needs of present and future generations with respect to economic, social, environmental as well as cultural aspects” (ITU)

We can summarize that the smart city concept has **interdisciplinary meaning**. Our opinion is that this concept should be associated with several main aspects: 1) improving the quality of life and providing better services for the citizens, 2) application of ICT for the means of gathering data, which analyzed can improve the decision-making processes; 3) improving the efficiency of the communication between citizens and local administration, and encourage the citizens to take active

role in the administration of cities; 4) providing sustainable development of cities; 5) improving the competitiveness of cities and 6) combining different processes in one interconnected urban ecosystem.

According to the European Union, the smart city can be defined along six axes (or dimensions) – smart economy, smart mobility, smart environment, smart people, smart living, and smart governance (European Parliament, 2014). The researcher in this field Anthopoulos adds the smart infrastructure, as well as the smart services (Anthopoulos, 2017, pp. 5-45). Different authors also view as axes of the concept the big data, the open data, the innovations, etc.

The dimensions are as follows:

- *Smart economy* – application of different ICT technologies and innovations with the goal to stimulate the businesses expansion, offering new jobs and improving the competitiveness of cities;
- *Smart mobility* – improved means for monitoring and control of the traffic in real-time;
- *Smart environment* – application of innovative IT solutions in order to protect and manage different resources in a more efficient manner (for example, systems for monitoring and management of carbon emissions, garbage collecting systems, and others);
- *Smart people* – integrating solutions to help the productivity of people, their creativity and encourage innovations;
- *Smart living* – application of innovations in order to improve the quality of life in cities;
- *Smart governance* – use of different solutions in order to help planning and managing the urban processes, resulting in sustainability, as well as improving the communication between citizens and local administration;
- *Smart infrastructure* – integration of different sensors in existing and new infrastructure (energy and water supply network, streets, facilities, etc.);
- *Smart services* – use of ICT solutions in healthcare, education, tourism, and other key sectors of the economy.

In order to better understand the nature of the smart city and the **differences** between conventional and smart cities, let us review some of their practical aspects, as shown in table 2.

Table 2

Comparative analysis between conventional and smart city

Practical aspect	Conventional city	Smart city
Energy consumption	The high energy consumption could lead to a shortage of energy resources and black-outs	Monitoring the consumption will provide more efficient use of resources, resulting in flawless service
Mobility and air quality	Traffic jams and congestions, poor traffic planning and management, lack of park spaces, air pollution	Better traffic management and air quality
Long-term city planning	Ineffective planning that can lead to increased costs	The monitoring can lead to effective planning, therefore fewer costs
Water supply management	The lack of monitoring of the water supply can lead to an outage of the service and financial loss for the operating company	The efficient monitoring will result in flawless service and cut the loss of water resources caused by faulty infrastructure
Health services	Lack of a unified register containing health records and medical history about the patients. Inability to adapt the latest ICT innovations in the health services	A unified register containing health records and medical history about the patients. Ability to adapt the latest ICT innovations in the health services, such as remote monitoring of patients' condition
E-government	A small number of e-services and poor dialogue between citizens and city's administration caused by lack of appropriate communication channels	Variety of e-services and improved dialogue between citizens and the city's administration
Competitiveness (from investors' point of view)	Lack of detailed information about the potential of the city from the investors' point of view	Competition between cities in order to attract investments, incl. facilitating the regulatory procedures

The application of the smart city concept can be found in different aspects of the urban environment – infrastructure, mobility, services,

living, and others. The services are only one of the components of the smart city.

The **meta-architecture** of the smart city, according to ITU, is visualized in figure 1 (ITU, 2015). The natural environment unites all the natural resources located on the territory of the city. The hard infrastructure (Non-ICT based) represents all urban infrastructure objects, incl. roads, facilities, bridges, etc. The hard infrastructure (ICT-based) is all the hardware that supports the smart services and by which they are delivered to the end-customers. Among the ICT-based infrastructure are data centers, communication networks, IoT devices, sensors, and others. Smart services are delivered using the two types of infrastructure. Some of the smart services are smart transportation, smart healthcare, smart waste management, and others. The soft infrastructure, on one hand, relates to the people living in the cities, and on other, to different business processes, software applications, and data.

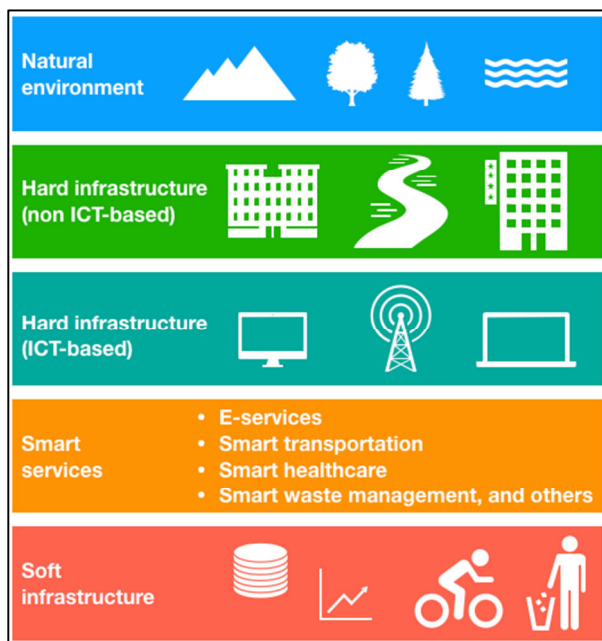


Figure 1. Smart city meta-architecture

Source: ITU (2015), adapted by the author.

The **technological aspects** of the smart city concept are connected mainly to solutions that provide the ability to collect, analyze, visualize and integrate data into a variety of services. The data lay in the core of the smart city concept. It is collected using different means – sensors (based on RFID, lasers, infrared, ultrasound), CCTV, information systems, social media, and others. In order to help the decision-making processes, the data should be processed into information, knowledge, and wisdom. The raw data is a collection of text, symbols, numbers and facts about events, people, situations, and others. After it is processed, it becomes information that has a specific structure and reveals certain links and associations. Organized information is called knowledge. It is context-oriented and synthesized and reveals patterns and models. Wisdom is knowledge applied in a certain field. It has practical meaning, reveals principles and supports the decision-making processes using the gathered knowledge and experience. In our opinion, smart cities can indeed help the cities' administration make more information-oriented decisions. The generated and then analyzed data can be used in real-time: 1) for integration in so-called dashboards which contain various information and reports that can be used by local administration and 2) as open data which is fundamental to the processes of creating Internet-based applications that can be used by the citizens.

There are two main **approaches** for the practical implementation of the concept: 1) application of the concept in new cities where the necessary ICT infrastructure is planned and built at the beginning, and 2) application of the concept in existing cities in which the ICT infrastructure should be additionally designed and constructed. In comparison, the first approach provides an easier way of integration of different kinds of smart services. Although there are practices of building entirely new cities in China in India, the more common approach is to implement the concept in existing cities that have been developing for hundreds of years.

The application of the smart city concept is a complex and resource-intensive process. Given its complexity, the practical implementation of the idea should be realized in stages, the process being iterative. Not least, when applying the concept, the respective city's policies, goals, and opportunities for financing the initiatives should be taken into account.

4. SMART CITIES – ADVANTAGES AND CHALLENGES

Different **examples** of good practices in the field of smart cities exist, including Barcelona, Amsterdam, Helsinki, Hamburg, London, Paris, Nice, Moscow, New York, Chicago, Seattle, Singapore, and others. Some countries such as South Korea have already begun constructing new cities from scratch, for example, Songdo, that will implement all the latest ICT technologies at the stage of initial planning and designing. The European Union also encourages the development of smart cities through different programs and possibilities of funding projects in this field.

The **advantages** of smart cities can be classified as follows:

- Better and more convenient services for the citizens;
- Better management of cities and their resources;
- Improving the quality of life;
- Improved city's infrastructure;
- Turning cities into greener spaces.

In addition, smart cities can create more jobs and encourage investments. These spaces can also help improving the dialogue between citizens and the city's administration. According to different studies, the smart applications can reduce accidents in cities by 10 percent, solutions related to mobility can save between 2 and 4 hours every day that can be used for other purposes, smart cities can also reduce carbon emissions by 15 percent, can save up to 80 liters of waters per day, and so on (Andrews, 2018).

Along with the advantages, there are some **challenges** before the concept:

- *Limitations related to the mass usage of IPv4* – this standard cannot meet the increasing number of devices that need a unique IP address in order to be addressable in the global network. This is the reason why worldwide migration to IPv6 is required;

- *The necessity of standardization* – as already mentioned, the data in smart cities is collected using a variety of objects that use different communication mechanisms and data models. Efforts should be put on the interoperability issue;

- *Potential attacks* – the variety of connected devices and the transmitted data can be potential targets of different kind of attacks;

- *Lack of confidence in the concept* – given the sensitive data that is transmitted, a high level of security should be guaranteed for citizens to gain trust in the concept;

- *High costs* – the practical implementation of the smart city concept is a complex process that requires time, budget and human resources. These can be limiting factors for some cities.

Nevertheless, at the moment smart cities present a possibility to deal with the challenges caused by migration, globalization and the increasing number of urban population worldwide.

CONCLUSION

We think that the smart city concept will have a key role in the future development of cities and in the process of urban transformation. The main benefits of implementing such technology can be found in two areas: 1) improving the quality of life, and 2) helping the decision-making processes in order to make information-based decisions. These benefits can not only lead to but also ensure the sustainable development of cities in the future.

The smart city topic is very promising and will be further analyzed by the author. In future works, the practical side of the problem can be studied in-depth, including the implementation of the concept in a particular city and the effects that it will have on the quality of life and the decision-making processes.

REFERENCES

1. ANDREWS, J. (2018) How Smart City initiatives can improve quality-of-life indicators by 10-30%: report [Online], Available from: <https://news.itu.int/smart-cities-improve-quality-of-life-says-mckinsey/> [Accessed 15/08/2019]
2. ANTHOPOULOS, L. (2017) The Rise of the Smart City. In: Understanding Smart Cities: A Tool for Smart Government or an Industrial Trick?, Book series: Public Administration and Information Technology Vol. 22, Springer, pp. 5 – 45.
3. ASHTON, K. (2009) That ‘Internet of Things’ Thing. RFID Journal [Online] Available from: <http://www.rfidjournal.com/article/view/4986> [Accessed 07/08/2019]

4. ATZORI, L., IERA, A. and MORABITO, G. (2017) Understanding the Internet of Things: definition, potentials, and societal role of a fast evolving paradigm. *Ad Hoc Networks*, Vol. 56, pp. 122 – 140.
5. CISCO. What is a smart city? [Online] Available from: <https://www.cisco.com/c/en/us/solutions/industries/smart-connected-communities/what-is-a-smart-city.html> [Accessed 12/08/2019]
6. CLERICI, A and MIRONOWICZ, I. (2009) Are landmarks essential to the city – its development? In: Schrenk, M. et al. (Ed.) *Proc. of REAL CORP 2009 CITIES 3.0 – Smart, Sustainable, Integrative. Strategies, concepts and technologies for planning the urban future*, pp. 23-32.
7. DORSEMAINE, B. et al. (2015) *Internet of Things: A Definition & Taxonomy*. 9th International Conference on Next Generation Mobile Applications, Services and Technologies, Cambridge, pp. 72 – 77.
8. EUROPEAN PARLIAMENT (2014) Directorate General for Internal Policies. *Mapping Smart Cities in the EU*.
9. FERNÁNDEZ, F. (2013) Report reveals \$400 billion smart cities opportunity globally [Online] Available from: <https://www.arup.com/news-and-events/news/report-reveals-%24400-billion-smart-cities-opportunity-globally> [Accessed 10/08/2019]
10. INFORMATION AGE (2012) IBM, Cisco and the business of smart cities [Online] Available from: <https://www.information-age.com/ibm-cisco-and-the-business-of-smart-cities-2087993/> [Accessed 10/08/2019]
11. ISO (2014) Smart cities – Preliminary Report [Online] Available from: https://www.iso.org/files/live/sites/isoorg/files/developing_standards/docs/en/smart_cities_report-jtc1.pdf [Accessed 12/08/2019]
12. ITU (2015). Setting the framework for an ICT architecture of a smart sustainable city – Focus Group Technical Specifications [Online] Available from: https://www.itu.int/en/ITU-T/focusgroups/ssc/Documents/website/web-fg-ssc-0345-r5-ssc_architecture.docx [Accessed 15/08/2019]
13. ITU. Smart Sustainable Cities at a Glance [Online] Available from: <https://www.itu.int/en/ITU-T/ssc/Pages/info-ssc.aspx> [Accessed 12/08/2019]

14. LEÓN, M. (2016) Digital Transformation in the new Smart City models: challenges and opportunities.
15. MEREDITH, S. (2018) Two-thirds of global population will live in cities by 2050, UN says [Online] Available from: <https://www.cnn.com/2018/05/17/two-thirds-of-global-population-will-live-in-cities-by-2050-un-says.html> [Accessed 04/08/2019]
16. NEWMAN, D. (2017) The Top 8 IoT Trends For 2018 [Online] Available from: <https://www.forbes.com/sites/danielnewman/2017/12/19/the-top-8-iot-trends-for-2018/#19880c7167f7> [Accessed 07/08/2019]
17. ROSER, M. and ORTIZ-OSPINA, E. (2019) World population growth [Online] Available from: <https://ourworldindata.org/world-population-growth> [Accessed 04/08/2019]

ONTOLOGY-BASED MATERIALIZED VIEWS FOR REVERSE ENGINEERING: TOWARD A SET OF CONCEPTUAL GUIDELINES

*Oumar Sy*¹

¹ *Université Gaston BERGER/U.F.R
Sciences Appliquées et de Technologie (Comp. Sc.),
Saint-Louis, Sénégal, oumar.sy@ugb.edu.sn*

Abstract

Data reuse and meta-data handling remain a tricky problem for both the designers and managers, especially when schemas reverse engineering is on demand and that some data are stored in materialized views. In this paper, we tackle such problem by using ontology-based meta-materialized views. Indeed, ontologies, which are semantics-based, ensure the stability of the underlying schemas of the data repositories and to ease the overall access and processing of the data and meta-data. Our proposal is sustained by a set of conceptual guidelines and outlined through a case study example.

Keywords: *Relational database, conceptual modeling, domain ontology, reverse engineering, materialized view, SQL.*

INTRODUCTION

The reuse of data and meta-data remain a tricky problem, especially when the reverse engineering of schemas is on demand. Yet, legacy systems such as traditional databases and the advanced databases systems, and the ontologies, must permanently be up to date. In this context, reverse engineering is useful for the understanding of existing systems. Already, in the early 90s, CASE (Computer-Aided systems-Engineering) tools were used to maintain and enhance existing systems (Chikofsky and Cross, 1990). Now, according to Gomez-Perez and Ma (Gomez-Perez and Ma, 1999, pp. 139-156), reverse engineering consists of producing a semantic model based on the code in which the data model or the knowledge model (e.g., an ontology) is implemented. On the other hand, the use of ontologies for creating more intelligent and effective enterprise information systems has increased considerably in recent years (Reynares, Caliusco, Galli, 2015). Indeed, ontologies allow the

specification of high-level formal conceptualizations (Gomez-Perez, 1999, pp. 6.1.1 - 6.1.18). However, most of the proposed approaches are not domain independent, such that the developed models are generally specific to a given domain of real world. Therefore, even for the same field of application, the reuse is not easy to achieve or not achieved at all because the common trend is to approach the problem of reverse engineering with object-oriented techniques, e.g., (Reynares, Caliusco, Galli, 2015; Gomez-Perez, 1999; Ristić, Aleksić, Celiković, Dimitrieski, Luković, 2014; Zhang, Karcher, 2012; Müller, Jahnke, Smith, Storey, Tilly and Wong, 2000; Shatnawi, Seriai, Sahraoui, Alshara, 2016; Ali, Fernandes, Paton, 2000; Brenner and Glimm, 2017; Koch, Lupei, Tannen, 2016). Conceptual modelling approaches of ontologies for data integration and reuse have been well developed in (Ristić, Aleksić, Celiković, Dimitrieski, Luković, 2014; Alberts & Franconi, 2012; Zedlitz and Luttenberger, 2014), but materialized views were not considered.

Here, we tackle the problem from the semantic models point of view and mainly when ontologies are considered. Indeed, to the best of our knowledge, only (Glimm, Kazakov, Liebig, Tran, and Vialard, 2014, pp.180-195; Volz, Staab & Motik, 2003, pp. 707-724; Volz, Staab & Motik, 2005, pp.1-34) have dealt with the problem of ontologies' materialization. However, rather than the use of materialized views with ontologies for reverse reengineering, the authors only discussed ontologies' storage, without conceptualization of materialized views in the design process. Moreover, it is not clear how ontologies' models are linked to business-data. Thus, we aim to address criticisms on the above works in the related works.

In the 1990's, Chikofsky and Cross (Chikofsky and Cross, 1990) argued that the term "reverse engineering" originate from the analysis of hardware, and defined six terms, namely (i) forward engineering, (ii) reverse engineering, (iii) redocumentation, (iv) design recovery, (v) restructuring and (vi) reengineering. The relationships between these conceptual terms have been rigorously related and defined based on a three-level's abstraction architecture, conceptually quite similar to the ANSI/X3/SPARC architecture dedicated to database schemas design. Such an architecture stands as the cornerstone in our approach, when aiming

to conciliate databases (DB) and ontologies, without redefinition of terms, concepts and methodologies (Sy, Lo, and Duarte, 2010, pp. 269-276; Sy, Duarte and Dal Bianco, 2018). In (Müller, Jahnke, Smith, Storey, Tilley & Wong, 2000, pp. 47-60) ontologies are not mentioned. Despite efforts made in the field of reverse engineering (Gómez-Pérez and Ma, 1999, pp. 139-156), ontologies’ storage within DB repositories has not been sufficiently investigated. In a previous work (Sy, Duarte and Dal Bianco, 2018), the meta-model for DB and ontologies designed in a single lifecycle can efficiently ensure the reverse engineering of both DB and domain ontology by using materialized views because of its high semantic-based abstraction degree. Indeed, the better way of achieving this is to store the roles’ names in the underlying database (Sy, Duarte and Dal Bianco, 2018), even though UML does not provide an “explicit marking of two arbitrary associations as inverses of each other” (Zedlitz and Luttenberger, 2014).

Running example: Let $U = \{R, E, L\}$ where, respectively, R, E, and L represent Rice Seed, Cereal Seed and Plot-land, and U is the universe of discourse of the application domain. Let VARIETY (V), STAMP (S), PURITY (P) and GERMINATION (G) be the properties of E, as shown at Figure 1.

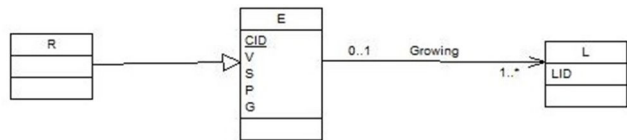


Figure 1. UML-based semantic model of the universe U

Knowing the value of S and P, the database manager may update column G of the underlying table E. using the following instruction:

```
UPDATE E SET G=93 WHERE (S='J') AND (P='A');
```

Moreover, knowing that rice seed is a cereal seed, one would like to propagate those update to all data related with the updated table. This is feasible thanks to the capabilities of relational databases' management systems (RDBMS) through triggers, based on keys and foreign keys. A general and simplified syntax of such triggers is as follows:

```

CREATE TRIGGER <trigger_name>
BEFORE | AFTER INSERT | UPDATE | DELETE
ON <table_name> FOR EACH ROW
<trigger_body>

```

Therefore, let R_i (K_i, A_i, \dots, B_i) and R_j ($K_j, B_j, \dots, B_j, \#K_i$) be two n -ary relations that represent the concepts C_i and C_j of a given UML-based semantic model, with $C_i \leq C_j$ meaning that C_j inherits of all the properties and features of C_i . Clearly, $T_j \subseteq T_i$, where T_i and T_j are respectively instances of R_i and R_j .

Now, let denote by M_v (CID, V, G, LID) a materialized view where LID represents the plot-land identifier. Accordingly, a state of M_v contains the instances of all cereal seeds and plot-lands where they have been seeded.

The posed problem is how reverse engineering – rebuilding – the UML-based semantic model of the universe U . Without loss of generality, the most reliable solution resides on the use of ontologies; with the assumption that there exists a well-designed meta-model that link the DB to the ontologies. Accordingly, we propose here a set of conceptual guidelines for schemas' reverse engineering by using ontology-based materialized views (OBMMV4RE).

The rest of the paper is organized as follows. In Sections 1, 2 and 3, we give a theoretical background comprising conceptual foundations of the notions of view and of materialized view, the notion of ontology, and their use in schemas reverse engineering process and reengineering, as well. Section 4 presents the related work. In Section 5, we discuss the OBMMV4RE approach based on a set of conceptual guidelines. A case study is presented in section 6. Finally, in Section 7, we conclude and present the direction of our future work.

1. DATA VIEW VS. MATERIALIZED VIEW

A typical data view is a simple named SPJ query to be executed by the RDBMS. Such named query, consisting of a set of n -ary tuples, is each time accessed by a user, and recomputed. For example, the code (under PostgreSQL) below returns, if any exists, the name N , the germination G and the cultivated area of all rice seeds and the plot of lands where they are growing.


```
SELECT "N" As "Name", "LID" AS "Plot land", "G" As Germination, "Cultivated_area" FROM "Seed", "Cultures"
WHERE "Seed"."CID"="Cultures"."CID";
```

The relational data view (Codd, 1970, 377-387) is shown at Figure 2.

Data Output				
	Name text	Plot land integer	germination integer	Cultivated_area numeric(6,2)
1	Oryza sativa	3	92	7.00
2	Oryza sativa	1	95	3.50
3	Oryza sativa	2	95	2.00
4	spanish	4	93	1.00

Figure 2. Data view dvGrowingSeeds

Now, let assume that a new tuple of an existing seed variety, e.g. corn, has been sowed in the plot of land number 5. Thus, whenever these data are added, to know the current state of the cultivated cereals and the related data (e.g., plot of lands and cultivated areas), the user must re-run the query dvGrowingSeeds above. Yet, according to rule 6, ‘all views that are theoretically updatable must be updated by the system’ (Codd, October 14, 1985; Codd, October 21, 1985; Rick, 2014). Thus, materialized views can be useful by caching the result of queries and allow refreshing them, once a user needs to access their content.

As example, the dvGrowingSeeds data view, theoretically updatable, and that we call mvGrowingSeeds, is materialized as follows:

```
CREATE VIEW public."mvGrowingSeeds"
WITH (check_option=local) AS
SELECT "N" As "Name", "LID" AS "Plot land", "G" As Germination, "Cultivated_area" FROM "Seed", "Cultures"
WHERE "Seed"."CID"="Cultures"."CID";
```

The above view is easily maintainable up to date based on foreign keys and triggers. However, difficulties can occur for the practical up-datability with respect to the data and meta-data. Moreover, when facing with more complex queries, which are frequent in the advanced and sophisticated information systems such as the data warehouses and or Business Intelligent environments, the answer may take more time. Thus, novelty in the methodologies, rather than in the concepts, is neces-

sary for efficient solutions. Accordingly, our proposal is the use of ontology-based materialized views not for overcoming the complex queries but especially for ensuring reuse and reverse engineering for recovering or improving conceptual schemas.

As example, let us assume that seed of wheat (W) have to be added (See Figure 3, dashed line) and, by the same time, a new tuple of sowed wheat is added. Given the sub-schema that describes the context, mvGrowingSeeds is still updatable, and from it is possible to rebuild the semantic model of Figure 1.

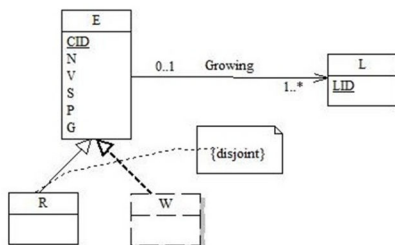


Figure 3. UML-based semantic model of the universe U (updated)

2. REVERSE ENGINEERING

According to Rick (Rick, 2014), there is no universally accepted definition of software re-engineering. However, there is a confusion between reengineering and ‘reverse engineering’ (Chikofsky and Cross, 1990). Indeed, ‘reverse engineering’ originates from the analysis of hardware. Precisely, it was defined as the ‘‘process of developing a set of specifications for a complex hardware system by an orderly examination of specimens of that system’’ (Chikofsky and Cross, 1990). In the field of software systems, several formulations have been proposed for defining the notion of reverse engineering. As examples, in (Gómez-Pérez and Ma, 1999, pp. 139-156), reverse engineering is defined as ‘‘a process of analysing a subject system to identify the system’s components and their interrelationships and to create representations of the system at a higher level of abstraction’’. In (Chikofsky and Cross, 1990), it is defined as ‘‘the process of analysing an existing system in order to identify its components and their interaction, illustrating it in a different

or more abstracted form''. According to (Reynares, Caliusco and Galli, 2015), a reverse engineering approach includes the extraction of information from source codes, and documentations; the abstraction of the extracted information; and the representation of the above abstraction.

In this work, we consider reverse engineering as the process, which aims to produce a conceptual model based on the code in which the ontology has been implemented. Moreover, we emphasize that reverse engineering is the opposite process of forward engineering that is traditional process for schemas design starting from the requirements analysis to the internal level. For a better understanding, borrowing the schema given in (Chikofsky and Cross, 1990), these two processes are illustrate at Figure 4.

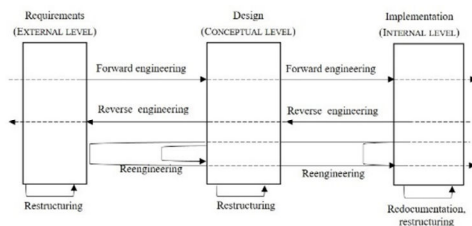


Figure 4. Reverse engineering and related processes between or within abstraction levels

Accordingly, reengineering is a process between or within abstraction levels while reverse engineering is a process that covers the entire three-level architecture, starting from the internal level up to the views' level.

3. ONTOLOGY-BASED MATERIALIZED VIEW

In the Web Semantic community literature, ontology is defined as a "formal, explicit specification of a shared conceptualization" (Studer, Benjamins and Fensel, 1998, pp. 161-197). However, there is no consensually accepted definition of the notion of ontology. Thus, in this paper, an ontology is understood as "a formalization of the universe of the discourse as an organized structure, and constrained by a set of axioms, according to the knowledge domain" (Sy, Duarte and Dal Bianco, 2018). More precisely, such formalization is the set $O = \{\Sigma, \tau \leq, A, \Omega, \rho, \phi\}$,

where: Σ is a set of concepts $\{C_1, \dots, C_n\}$ assigned with a taxonomic partial order relation $\tau \leq C_i, C_j$; Aa finite set of attributes $\{A_i, \dots, A_m\}$ describing the concepts $C_i \ 1 \leq i \leq n$; Ω is a set of relations; ρ a relation over Ω assigning to each $R_i \in \Omega$ a domain $\rho \text{ dom}: R_i \rightarrow \Sigma \times \Sigma$ and a range $\rho \text{ range}: R_i \rightarrow \Sigma \times \Omega$; and ϕ a set of axioms which hold with the concepts in Σ and the relations over Ω .

Accordingly, an ontology-based materialized view (OBMV), can intuitively be seen as a view built on the basis of O such that it contains semantic-based knowledge, e.g. rice seed is a cereal seed, or rule-based knowledge, e.g. a plot of land is a “rice crop”, a “farm land” or a “research station”. Now, since views are built upon concept-based relations over Ω which involve attributes of A , let F_c be the FROM clause of the SPJ query. We define an OBMV as follow.

Definition – A view V materialized with the CREATE VIEW instruction is an OBMV if and only if F_c satisfies the following condition:

$F_c \supset R_i \in \Omega \wedge C_j \in \Sigma \mid \tau \leq C_i, C_j \vee \tau \leq C_j, C_i$. i.e. F_c contains at least one concept-based relation and it exist a concept C_j such that C_j subsumes C_i or conversely.

Proposition 1 – Based on the properties of the subsumption relation $\tau \leq$, an OBMV is an updatable view.

Corollary 1- A well-defined OBMV is an OBMV such that its F_c contains two concept-based relations $R_i \ R_j \wedge C_i, C_j \in \Sigma \mid \tau \leq C_i, C_j \vee \tau \leq C_j, C_i$.

Yet, the semantic model M of an ontology contains at least one taxonomic relation.

Corollary 2 – The semantic external model MOV of an OBMV can be rebuilt using reverse engineering based on properties of the subsumption relation $\tau \leq$.

Corollary 3 – The ontology-based model MO of a given universe Σ can be reconstructed by reverse reengineering according to a unique well-defined OBMV and MOV.

Indeed, on the basis of $R_i (K_i, A_i, \dots, A_k)$ and $R_j (K_j, B_i, \dots, B_l, \#K_i)$ that represent C_i and C_j , with $C_i \leq C_j$ (or conversely), the underlying UML-based sub-model can be redesigned by using reverse engineering. Moreover, such inheritance deriving from the taxonomic relation $\tau \leq$

intrinsically implements the well-known functional dependencies (FD) in RDBs. Yet, any FD graph constitutes a theoretic access structure to the DB content, including the met-data of the RDBMS. Accordingly, the relational schema can be rebuilt.

4. RELATED WORK

As previously emphasized, only (Glimm, Kazakov, Liebig, Tran, and Vialard, 2014, pp.180-195; Volz, Staab & Motik, 2003, pp. 707-724; Volz, Staab & Motik, 2005, pp.1-34) tackled ontologies' materialization. However, materialized views were not used for reverse reengineering. In (Glimm, Kazakov, Liebig, Tran, and Vialard, 2014, pp.180-195), reasoning over “large ABoxes” by “abstraction refinement” was the focus. However, the content – state – of an ontology ABox, obviously, is larger than the number of axioms in its TBox, which stands as conceptual schema. Besides, the comparison between “the number of different concept names” and “the number of different individuals” is fuzzy.

Incremental maintenance of materialized ontologies was discussed in (Volz, Staab & Motik, 2003, pp. 707-724; Volz, Staab & Motik, 2005, pp.1-34). The work in (Volz, Staab & Motik, 2003, pp. 707-724) also focused on reasoning over ontologies through materialized views. However, the work is especially declined for the Web Semantic ontologies, namely RDF/RDFS and the well-known Ontology Web Language (OWL). In (Volz, Staab & Motik, 2005, pp.1-34), the authors are mainly interested on ontology-based materialization views, but the technical approach is based upon logic databases. Nevertheless, the ontology-based materialized views are not used or related to schema reverse engineering, which is the main goal of this paper. Accordingly, we propose a set of conceptual guidelines for OBMMV4RE with the aim to create the missing link between the use of materialized views, ontologies, and reverse engineering.

5. OBMMV4RE: TOWARD A SET OF CONCEPTUAL GUIDELINES

In the aim to develop an efficient methodology for reverse engineering using ontology-based materialized views, we propose a first set of useful conceptual guidelines as follows:

- Guideline 1: The DB catalog contains meta-data related to the ontology meta-model, which completes and enriches the Db schema.

Such ontology meta-model called KBSM (Knowledge-based Semantics Data Model) was proposed in (Sy, Duarte and Dal Bianco, 2018).The KBSM (See Figure 5) is a common meta-schema establishing a logical link between RDBs and ontologies.

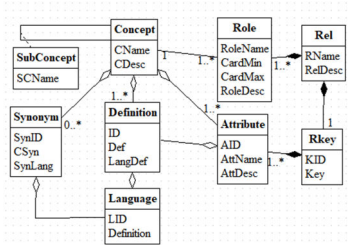


Figure 5. Knowledge-based Semantics Data Model (KBSM) of Domain ontologies

- Guideline 3: Based on Corollary 2, building of a semantic external model of M^{OV} , by reverse engineering.
- Guideline 4: According to Corollary 3, and based on the meta-data of the DB catalog enriched by the KBSM whose stored instances contain, re-build the global model M^O .

6. OBMMV4RE: A CASE STUDY EXAMPLE

An overview of our application domain (certified cereal seeds) is given at Figure 6. In this context, e.g., given a variety V_i of certified rice seed, one would like to know its breeder (maintainer) and the plot of land where it is growing or where it has been produced.

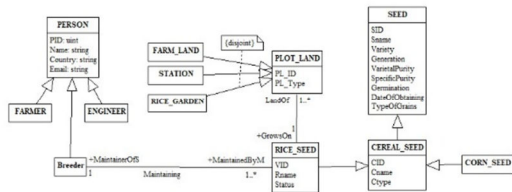


Figure 6. A case study: an overview

Accordingly, let mvBreeders be the materialized view that stores all farmers (name, country, e-mail) who are rice seed breeders (rice name). The SQL code is the following:

```
CREATE VIEW public."mvBreeders"  
WITH (check_option=local) AS  
SELECT "Name", "Country", "Email", "RName" FROM "Breeder"  
As "B"  
WHERE EXISTS (SELECT 1 FROM "Rice_Seed" AS "R"  
WHERE "R"."PID"="B"."PID" AND "B"."TYPE"='F');
```

As the relation Breeder in the FROM clause Fc of the mvBreeders view is concept-based and it is subsumed by the concept Person, mvBreeders is an OBMV. Thus, mvBreeders is updatable. Furthermore, based on the KBSM's DB (See Figure 5) and corollaries 2 and 3, the semantic model of Fig. 6 can be rebuilt.

CONCLUSION

In this paper, we tackled a tricky problem in information systems and software systems maintenance, namely data and meta-data reuse, especially schemas reverse engineering based on ontology-based materialized views.

Hence, we first have shown that most of the works, which have dealt with these topics, were inclusive, i.e. either ontologies or materialized views were left out when reverse engineering was on demand. Moreover, we clarified one again the use and the signification of the term 'reverse engineering' often confused with reengineering.

Finally, in our aim to improve the process of reverse engineering, we proposed a set of conceptual guidelines as useful methodology for overcoming the existing drawbacks.

However, actually these guidelines are not implemented. Accordingly, the next step in our future work is the experimentation of their theoretical foundations based on CASE tools that allow reverse engineering of advanced conceptual schemas as ontology-based ones.

REFERENCES

1. Alberts, R., & Franconi, E. (2012). "An Integrated Method using Conceptual Modelling to generate an Ontology-based Querying Mechanism". In *OWL: Experiences and directions workshop*.
2. Anas Shatnawi, Abdelhak-Djamel Seriai, Houari Sahraoui (2016). "Recovering software product line architecture of a family of object-oriented product variants", In *The Journal of Systems and Software* (2016), <http://dx.doi.org/10.1016/j.jss.2016.07.039>.
3. Anas Shatnawi, Abdelhak-Djamel Seriai, Houari Sahraoui, Zakarea Alshara (2016). "Reverse engineering reusable software components from object-oriented APIs", In *The Journal of Systems and Software* (2016), <http://dx.doi.org/10.1016/j.jss.2016.06.101>.
4. Asunción Gómez-Pérez, Ma Dolores Rojas-Amaya (1999). *Ontological Reengineering for Reuse*, D. Fensel and R. Studer (Eds.): EKAW'99, LNAI 1621, pp. 139-156, 1999, ©Springer-Verlag Berlin Heidelberg 1999.
5. Christoph Koch, Daniel Lupei, Val Tannen (2016). "Incremental View Maintenance For Collection Programming", *PODS'16*, June 26-July 01, 2016, San Francisco, CA, USA 2016 ACM. ISBN 978-1-4503-4191-2/16/06, DOI: <http://dx.doi.org/10.1145/2902251.2902286>.
6. Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377-387.
7. Codd, E.F. (1985). "Is Your DBMS Really Relational?". *Computerworld*, October 14, 1985.
8. E. Chikofsky and J. Cross (1990). "Reverse engineering and design recovery: A taxonomy.", *IEEE Software*, 7(1):13-17, January 1990.
9. E. Reynares, M.L. Caliusco, M.R. Galli (2015). "A set of ontology design patterns for reengineering SBVR statements into OWL/SWRL ontologies", *Expert Systems with Applications* 42 (2015) 2680-2690.
10. Edgar Frank Codd, "Does Your DBMS Run By the Rules", *ComputerWorld*, 21 octobre 1985.
11. Glimm B., Kazakov Y., Liebig T., Tran TK., Vialard V. (2014). *Abstraction Refinement for Ontology Materialization*. In: Mika P. et

- al. (eds) The Semantic Web – ISWC 2014. ISWC 2014. Lecture Notes in Computer Science, vol 8797. Springer, Cham.
12. Gómez-Pérez A (1999). Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases, Banff Knowledge Acquisition for Knowledge-Based Systems (KAW'99), Proceedings of the Banff Knowledge Acquisition for Knowledge-Based Systems, Workshop. KAW'99 Volumen: II, Editor: Gaines, B.R.; Musen, M., Editorial: University of Calgary, Alberta, Canadá, 16-21 de octubre de 1999., Banff, Alberta, Canadá., Pages: 6.1.1 - 6.1.18.
 13. Jesper Zedlitz and Norbert Luttenberger (2014). "Conceptual Modelling in UML and OWL-2", International Journal on Advances in Software, vol 7 no 1 & 2, year 2014, <http://www.iariajournals.org/software/>.
 14. Markus Brenner and Birte Glimm (2017). "Incremental Materialization Update via Abstraction Renement", In the Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, July 18-21, 2017.
 15. Müller, H. A., Jahnke, J. H., Smith, D. B., Storey, M. A., Tilley, S. R., & Wong, K. (2000, May). "Reverse engineering: A roadmap.", In Proceedings of the Conference on the Future of Software Engineering (pp. 47-60). ACM.
 16. Oumar Sy, Denio Duarte, Guilherme Dal Bianco (2018). "Ontologies and Relational Databases Meta-Schema Design: A Three-Level Unified Lifecycle", 5th International Conference on Control, Decision and Information Technologies, CoDIT'2018 10-13 April, Thessaloniki - Greece.
 17. Oumar Sy, Moussa Lo, Denio Duarte (2010). "Integrating Ontologies in Database Scheme: Ontology-Based Views Conceptual Modeling", In the Proceedings of the 6th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS, Kuala Lumpur, Malaysia, December 15-18, 2010, pp. 269-276.
 18. Prabhot Kaur Chahal, Amritpal Singh (2014). "Software Reuse and Reengineering: With A Case Study", In International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-6, January 2014.

19. Qi Zhang, Andreas Karcher (2012). "System Reverse Engineering to Requirements and Tests", The Seventh International Conference on Systems ICONS 2012, Copyright (c) IARIA, 2012. ISBN: 978-1-61208-184-7 35.
20. Rick F. van der Lans (2014). "Is Actian PSQl a Relational Database Server?", A Technical Whitepaper, March 2014, Copyright © 2014 R20/Consultancy, all rights reserved, http://www.pervasive.com/Portals/55/documents/WP_IsPSQLaRelationalDatabaseServer_R20_March2014.pdf
21. Rudi Studer, V. Richard Benjamins, Dieter Fensel (1998). Knowledge Engineering: Principles and methods, Data & Knowledge Engineering, 25 (1998) 161-197, © 1998 Elsevier.
22. Sonja Risti, Slavica Aleksi, Milan Celikovi, Vladimir Dimitrieski, Ivan Lukovi (2014). "Database reverse engineering based on meta-models", Cent. Eur. J. Comp. Sci. • 4(3) • 2014 • 150-159, DOI: 10.2478/s13537-014-0218-1.
23. Volz, R., Staab, S., & Motik, B. (2003, November). Incremental maintenance of materialized ontologies. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 707-724). Springer, Berlin, Heidelberg.
24. Volz, R., Staab, S., & Motik, B. (2005). "Incrementally maintaining materializations of ontologies stored in logic databases". In Journal on Data Semantics II (pp. 1-34). Springer, Berlin, Heidelberg.

International Conference
INFORMATION AND COMMUNICATION
TECHNOLOGIES IN
BUSINESS AND EDUCATION
Conference proceedings

Предпечатна подготовка *Мария Янчева*

Дадена за печат 14.X.2019 г. Печатни коли 25,62
Излязла от печат 17.X.2019 г. Издателски коли 24,33
Формат 60×90/16 Тираж 50

Издателство „Наука и икономика”
Икономически университет – Варна
ул. „Евл. Георгиев” 24

Печатна база на ИУ – Варна

ISBN 978-954-21-1004-0

ISBN 978-954-21-1004-0