University of Novi Sad
Technical faculty "Mihajlo Pupin" Zrenjanin

Proceedings of the 10th International Conference on Applied Internet and Information Technologies AIIT 2020

16 October, 2020, Zrenjanin, Serbia

AIIT
International Conference

University of Novi Sad
Technical faculty "Mihajlo Pupin"
Zrenjanin, Republic of Serbia

X INTERNATIONAL CONFERENCE ON APPLIED INTERNET AND INFORMATION TECHNOLOGIES

AIIT2020

PROCEEDINGS

October 16, 2020
Zrenjanin, Serbia

Publisher and organizer of the conference:
University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia

For publisher:
Dragica Radosav, PhD, Professor
Dean of Technical faculty "Mihajlo Pupin", Zrenjanin,
Republic of Serbia

Conference chairman:
Dalibor Dobrilovic, PhD

Conference co-chairmen:
Andrijana Bocevska, PhD
Evgeny Cherkashin, PhD

Proceedings editors:
Dalibor Dobrilovic, PhD – main editor
Evgeny Cherkashin, PhD
Andrijana Bocevska, PhD

Cover design:
Višnja Ognjenović, PhD

Technical preparation of proceedings:
Dalibor Dobrilovic, PhD
Maja Gaborov, MSc
Milica Mazalica, MSc

e-Proceedings
ISBN 978-86-7672-342-3

CIP - Каталогизација у публикацији
Библиотеке Матице српске, Нови Сад

004(082)(0.034.4)

INTERNATIONAL Conference on Applied Internet and Information Technologies (10 ; 2020 ; Zrenjanin)
Proceedings [Elektronski izvor] / X International Conference on Applied Internet and Information Technologies AIIT 2020, Zrenjanin, October 16, 2020. - Zrenjanin : Tehnički fakultet "Mihajlo Pupin", 2020. - 1 elektronski optički disk (CD-ROM) ; 12 cm

Naslov sa nasl. ekrana. - Bibliografija uz svaki rad.

ISBN 978-86-7672-342-3

а) Информационе технологије - Зборници

COBISS.SR-ID 27477513

INTRODUCTION

The objectives of the International conference on Applied Internet and Information Technologies are aligned with the goal of regional economic development. The conference focus is to facilitate the implementation of Internet and Information Technologies in all areas of human activities. The conference provides a forum for discussion and exchange of experiences between people from government, state agencies, universities, research institutions, and practitioners from industry. Information technologies change during time and this year AIIT conference addressed the diversity of ICT application areas and relevant research topics such as:

- Information systems
- Software engineering and applications
- Data science and big data technologies
- Business intelligence and IT support to decision-making
- Communications and computer networks
- Data and system security
- Distributed systems
- Internet of Things
- Embedded systems
- Software quality
- Software maintenance
- Computer graphics
- IT management
- E-commerce
- E-Government
- E-Education
- Internet marketing
- ICT practice and experience

Information technologies enable collaboration across the globe. This year the conference was successfully co-organized by 5 institutions from 4 countries - Serbia, North Macedonia, Russia, and Bulgaria. It has been managed in collaboration with 3 co-chairmen from Serbia, North Macedonia, and Russia.

International Conference on Applied Internet and Information Technologies (AIIT) is an annual conference that was held since 2012, based on successful results of the International Conference on Information and Communication Technologies for Small and Medium Enterprises in 2011. This year, AIIT2020 was held on October 16, 2020, in Zrenjanin, Serbia.

Due to a COVID-19 pandemics, the conference is held in virtual form, with online presentations with Google Meet, and streaming video and poster presentations available at the web site of the conference (http://www.tfzr.uns.ac.rs/aiit/). There were 54 accepted papers with 125 authors from 21 countries (Serbia, North Macedonia, Montenegro, Bosnia and Herzegovina, Croatia, Slovenia, Romania, Hungary, Bulgaria, Slovakia, Russia, Sweden, United Kingdom, USA, Canada, India, Sri Lanka, Japan, China, Egypt, and Iraq). The papers are presented online, or in the video stream and poster sessions. Within the video presentation session, there is a presentation of IT company ACS – Advanced Cyber Security, Belgrade, Serbia.

The AIIT 2020 organizing committee would like to thank the authors of the papers for their contribution. All submitted papers were peer-reviewed by the members of the AIIT2020 program committee. Each submitted paper was assigned to at least two reviewers from different countries and the paper analysis was conducted as a double-blind review.

Special gratitude is addressed to many reviewers from co-organizing institutions that made a great impact on the quality of papers. The AIIT organizing committee especially appreciates the IT company's efforts in supporting the conference by its participation.

Information technologies are integrated with every human activity. IT application enhancements are encouraged by university research, business organizations, public institutions, and the IT industry. The AIIT organizing committee welcomes future presentations of work in this field at the next AIIT conference, hoping that all of us will meet again in the real conference event.

Conference chairs:

Dalibor Dobrilović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

co-chairman Evgeny Cherkashin, *Institute of High Technologies, Irkutsk,Russia*

co-chairman Andrijana Bocevska, *Faculty of Information and Communication Technologies - Bitola, North Macedonia*

Main Organizing Institution

Technical Faculty "Mihajlo Pupin" Zrenjanin
University of Novi Sad
SERBIA
http://www.tfzr.uns.ac.rs/

Organizing Partner Institutions

Faculty of Information and Communication Technologies - Bitola
"St. Kliment Ohridski" University - Bitola
North MACEDONIA
http://www.fikt.edu.mk/

Irkutsk National Research Technical University
Institute of High Technologies, Irkutsk
Matrosov Institute for System Dynamics and Control Theory SB RAS
Irkutsk, RUSSIA
http://www.istu.edu/

Irkutsk State Transport University
Irkutsk, RUSSIA
www.irgups.ru/en/about-university

Faculty of Engineering
South-west university "Neophyte Rilsky"-Blagoevgrad
BULGARIA
http://www.swu.bg/

Conference Supporting Institutions

Municipality of Zrenjanin, Serbia
Regional Chamber of Commerce, Zrenjanin, Serbia
Regional Center for development RCR Banat, Zrenjanin, Serbia
Zrenjaninski IKT Klaster, Zrenjanin, Serbia
Business Incubator, Zrenjanin, Serbia

Companies Presented at AIIT2020

ACS – Advanced Cyber Security
Belgrade, Serbia
https://www.acs.co.rs

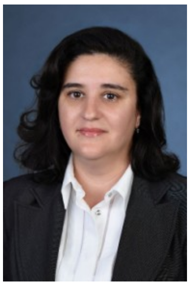
CONFERENCE CHAIRS

Dalibor Dobrilović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

co-chairman Evgeny Cherkashin, *Institute of High Technologies, Irkutsk,Russia*

co-chairman Andrijana Bocevska, *Faculty of Information and Communication Technologies - Bitola, North Macedonia*

Dr. Dalibor Dobrilovic is an associate professor at the Information Technology department at the University of Novi Sad, Technical Faculty „Mihajlo Pupin", Zrenjanin. He has received his PhD in Information technology in 2012. His teaching areas are in the field of Computer Networking, Communication Systems, and Data and computer systems security. His research interests are in the area of IoT, Smart Cities, Wireless communications, Wireless sensor networks, Computer Networking, Engineering education, etc. Dr. Dalibor Dobrilovic has more than 120 research articles published in international journals and conferences and he has participated in several EU and national funded projects. He is a member of IEEE and ACM societies. Since 2019 he is the president of the Council of the Technical Faculty "Mihajlo Pupin" Zrenjanin.



Dr. Andrijana Bocevska is an Associate Professor at the Faculty of Information and Communication Technologies, "St. Kliment Ohridski" University – Bitola, R. North Macedonia. She received her MSc and PhD degrees in Mechanical Engineering in December 2001 and October, 2012, respectively. Her research areas include: Integrated computational methods and applications, Computer integrated manufacturing, Product engineering, technology and systems. Dr. Andrijana Bocevska has published 6 books and more than 35 research articles published in international journals, conferences and congresses and she has participated in several EU and domestic funded projects. Dr. Andrijana Bocevska currently teaches subjects in: Application software, Solid modeling, Computer integrated manufacturing, Scientific visualization in virtual environments, Product Lifecycle Management. Associate Professor Andrijana Bocevska was appointed to the position of the Vice-dean for teaching and international cooperation on 01 March 2018.

Dr. Evgeny Cherkashin has graduated from Irkutsk State Technical University at 1996, at 1999 defended dissertation "Quant/2 system for automatic theorem proving" on application new logical calculus for control technical systems. After that, he mostly deals with application first-order logical inference systems for model identification algorithm synthesis, software model transformations. Most of the scientific activity is carried on in Institute for Systems Dynamics and Control theory of Siberian Branch of Russian Academy of science, at Laboratory of Complex information systems. E.Cherkashin instructs students of two Irkutsk universities programming, software design, real-time system engineering and artificial intelligence. He is author more of 160 scientific papers.

TABLE OF CONTENTS

APPENDIX

INVITED PAPERS

Training in Parallel and Distributed Computing Based on the Orlando Tools Framework

S.A. Gorsky, A.G. Feoktistov
Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia
gorsky@icc.ru, agf@icc.ru

Abstract - The paper presents new methods and tools for training parallel and distributed computing. We use the Orlando Tools framework for training users of various categories in designing and applying scientific applications in a heterogeneous distributed computing environment. The main attention is focused on the scalability of computations depending on the problem formulation, problem-solving scheme (workflow) selection, continuous integration of applied software, environment configuration, and fault tolerance. Orlando Tools provides hands-on training skills for end-users. Application of training is shown by the example of solving practical problems.

I. Introduction

Parallel and distributed computing is an important area of Computer Science Curricula [1]. Some curricula related to challenges that need High-Performance Computing (HPC), but not all of them have special courses dedicated to parallelism.

Experience in teaching students shows that training is essential in studying practical issues of parallel and distributed paradigms of programming. As a rule, the training is based on the methods and tools used in practice. Thus, the inclusion of such training oriented to developing and using parallel and distributed software into the education process significantly improves an understanding of many aspects in developing real scalable scientific applications.

Nowadays, specialists in subjects domains create complex problem-oriented applications based on workflows to carry out large scientific experiments in distributed environments. In such environments, they obtain, save, manage, and process big data [2]. Environments can be heterogeneous and combine computational resources from different platforms (grid, cloud, public access centers, etc.).

A workflow is defined as a specification of stages in executing and managing scientific applications [3]. Usually, a modular approach is applied in the workflow design. Each computation process stage in scalable scientific applications may require HPC.

Organization of efficient parallel and distributed computing is a non-trivial problem. Therefore, there is a need for the use of specialized tools.

A Workflow Management System (WMS) is system software for specifying, managing, and executing workflow processes on computational resources [4]. A large spectrum of the well-known WMSs are used in practice. Among them are Askalon, Condor DAGMan, Pegasus, Taverna, etc. Many scientific problems are successfully solved applying WMSs [5-8].

The other two well-known approaches that can be used to create scientific applications are proposed to help end-users with the development and use of problem-oriented software:

- Problem Solving Environments (PSE) [9, 10],
- Integrated Development Environments (IDE) [11, 12].

PSE is a specialized applied software oriented to solving specific classes of problems by combining automated problem-solving methods with human-oriented tools for guiding the problem resolution.

IDE is oriented to programmers for developing, modifying, compiling, and debugging software for mathematical modeling in problem-solving. It is based on the integration of general-purpose facilities such as Eclipse or Visual Studio.

We use the Orlando Tools framework (OT) for the development of distributed applied software packages. A set of scalable science applications is a kind of such packages [13]. OT implements the main features of WMSs and systems designed within the framework of PSE and IDE.

Table I summarize the main supported capabilities of the compared approaches. The markers ‘+’ and ‘–‘ mean that all systems supported the compared approach have and do not have the marked capabilities, respectively. The marker ‘+/–’ means that some systems may have and some do not have them.

The use of OT gives its end-users encouragement to experiments. An important challenge for our training of scalable application is the integration computing skills and problem-oriented knowledge.

Hands-on training the skills of distributed computing in these approaches is based on various computational experiments in a computing environment [14]. The training level is determined by a spectrum of tools for the computational processes study.

TABLE I. SUPPORTED CAPABILITIES

Capability	Approach			
	WMS	***PSE***	***IDE***	***OT***
Conceptual programming support	+/–	+/–	+/–	+
Procedural problem formulation	+	+	+	+
Non-procedural problem formulation	+/–	–	–	+
Static problem decomposition	+	+	+	+
Dynamic problem decomposition	+/–	+/–	+/–	+
Tools for continuous integration	–	–	+/–	+
Workflow execution monitoring	+	+	+	+
Resource monitoring	+	+/–	+/–	+

The classical distributed computing learning involves the study of special courses on mathematical foundations of parallel computing, architectures of parallel computing systems, technologies of parallel programming and parallel algorithms [15-17]. Usually, problems of the learning are the integration of this knowledge and full-scale hands-on training.

In this paper, we represent the OT applying as an experimental training environment to teach the scalable application development and use.

II. SCALABLE SCIENTIFIC APPLICATIONS

A scalable application focuses on parallelism provided by the growth in the number of computational units (cores, processors, nodes, etc.) of its execution environment.

The collaborative team of mathematicians, applied and system programmers and resource administrators design the scalable scientific application. As a rule, it includes the follows applied and system software:

- Library of applied programs for problem-solving with varying degrees of detail in formulating problems,
- Library of system programs for supporting computation planning, jobs forming for workflows executing under the LRMs control, computing monitoring, etc.,
- Library of auxiliary programs for data pre-processing and post-processing.

A significant characteristic of computation scaling is a problem-solving makespan. In scalable applications, it decreases with increasing the number of computational nodes used.

A modular approach use provides the opportunity to decompose a scalable problem into sub-problems, which can be solved independently in a computing environment. This is especially evident within the framework of parameter sweep computations [18]. Moreover, the environment can be distributed and heterogeneous.

Scalability depends of both the properties of applied software and hardware characteristics including features of system software located in environment nodes. An environment can consists specialized computational nodes or Personal Computers (PCs), PC-clusters, HPC-servers, HPC-clusters, etc. The owners of such computing systems are interested in their efficient and balanced use.

Therefore, when using scalable applications, it is important to understand in what cases computation speedup and efficient use of resources is ensured.

III. ORLANDO TOOLS

OT is intended for describing and executing problem solving processes (workflow) using distributed applied software packages. Packages run in a heterogeneous distributed computing environment. OT supports version control based on continuous integration, delivery, and deployment.

The packages are related to intelligent systems based on subject knowledge. In addition to knowledge about the subject domain and algorithms for problem-solving, they include: tools of collecting, storing, and analyzing information about the computing processes execution and environment resources based on their monitoring.

OT includes the following key components:

- Conceptual model designer for describing the subject domain of a package in the text or graphic form (Fig. 1),
- Designer of libraries for applied modules of a package,
- Tools for continuous integration of package software for debugging and testing modules and workflows, controlling their versions, and placing software on computational resources of an environment,
- Configurator of a computing environment, which provides connection of computing resources and their setting,
- Repositories, knowledge bases and databases for saving and processing package descriptions, data, and computations results.

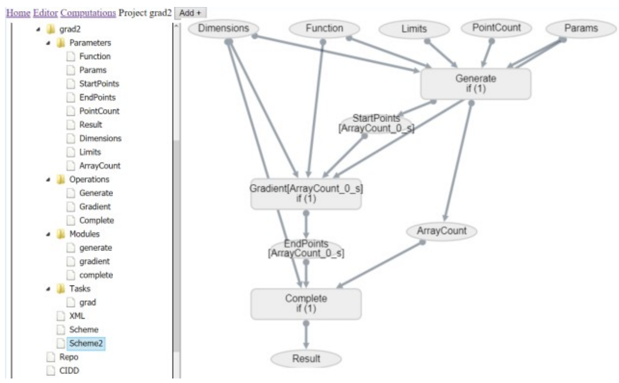


Figure 1. Screenshot of interface for the conceptual model designer

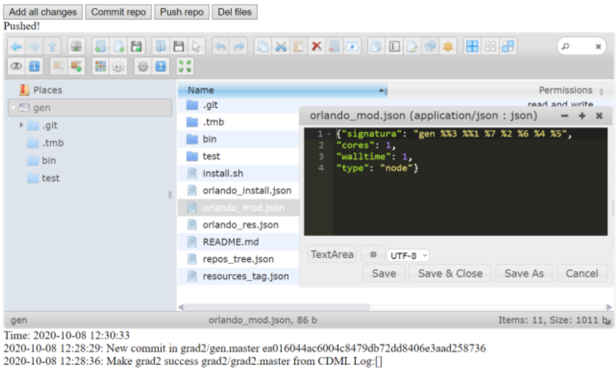


Figure 2. Screenshot of the software manager interface

A key feature of OT is continuous integration of package software. Using the GitLab system allows us to manage versions of modules, workflows, and packages in the whole. Support of their debugging and testing in nodes of the environment with different computational characteristics can significantly reduce the package development times, increase the reliability of computations, and decrease the makespan of large-scale scientific experiments. The screenshot of the software manager interface is shown in Fig.2.

IV. TRAINING

A scheme of hands-on training in OT is shown in Fig. 3. Our hands-on training is oriented to various categories of end-users:

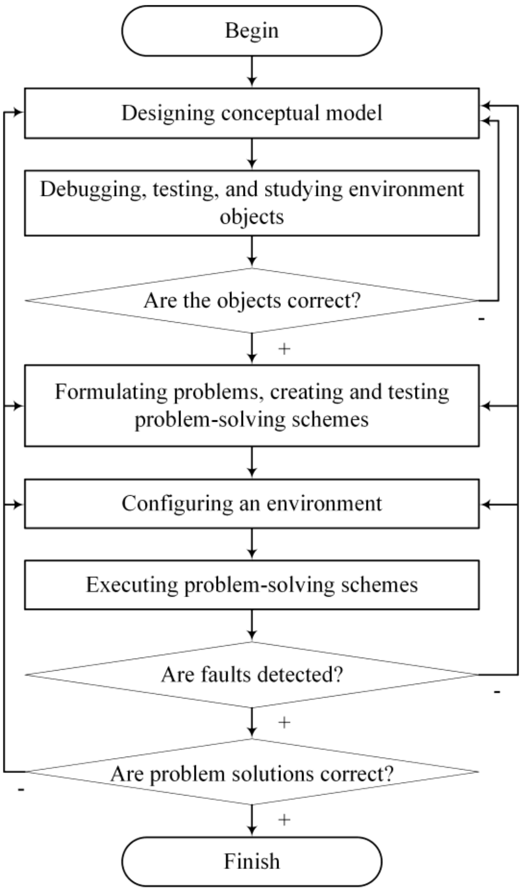


Figure 3. Trainning scheme

- Application developers,
- Environment administrators,
- Application users.

Developer training may include all stages of training.

Administrator training involves tools for configuring the environment and testing its nodes.

Application user training supposes operating with the conceptual model of application and carrying out computational experiments in a heterogeneous distributed computing environment.

Users train on the standard models and algorithms with default parameters of experiments from the OT knowledge. We provide them with two educational applications for searching the global minima of multi-extreme functions [19] and optimizing warehouse logistics [20].

Advanced users may develop their own models and algorithms for more complex problems. In addition, they may plan experiments with special parameters.

The following OT components are used in training (Table II):

TABLE II. OT COMPONENTS USED IN TRAINING

Component	Training	Practical Skills
Conceptual model designer	End-users study structural analysis of a subject domain for the solved problem. They learn to identify important objects of a subject domain, define relations between them, formulate problems, and form problem-solving schemes (workflows). The results of structural analysis, formulations and problem-solving schemes are reflected in a graphical editor.	Experience in constructing computational models, formulating procedural and non-procedural problems, forming problem-solving schemes in accordance with the specified criteria.
Subsystem for continuous integration of applied software on resources	End-users train to operate with repositories for storing and controlling software versions, tools for debugging and testing software in heterogeneous nodes of the environment.	Skills in managing software versions in heterogeneous distributed computing environments.
Environment configurator	Through configuring the computing environment, end-users can study how each configuration affects the computation speedup and resource use efficiency for a specific problem.	Knowledge about configuring computing environments for solving specific problems, taking into account their properties.
Computation Manager	Operation with the executive subsystem allows end-users to obtain information about the work with LRMs queues in environment nodes, job priorities, computational history of solving problems, and features of providing resource quotas.	Knowledge about the LRMs operation features and administrative policies for different environment resources.

- Conceptual model designer,
- Subsystem for continuous integration of applied software on resources,
- Environment configurator,
- Executive subsystem.

In Table II, we briefly describe the training process and its result (obtained practical skills).

The training allows end-users to significantly reduce uncertainties in selecting a high number of criteria in problem-solving (algorithms, values of their control parameters, inputs, problem-solving schemes, resources, etc.). The main aim of such selection is to optimize distributed computing.

The success of the OT training in the designing and applying of scalable science applications is due to several factors. OT supports team training, in which end-users of different categories interact at the intersection of their subject domains. The OT components provide end-to-end training from the application design to practical problem-solving.

V. Example

In this Section, we consider a problem of modeling loading and unloading logistic operations at a warehouse. This problem is one of the problems solved with the aforementioned educational application [20].

Depending on the detail of the problem formulation, it may require experiments with a different number of variants of the initial data. Workflow for solving the problem includes the modules for data preprocessing, simulating warehouse operations through parameter sweep computations, selecting optimal data variants, and data postprocessing. The module for carrying out parameter sweep computations is implemented in the General Purpose Simulation System (GPSS) language [21].

The computing environment includes PC (Intel Core i3-4160, 2 core with hyper-threading, 3.6 GHz, 4 GB RAM), PC-cluster (16 nodes with 1 processor Intel Core i3-4000M, 2 core with hyper-threading, 2.4 GHz, 2 GB RAM), and HPC-cluster (32 nodes with 2 processors AMD Opteron 6276, 16 core, 2.3 GHz, 64 GB of RAM). All above-listed computing systems have different computational characteristics. There are the following resource use quotas for one experiment: 10 nodes for a PC-cluster and 20 nodes for a HPC-cluster.

In the paper, we skip the training stages at the design and continuous integration of applied software since we are using the educational application.

Therefore, the purpose of the training is to show end-users the change in the computation speedup with the different environment resources when the detail of the problem formulation changes. A more detailed formulation requires the study of a larger number of variants of the initial data and leads to a greater load on resources.

VI. Empirical Research

In Table III, we can see the results of experiments carried out by end-users. These results are generated for users by the OT executive subsystem based on computational history of the workflow execution.

TABLE III. Experiment Results

Experiment	*n*	t_1	t_2	t_3	s_1	s_2
1	30697	32.18	4.57	0.12	7.04	268.17
2	61422	64.36	9.11	0.23	7.06	279.83
3	122843	128.73	18.21	0.46	7.07	286.07
4	245701	257.48	36.53	0.90	7.07	289.30
5	498739	513.95	67.63	1.77	7.06	290.73

In the experiment specification, end-user can determine alternative resources, including a basic resource, for the workflow execution with the speedup calculating in relation to the basic resource.

Variants of initial data and their number are determined in a job for the workflow execution. Thus, Table III provides the results of five experiments for one end-user's problem with the different number of initial data variants.

The columns of Table III show the values of the following parameters:

- Number n of variants,
- Experiment makespan t_1 on the PC,
- Experiment makespan t_2 on the PC-cluster,
- Experiment makespan t_3 on the HPC-cluster,
- Speedup s_1 with PC-cluster,
- Speedup s_2 with PC-cluster.

The computation speedup with the PC-cluster grows with an increase in the number of data variants up to a certain point (Experiment 3, $n = 122843$). Then the growth stops and begins to decrease. At the same time, the computation speedup with the HPC-cluster monotonically increases with the number of data variants. This is due to the high-performance of the cluster nodes and the higher network bandwidth in comparison with the PC-cluster.

This short training demonstrates the capabilities of different computing systems when performing distributed computing. In addition, end-users can distinctly see how the detail of the problem formulation affects the experiment makespan. For example, they can formulate problems related to first four experiments and carry out them only on HPC-cluster, when the time limit for problem-solving is 1 hour.

VII. Conclusion

Nowadays, applying the potential of HPC allows us to effectively solve difficult scientific and practical problems. At the same time, this requires special knowledge and skills in this field. To this end, we propose new methods and tools for designing scalable scientific applications and training of their use in heterogeneous distributed computing environments.

The proposed methods and tools are implemented in the Orlando Tools framework. Applied specialists, programmers, administrators, and end-users with different levels of skills and needs collaborate in an integrated process of design and use of distributed applied software packages (scientific applications).

The proposed framework is used for supporting various academic disciplines within scientific specialty 05.13.11 «Mathematical and software of computers, complexes and computer networks» for postgraduate students at the Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences.

It is important to note that such training within classical learning is not easy, because it is difficult to organize solve the real large-scale problems for students simultaneously. Our training allows us to individually improve student skills in parallel and distributed computing. Thus, such training is a good supplement to the classical education program.

ACKNOWLEDGMENT

The study is supported by the Basic Research Program of the Siberian Branch of the Russian Academy of Sciences, project no. IV.38.1.1.

REFERENCES

[1] Computer Science Curricula 2013. Joint Task Force on Computing Curricula, ACM Press and IEEE Computer Society Press, 2013.

[2] D. Hollinsworth, “The workflow reference model,” in Workflow Management Coalition, Document Number TC00-1003, vol. 19, no. 16, pp. 224, November 1994.

[3] J. Yu, R. Buyya, “Scheduling scientific workflow applications with deadline and budget constraints using genetic algorithms,” Sci. Program., vol. 14, no. 3-4, pp. 217–230, December 2006.

[4] G. Alonso, H. J. Schek, “Research issues in large workflow management systems,” in Proceedings of the NSF Workshop on Workflow and Process Automation in Information Science, NSF, 1996, pp. 126–132.

[5] D. Talia, “Workflow Systems for Science: Concepts and Tools,” Int. Sch. Res. Notices, vol. 2013, Article ID 404525, pp. 1–15, January 2013.

[6] X. Li, J. Song, B. Huang, “A scientific workflow management system architecture and its scheduling based on cloud service platform for manufacturing big data analytics,” Int. Adv. Manuf. Tech., vol. 84, no. 1-4, pp. 119–131, April 2016.

[7] J. T. Krogel, “Nexus: A modular workflow management system for quantum simulation codes,” Comput. Phys. Commun., vol. 198, pp. 154–168, January 2016.

[8] F. Marozzo, D. Talia, P. Trunfio, “A workflow management system for scalable data mining on clouds,” IEEE T. Serv. Comput., vol. 11, no. 3, pp. 480–492, July 2016.

[9] J. R. Rice, R. F. Boisvert, “From Scientific Software Libraries to Problem-Solving Environments,” IEEE Comput. Sci. Eng., vol. 3, no. 3, pp. 44–53, 1996.

[10] C. Johnson, C. Johnson, S. Parker, D. Weinstein, S. Heffernan, “Component-based, problem-solving environments for large-scale scientific computing,” Concurrency and Computation: Practice and Experience, vol. 14, no. 13-15, pp. 1337–1349, November-December 2002.

[11] H. J. Bungartz, P. Neumann, W. E. Nagel. Software for Exascale Computing-SPPEXA 2013-2015. Lect. Notes Comp. Sci., Springer, Cham, vol. 113, 2016.

[12] V. Il'in, “Artificial Intelligence Problems in Mathematical Modeling,” Comm. Com. Inf. Sc., vol. 1129, pp. 505-516. 2019.

[13] A. Feoktistov, S. Gorsky, I. Sidorov, I. Bychkov, A. Tchernykh, A. Edelev, “Collaborative Development and Use of Scientific Applications in Orlando Tools: Integration, Delivery, and Deployment,” Comm. Com. Inf. Sc., vol. 1087, pp. 18–32, February 2020.

[14] M. Murshed, R. Buyya, “Using the GridSim Toolkit for Enabling Grid Computing Education,” in Proceedings of the Int. Conf. on Communication Networks and Distributed Systems Modeling and Simulation, Society for Computer Simulation, 2002, pp. 27–31.

[15] L. Ferreira, F. Lucchese, T. Yasuda, C. Y. Lee, C. A. Queiroz, E. Minetto, A. Mungioli. Grid Computing in Research and Education. IBM redbooks, 2005.

[16] N. Stojanovic, E. Milovanovic, “Teaching introductory parallel computing course with hands-on experience,” Int. J. Eng. Educ., vol. 31, no. 5, pp. 1343–1351, 2015

[17] D. J. John, S. J. Thomas, “Parallel and Distributed Computing across the ComputerScience Curriculum”, in Proceedings of the Workshop on Parallel and Distributed Computing Education (EduPar-14), IEEE, 2014, pp. 1085–1090.

[18] R. Buyya, M. Murshed, D. Abramson, S. Venugopal. “Scheduling parameter sweep applications on global Grids: a deadline and budget constrained cost-time optimization algorithm,” Software Pract. Exper., vol. 35, no. 5, pp. 491–512, January 2005.

[19] I. V. Bychkov, G. A. Oparin, A. N. Tchernykh, A. G. Feoktistov, S. A. Gorsky, R. Rivera-Rodriguez, “Scalable Application for the Search of Global Minima of Multiextremal Functions,” Optoelectron. Instrum. Data Proces., vol. 54, no. 1, pp. 83–89, April 2018.

[20] A. Tchernykh, A. Feoktistov, S. Gorsky, I. Sidorov, R. Kostromin, I. Bychkov, O. Basharina, A. Alexandrov, R. Rivera-Rodriguez, “Orlando Tools: Development, Training, and Use of Scalable Applications in Heterogeneous Distributed Computing Environments,” Comm. Com. Inf. Sc., vol. 979, pp. 265-279, March, 2019.

[21] T.J. Schriber, Simulation using GPSS. New York: John Wiley, 1974.

Collaborative robot applicability analysis on the place of the manual welder

A. Kafi*, Z. Nyikes**, T. A. Kovács **
* Óbuda University / Doctoral School on Safety and Security, Budapest, Hungary
** Óbuda University / Donát Bánki Faculty of Mechanical and Safety Engineering, Budapest, Hungary
abdallah.kafi@uni-obuda.hu; nyikes.zoltan@uni-obuda.hu; kovacs.tunde@uni-obuda.hu

Abstract - The collaborative robots applications are widely increasing in the industrial application. In the welding process, the robot application area is continuously growing. The manual welding is a professional but really hard physical work, the replacing by collaborative robots will be a big advance for the industry. The collaborative robots application in the welding tasks is not solved yet.

In this article, the authors want to analyse by risk assessment the dangers for the human in the welding workplace. On the base of the results determines the danger zones of the collaborative welder robot workplace on the base of the dangers level and kind. Also in this applicability analysis, the authors count the requirable sensors for the collaborative welder robot in the welding to assure labour safety.

On the base of the labour safety requirements the aim of this work to establish a technical aspect system to found the launch of the collaborative robot welding for the industrial application.

I. Introduction

The fusion welding is a professional work when the metal melting on high temperature and the melted metal establish a metallic cohesion joint [1, 2]. This process is widely used in the great number of tasks for metal manufacturing. On the base of the heat source during the fusion welding process, it can determine several dangers. Labour safety requirement the supporting of human welder safety and healthy by the minimization of the dangers. The ergonomic manual welding is almost impossible, the labour safety tries to find the solution to prevent accidents and health damages [3].

II. The Collaborative robots

The collaborative robots in our age are widely used for several workplaces. The robotic rules are applied for them also. Asimov defined the robotic rules in his famous novel [4]. The robotics increased rapidly in the last decade and we learned to live and work together with them. The first robots usually some automats were [5]. They made their program continuously without any communication with their environ observes the rules of Asimov. The nowadays used robots are different, they can communicate with their environ. The robots use sensors to pick information from their environ. The sensor technology is also increasing with the robotics to support the robot designer pretensions. Collaborative robots continuously monitoring environmental data [6].

The available sensors number almost uncountable. The sensor technology base to use the picked physical properties data of the environ (f.ex. temperature, waves, currents, etc.). These are the heat sensor, the moving sensor, light sensor, arc sensor etc. Also the motion of some robots and collaborative vehicles supported by radar and GPS technology [7]. The sensor, radar, GPS and wireless technologies are necessary to build a suitable and safe robot system [8]. The base of the robot programming is the coordinate geometry. The collaborative robots are safe for a human co-worker, means that the robot and the humans work together in the same work area [9]. The human-robot collaboration levels shown in Fig.1. In the case of the lowest level collaboration when the human and robot workplace is separated but they work in the same all. Middle-level human-robot interaction when human and robot are sharing the workplace, but not the same time. Highest level of the human-robot interaction is the real collaboration when the workplace is fully shared and the contact not only possible but desired.

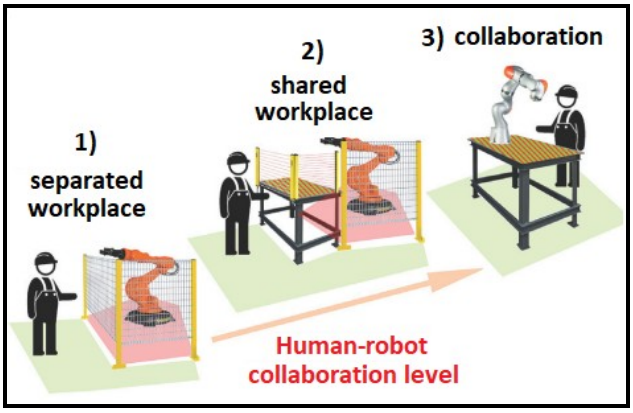


Figure 1. Integration levels of the human-robot collaboration [9]

As a part of Industry 4.0. strategy the industrial work needs to be automated with the collaboration of the robots and human workers. The collaborative robot's task supported by several sensors. For the robot control, the used sensors kind and sensitivity is very important, it needs to be suitable for the robot task and security requirements [10]. The suitable collaborative robot system (sensors included) environment needs to be fully secured. The industrial robots stipulated by the ISO standard 10218 [11-13]. The robot's control in the case of the collaborative robots realizes by wireless technology. This technology assures suitable, rapid data transfer between the robot and the controller. Unfortunately, industrial robots wireless data transfer is vulnerable [14]. The robot's programming and control as a function of the robot function are almost solved, it can find several examples and manual in the literature [15, 16].

Nowadays it can find several robots in the welding tasks. Most commons are the spot welder robots in car industrial processes, and also it can find welder robots in the fusion welding area too. Commonly the automatised fusion welding is the Gas Metal Arc Welding (GMAW) process [17, 18]. The guiding and control of this welding process by a computer program is solved.

Hence the aim of the collaborative robot applicability analysis on the place of the manual welder is a complex work to establish a technical aspect system what include the dangers and requirements of the welding and the robotics.

III. COLLABORATIVE WELDER ROBOTS

The manual welding is hard physical work for the human welder. The welded joint quality depends on the welder knowledge and experiences. The manual welder is limited by his physical and environmental conditions. The welding speed, the applicable power source namely the productivity of the welder is limited by these parameters. The industry is expecting higher and higher productivity what is impossible to perform by manual welding. Also, labour safety requirements are increased to save workers health. In this aspect truly understandable the expectation to replace the manual welders by collaborative welder robots. These collaborative welder robots need to be integrated between manual welders and welding inspectors. To satisfy this expectation it needs to define the probable dangers for human during the welding task of the collaborative welder robots.

A. The danger of arc welding

During the GMAW process, it can find several dangers from the process specification and the process of metal transfer. The base of the arc welding is the melted metal transfer between the electrode and the weld metal pool to establish a metallic joint. The melting of the metal made by an electrical arc. In the high-temperature electrical arc, any metal can be melting. The melted metal temperature is much over than the melting point of the metal. The melted metal during the metal transfer process is covered by shielding gas to isolate the melted metal from the environmental pollutions.

Dangers of the GMAW process:

- ✓ Heat (electrical arc heat, high-temperature product)
- ✓ UV light (electrical arc)
- ✓ Spattering (melted metal drops)
- ✓ Fume (established gas mix from the metal component and shielding gas)
- ✓ Robot „arm" (movement of the robot)

On the base of the international and national laws in the case of any welding, manufacturing needs to observe the Welding Safety Regulation requirements. This regulation contains the rules for automated welding workshop. Regulated the distance between the robot standard and moving parts and the workshop walls, pillars and other devices. On the workplace of the robot can stay only the educated operator. To enter in the robot risk zone to enter forbidden. The configuration of the robot environs it needs to be suitable to observe these rules. In the robot risk zone, only the robot maintainer staff can stay during robot installation, calibration and maintenance. The collaborative human worker needs to wear protective clothes, gloves and helmer like in the case of manual welding. The GMAW robot operator can serve and/or collaborate in the work of the robot only in the dangerous light separated area. The GMAW used electrical arc dangerous light emission (ultraviolet (UV) and infrared (IR)) depends on the welding parameters. The light spectrums are shown in Fig. 2. between 750 nm to 400 nm. The visible light spectrums area for a human is limited. Also, dangerous UV and IR lights are defined for human eyes and skin.

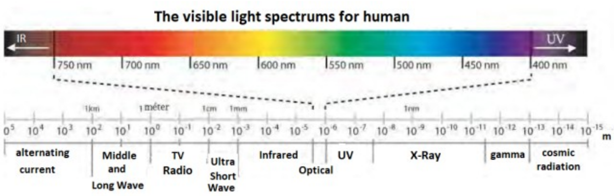


Figure 2. The visible light spectrums for human [19]

The operator and the GMAW robot can be working together in the case of suitable air exchanging or air ventilation. In the case of any alarm, the operator needs to go to the assigned safe area [19].

B. Safety of collaborative robots

The standard ("ISO/TS 15066 Robotics and robotic devices - Collaborative robots,") contain the safety requirements for collaborative industrial robot systems [20]. The standard declares four important safety-related monitorable components of the collaborative human-robot work. The most important two components are the "speed and separation monitoring" in the human-robot collaboration. Continuously needs to assure the safe separation distance between the robot and human during the collaborative work. The second two components are the "power and force limiting" limiting the robot's transfer of pressures and forces onto the human body [21, 22].

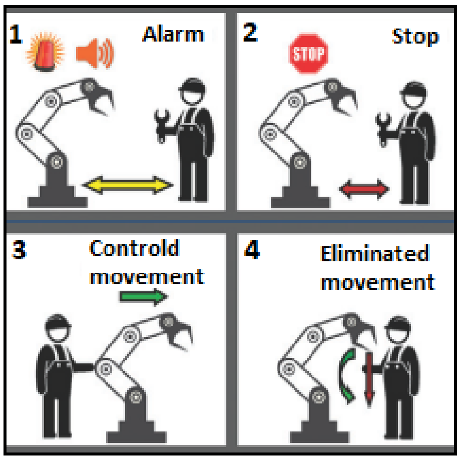


Figure 3. The reaction of the robot in the case of foreign object detection [9]

The collaborative robot by several sensors continuously monitoring of his work area. In the case of the detection of any foreign object, it needs to respond to the event. Fig. 3. shown some robot reaction in the case of foreign object detection. The robot can react by light and sound alarm (1) when detecting the unsuitable distance between itself and the foreign object. In the case of dangerous distance, the robot stops itself (2). The foreign object (operator) can the movement pushing off (3). The robot can react in the case of the foreign object detection by to controlled collision eliminated movement (4).

IV. SAFE COLLABORATIVE GMAW ROBOT WORK AREA

On the base of the collaborative robot safety standard and the risks of the GMAW it can recommend a safety rule system for collaborative welder robots workplace configuration.

The aim of the automatization for welding, to apply the robots on the place of the manual welders. The robots must work as a manual welder. The human has six senses of monitoring the environ around himself. Human is continuously deciding about his operation. Human will not stop his welder task when any foreign object is approaching his workplace.

A manual welder can be more dangerous than a robot for foreign humans because he wants to do a quality welded joint without any break because the welding task requires determined length welding to manufacture a suitable quality welded joint. It can't use the speed, power and force limitation in the welding manufacturing, because any limits during the welding process cause substandard product. This is the reason why it needs to support the undisturbed welding time. Breaking in the welding task cause unsuitable joint. It needs to configure the welder robot workplace on the base of the (ISO) regulation and support the welding quality availability.

The Asimov rules are suitable in robotics, robots can't effect any damage to humans not even than the welded joint will be unsuitable.

On the base of the robotics safety standards and the welding process technicalities, it needs to prevent the foreign human entering the welding zone. To support the collaborative welder robot in its undisturbed working, it can define risk zones. Fig. 4. The collaborative robot must monitor not only its workplace but must monitoring also the extended zone of its workplace.

The extended zone means a prevention zone. If the robot detects any foreign object in the extended zone, it can modify the foreign object movement with sound and lighting call. Also, the robot can do a difference between a foreign human or a foreign device entering. In the case of a foreign device, the robot doesn't need to stop its work, because the heat, UV light, fume and spatterings are usually not dangerous for technical devices or the protection is solved.

To realize the monitoring of the robot environs, the robot needs to use several sensors for a continuous collection of environmental data. The robot operation will spend on the collected and processed data. In the case of any foreign object located in its workplace area will block the operation.

In the case of a human infiltrator, the robot needs to warn the human to leave the extended work zone. The robot must control the human moving in the extended zone and continuously warn in the case of the approaching. Therefore the robot by sensors data needs to differentiate human and device in its working area.

In the case of the continuous human approaching, on the base of the measured average walking speed, the robot can calculate the arrival time to the dangerous working zone. Even that robot continuously warn the human about his unsuitable, dangerous movement, the robot can continuous its welding work to the arrival time to the danger zone.

The moment that the human enters the danger zone, the robot needs to stop welding to protect human safety. During this process, the robot has a chance to wake the human to modify the moving up to the entering for the danger zone while the robot may finish the welding task.

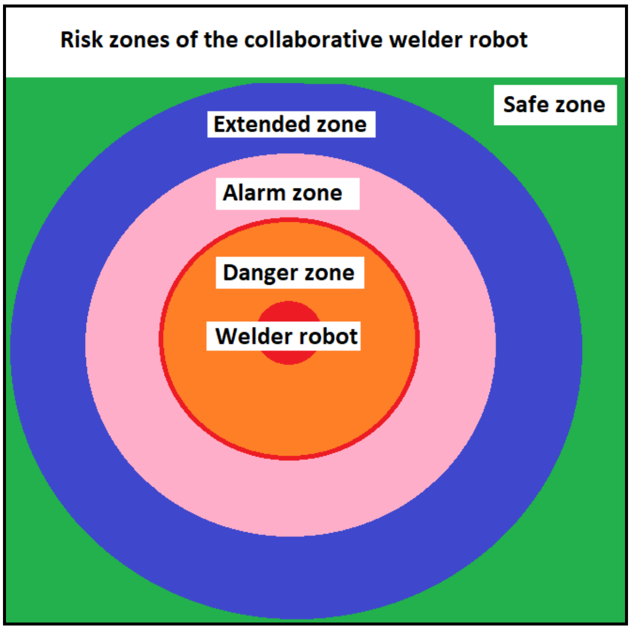


Figure 4. The collaborative welder robot risk zones

On the base of the visible light spectrum and danger level of them, it can calculate a safe distance from the emission source. The safelight distance is the distance between the human and the light emission source (welding arc) where the light intensity is enough low and can't cause any damage for human eyes and skin. The most dangerous lights are UV and IR.

Danger zone: the diameter of the danger zone D_D (m) calculated from the robot maximal arm reach A_L (m), the safe light distance L_{UV} (m) and the safety coefficient L_{S1} (m).

$$D_D=2*(A_L+L_{UV}+L_S) \text{ (m)} \quad (1)$$

Alarm zone: the diameter of the alarm zone D_A (m) calculated from the diameter of the danger zone D_D (m) and the average human walking speed v_H (m/s), the reaction time t_R (s) and safety coefficient L_{S2} (m).

$$D_A=D_D+2*(v_H*t_R+L_{S2})\ (m) \quad (2)$$

Extended zone: the diameter of the extended zone D_E (m) calculated from the diameter of alarm zone diameter D_A (m) and a safety coefficient L_{S3} (m).

$$D_E=D_A+2*L_{S3}\ (m) \quad (3)$$

The determined zones diameters depend on the robot maximal arm reach A_L (m) because the safe UV light distance L_{UV} (m) is constant.

V. Conclusions

The collaborative robots developments supported by the various sensors can enable the possibility of the collaborative welder robot availability. The welding task is metalworking where automatization facilitates and speeding up the process. It would be a great advantage for the industry if on the place of the human welder it can apply collaborative welder robots.

It can conclude a rule, that it needs to keep out the foreign human in the welding work. To realize this rule the collaborative robot needs to recognize the foreign human and use tools to keep out him to enter the dangerous area.

In this work, authors determined the necessity of the risk zones in the case of the collaborative welder robots. It can conclude that the calculation of the risk zones diameters depend on the maximal robot arm reach and the human walking speed and reaction time.

One part of the research work focuses to determine the safe (UV and IR) light distance and the safety coefficient in the different zones. To realize this plan it needs to do several tests in the industrial workplaces and analyse the human-robot interactions.

The collaborative welder robots application on the place of the manual welders will available soon, but to assure the welding quality and also the safety of the human co-workers needs to standardize the collaborative welder robot environ configuration.

Acknowledgement

The authors acknowledge the financial support of this work by the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 project.

References

[1] Kou, Sindo, Welding Metallurgy, John Wiley & Sons, Inc., Hoboken, New Jersey 2003

[2] D.H. Phillips, Welding engineering and introduction, John Wiley & Sons, Ltd, West Sussex 2016

[3] C.M. Schlick, Industrial Engineering and Ergonomics, Springer, Aachen 2009, pp.763-768

[4] J. Nagler, J. van den Hoven, D. Helbing, An Extension of Asimov's Robotics Laws, SSRN Electronic Journal August 2017

[5] B. Singh, N. Sellappan, P. Kumaradhas, Evolution of Industrial Robots and their Applications, International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 5, May 2013

[6] J. O'Neill, J. Lu, R. Dockter, T. Kowalewski, Stretchable, Flexible, Scalable Smart Skin Sensors for Robotic Position and Force Estimation, Sensors, MDPI 2018. https://doi.org/10.3390/s18040953

[7] J. Raible, M. Blaich, O. Bittel, Differential GPS supported navigation for a mobile robot, IFAC Proceedings Volume 43, Issue 16, 2010, pp. 318-323

https://doi.org/10.3182/20100906-3-IT-2019.00056

[8] V. Matellan, J. Marıa Canas, O. Serrano, WiFi localization methods for autonomous robots. 2006 Robotica. 24., pp. 455-461. http://doi.org/10.1017/S0263574705002468

[9] A. Vysocky, P. Novak, Human-Robot Collaboration In Industry, MM Science Journal, 2016 June, DOI: http://doi.org/10.17973/MMSJ.2016_06_201611

[10] International Federation of Robotics: Robots and the Workplace, Federation of Robotics, Frankfurt, Germany March 2018

[11] [EN ISO10218-1] EN ISO10218–1, Robots for industrial environments – Safety requirements- Part 1. ISO, Geneva (2011).

[12] [EN ISO2018-2] EN ISO2018-2, Robots and robotic devices – Safety requirements for industrial robots – Part 2. Robot systems and integration (2011).

[13] [EN ISO12100] EN ISO12100, Safety of machinery General principles for design – Risk assessment and risk reduction. ISO, Geneva (2010).

[14] R.E. Mahan et al., Secure Data Transfer Guidance for Industrial Control and SCADA Systems, Prepared for the U.S. Department of Energy under Contract DE-AC05-76RL01830, Pacific Northwest National Laboratory, 2011

[15] P. Lin, G. Bekey, K. Abney, Autonomous Military Robotics: Risk, Ethics, and Design, US Department of Navy, Office of Naval Research 2008

[16] K. Nilsson, R. Johansson, Integrated Architecture for Industrial Robot Programming and Control, Robotics and Autonomous Systems (1999) 29(4), pp. 205-226. https://doi.org/10.1016/S0921-8890(99)00056-1

[17] Z. Hou, Y. Xu, R. Xiao, et al., A teaching-free welding method based on laser visual sensing system in robotic GMAW. Int J Adv Manuf Technol 109, 1755–1774 (2020). https://doi.org/10.1007/s00170-020-05774-0

[18] Y. Xu et al., Welding seam tracking in robotic gas metal arc welding, Journal of Materials Processing Technology Volume 248, October 2017, pp. 18-30

https://doi.org/10.1016/j.jmatprotec.2017.04.025

[19] Hegesztés Biztonsági Szabályzat (Welding Safety Regulation) 143/2004. (XII. 22.) GKM rendelet

[20] ISO/TS 15066:2016: Robots and robotic devices — Collaborative robots

[21] J.A. Marvel, Sensors for Safe, Collaborative Robots in Smart Manufacturing, 978-1-5090-1012-7/17/2017 IEEE

http://doi.org/10.1109/ICSENS.2017.8234264

[22] R.N. Shea, Collaborative Robot Technical Specification ISO/TS 15066 Update, International Collaborative Robots Worshop, Robotic Industries Association

The use of various CAD / CAM software in architectural design process - users' experience

Andrej Šmid
Faculty of Civil Engineering, Transportation Engineering and Architecture,
University of Maribor, Slovenia
andrej.smid@um.si

Abstract - Since the beginning of use of CAD CAM software and computers in particular in the processes of architectural design and construction processes, the software has evolved into more and more usable and presentable products. Design as an activity to visualize ideas and concepts, buildings as virtually formed objects are nowadays presentable to the smallest detail and hence evaluated as separate presentations of the articulated ideas. This process has evolved into separate entities: the technical design is used for construction and the visualisations and rendered models as tools to present and decide for the variations of the designed buildings in advance.

I. INTRODUCTION

The tools for architectural design have altered in the last century considerably. The development of drawing tools and media was finished in the 1980es and the big switch was made from physical to virtual since personal computer was introduced into office practice.

At the beginning of the CAD software popularisation, in the end of 1980es, the things were limited to very simple computer hardware with its limitations. Getting from "graphic stations" to personal computer software, affordable to design studios, used to analogue drawing boards and rotring-to-transparent paper design, was a long and winding road to step on. The first personal computer approach to design studios was bound to DOS, simple text editors like WordStar and simple table editors like Lotus 1-2-3.

The architectural design offices and students started using CAD software with the beginning of the 1990es, when the AutoCAD[1] [1], ArchiCAD[2] [2] and similar programs were available for a reasonable price and for moderately priced personal computers. A bit later some of the offices started using Microstation[3] [3] and Nemetschek Allplan, other used a modification of AutoCAD, known as ACADBau, and they were running on pretty affordable personal computers.

There are two basic differences on how someone approach the architectural design. When architects use the drawings as a derived, abstract tool of capturing the building that is to be built and when the drawings are used mainly to instruct the building site, then the computers help the designers to produce almost the same drawings as they would using the ink tools and paper. There is but a difference to it: the alterations and changes to the drawings are time consuming and quite un-elegant - ink has to be removed, parts of the drawings cut away or erased with a razor blade, then new drawing parts are to be inserted. With the use of CAD software the drawing alterations were done much easier - the designer changed, erased and reconstructed parts of the drawing and the new variatn was printed.

The first step of the CAD software use was done with the knowledge that the designer had very little work when the contractor wanted design changes. The second very appealing function was that the software did a great job in very simply adding the dimensions to a properly constructed drawing. This was also done very neatly and accurately. To conclude this observation about "getting into the CAD process", there are very similar reasons why a designer converted to CAD constructor - it was much simpler to start a way of creating and to alter drawings.

At this point, nobody realized that there is much more to the process of creating the projects than merely creating the drawings - and this is the most important difference about architectural creativity: it is largely thought that architects create drawings in order to instruct how the buildings are constructed; during the last few decades this is proved wrong - architects prepair the creation of buildings using media as drawings and models.

II. FROM 2D DRAWINGS TO 3D MODELS

A. CAD / CAM as not suitable environment for architectural design

The great idea behind the creators of AutoCAD software was that "everything that can be drawn, can also be constructed". As a follower of various geometric and mathematic procedures, especially parallel projections, the AutoCAD version of 1988 alreday operated as a sophisticated tool for constructing proper and technically fully adequate drawings.

The most interesting designs of that time included 3D parallel projections of the St.Pauls' Cathedral of London and of the most expensive global vessel, the Space Shuttle. Both of them were created using basic 3D design tools (point, line), but with defined 3D attributes. The fascination of rotating the before mentioned virtual models caused hundreds of personal computers all over

[1] A famous product by Autodesk Inc.,
[2] A product by Graphisoft
[3] Microstation became famous in urban planning due to its ability to process complex large data, Bentley systems, Inc.,

the world to crash on a daily basis. The models were presented in two modes: as a “wireframe” and as “hidden lines” model. To “hide lines” was a process of minutes.

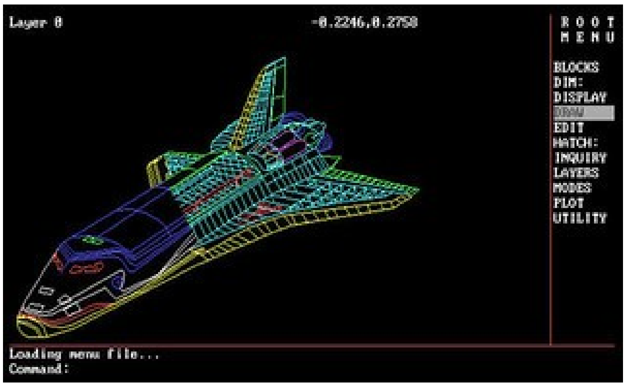


Figure 1. Screenshot by AutoCad with the 3D model of Space Shuttle, 1985, (Screenshot by the author)

AutoCAD is still in use, it is a perfect tool for creating 2D drawings since its universal ability to connect the various designers in the stage of constructing building (electrical, mechanical, HVAC, fire prevention, construction drawings). The development of creating 3D models is overwhelming, the forms derived from mathematical as well as from “organic” functions is astonishing.

B. *Architects, 3D models, perspective views*

The architects, designers of buildings, belong to a bizarre kin of engineers. Although their drawings serve the construction sites all over the world, one of the basic activities for an architect is to persuade, to impress clients.

Almost all the clients want to know how the building will look like in many details: will it correspond to their needs, will it represent their image, will it be constructed properly, will it meet their aesthetic standards. On the other hand, many clients have almost no ability to picture the building in their mind when seeing the design drawings, they are unable to connect the floor plans, sections and the facade drawings into a building in their mind. Mostly the clients can understand how the building is organized, how do the spaces, rooms and areas correspond to each other, how the building is tailored, what are the surfaces of the floors, but one cannot count on them to absorb the drawings and create a virtual model of it in their mind. That is exactly where personal computers and architectural software comes handy.

How was it done in the old days? In sixteenth century, the painters in Italy (it is disputed which church was painted with it as the first) invented the “perspective”. The human eye as a camera was studied and the perspective projection was invented. Knowing that it is a geometric procedure, beautiful, technical perspective drawings were produced for big construction ventures as soon as the 18^{th} century. In France, a special occupation was defined, and “monsieur Le Perspecteur” could produce a beautiful perspective drawing of a new mansion literally overnight.

Figure 2. A perspectivic drawing by Giovanni Battista Piranesi, 1720-1778, Cut-away view of the interior of the Basilica of S. Paolo fuori della Mura, 1749, Mount Holyoke College Art Museum, https://sites.google.com/a/mtholyoke.edu/math-120-the-mathematics-of-perspective-drawing/ Retrieved in October 2020.

The perspective projection is a part of the geometry studies at every architectural faculty in the world. It was only a question of time and effort to implement the geometrical rules of creating a perspective drawing using a smart software. The steps toward it were not in how to instruct the machine to calculate it. The problem was to teach the user that the perspective drawing is just a way to look at a 3D model. The procedure of making proper perspective drawings was hence to teach the users to think and construct in models, not only to produce drawings. It was achieved in the 1990es.

C. *Model thinking*

The first software that understood the constructional way of imagining buildings, was the Graphisoft ArchiCAD. Designed by a hungarian software company that included architects in the making, the ArchiCAD grew from the 1990es with many understandable novelties in each major version upgrade. The design process for architects was altered in the version 5.0 which introduced automatic section generation. During the learning process the ability to create models, not merely architectural drawings, was pursued. The philosophy behind working with the ArchiCAD software was pretty demanding: the user should create finely defined 3D models and then use them to cut or observe them. The procedure is so much different than creating 2D drawings that some users never used the ArchiCAD features to create models. The “model thinking” powered the architectural presentation skills: various tools were invented to alter the attributes of the ArchiCAD models and the results were exquisite - the models were fine adjusted in Artlantis[4] [4] rendering engine and the rendered perspective views of the 3D models were exquisite (it was the end of the 1990es). On the other hand, AutoCAD also invented a slightly other way of thinking: the 2D drawings were upgraded with the third dimension and the models that were produced this way, were transported to the 3D Studio[5] [5] engine where similar procedures (material finalisations) were applied and renders were calculated.

[4] Artlantis is a Abvent product since the 1990es
[5] 3d Studio and later 3DMax are Autodesk products

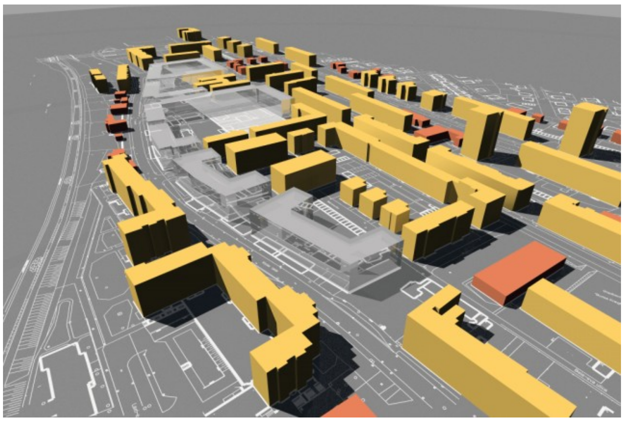


Figure 3. A 3D shaded model of the "Urban alteration study of Koroška cesta in Maribor Slovenia", render by Marko Rozman, (Komunaprojekt Maribor, 2003)

D. Model details and rendering procedures

Very soon, most architectural designers learned that the power of detail is the real secret of the 3D model quality. The models became very detailed and extremely power, time and RAM consuming. Very simple models were used for detailed urban plan presentations, but for fine architectural presentations, very fine models were developed. Hardware limits were pushed higher, but no one was thinking of dealing with the basic problems. If the models were detailed, the renders would take for hours. There were engines that helped in optimizing time in the rendering process but the models were still too consuming. The main art of the presentation was to develop a feeling of the future model use: the rough models were presented in the architectural competitions, the fine interior models were created to be applied in the interior presentations and similar. Somehow, the models failed to be realistic since the ability of the software was not sufficient for large amount of data processed.

Figure 4. A render of the "Duplek sports hall", render by Marko Rozman, Andrej Šmid and Nejc Gonza 2019 (Arhitekt Šmid, 2019)

It became clear that there is no CAD software that could help the user in all the various fields of architectural design, from creating 3D models to producing shop drawings and similar. Many of the powerful tools could do much of the needed, but there was no product that could serve all the needs. Creating 3D models and rendering them, creating short animations was one part of the architectural design and drawing plans and other drawings was quite something different.

These two productions were soon to be separated and the "monsieur le perspecteur" occupation was reborn.

In the process of architectural design two capital steps were achieved: to differ between the presentation skills and the construction drawings production. The two have been separated in order to fulfil the modus operandi of the building site on one hand and to respect the presentation procedures on the other. It seems that the software developers followed this separation. Today it is unthinkable that an office would produce shop drawings for the building site that are not perfect construction drawings. If sometimes a 3D presentation is brought to explain construction procedures, it is considered inadequate. Construction engineers at the site understand the drawings and are fully acquainted with the process of adding layers to the construction so the drawings have to show them this distinction between the construction elements and the layers added to it.

On the other hand, the clients care less for the construction procedures - they crave to see the "final outcome" of the building and are merely uninterested for the procedures that lead to it.

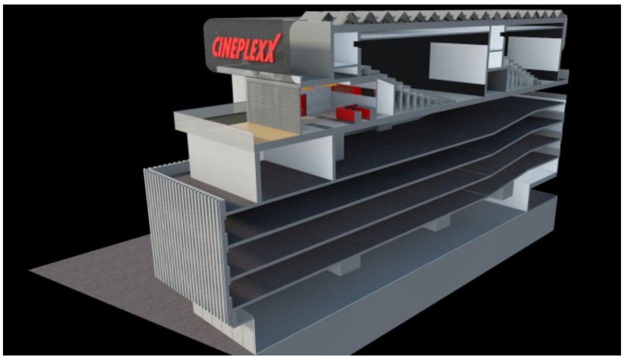


Figure 5. A cut away 3D model of the "Cineplexx Ljubljana concept design", render by Marko Rozman (Arhitekt Andrej Šmid, 2014)

For a better understanding of the outcome even the building site permit procedure is set up to include a "spatial presentation" of the building so the clerks can imagine what they are looking at and issuing a permit.

The model design achieved a new level with products that were launched a decade ago, especially the SketchUp[6] [6] software with its simplicity towards inexperienced users. Producing a model from simple volumes and using its wonderful presentation potential was a new step in presenting the architectural concepts to the client.

E. Separate presentations and new media

A very usable and affordable presentation media has occured a decade ago when the 3D printing devices became widely produced. Printing an architectural model using a thin filament was just the final step to present a 3D virtual model in a plastic form. Although time consuming, the 3D prints are the final step to present the architectural solutions in a (literally) handy manner.

[6] Created by @last software

When looking for a tool to produce shop drawings and to interchange them with other engineering coworkers, there are numerous software products that meet the needs since three decades. All of them are capable to produce decent drawings and there is hardly any need or intention of making them any better.

There are considerable changes in the 3D model productions in the last decade that cannot pass unmentioned. One is the production of more sophisticated models using new rendering engines and using a new approach to real time rendering and the other is the connection of the model to the future building management.

Figure 6. A presentation render of the renovation of the Slovenian national parliament small assembly hall, render by Nejc Gonza, Marko Rozman and Andrej Šmid (Arhitekt Šmid, 2019)

When 3D model of the building is used for an architectural presentation, there are two ways to go: until a decade ago the viewpoints were chosen and perspective views were calculated (rendered) so the final product was a perspective graphic of photographic quality which was considered (for a client) a well presented building from different angles. The use of a model was therefore limited to rendering a couple of views. In the last decade, new software products were developed with a completely new approach to 3D models: since they use calculation and rendering engines, borrowed from the simulation games, the movements of the models are run smoothly and the model rotations and view creations are quick.

The real time "passing through" the 3D model is so appealing that the recent versions of software products like Lumion[7] [7] or Twin motion[8] [8[are being used by the clients to "visit" the virtual building and exploring every area in it. Of course the software producers invented the lighter version, the "viewers" for the wide range of users. The appealing power of the "model visits" makes an architectural presentation individual for the user - it is no show anymore, the virtual building is explored by the client in an intimate manner and is similar to playing a simulation game. The presentations of the model often end in clients commands like "please go back into the main hall and turn right so we can see the view to the patio" - not to mention the presentations of the final materials: clients can choose between them, some production facilities working with wood or stone are practising huge scans of the real materials to be inserted in the final presentation models so perfect final simulations can be presented.

Figure 7. A render for the "Moto nautika" shop in Miklavž, Slovenia, a simple sketch for the investor, based on the streetview graphic data. Architecture and render Marko Rozman (Arhitekt Šmid, 2018)

The final connection of the future building was achieved through planning of all the technical facilities and devices in the future construction. All the lights, heating systems, windows, doors, entrance controls, loudspeakers and similar were joined in an intelligent system, known as BIM (Building information modelling). The architectural, construction and installations planning was upgraded in a way that the 3D model contents complete data and information about all the elements of the various systems. The BIM model is then used even after the building is completed to smartly manage its subsystems (heating, lighting, HVAC, sun shading, access control, elevators and similar). BIM planning and management was made possible with 3D modelling of the buildings and has been achieving great results for corporate users of new buildings.

Figure 8. A presentation render for a planned building in a highly conserved area of Maribor, Slovenia (Arhitekt Šmid, 2018)

III. Conclusion

A new age has begun. The architects started to develop their projects using software for 3D modelling and the clients embraced it. The procedures of architectural design have been completely altered in the last three decades. The software, developed through this time has altered the way of thinking.

The presentation and production were separated and the architects were equipped with powerful tools to express their knowledge and their ideas.

Hopefully this is not the end of the journey. The power of inventing and conceiving new ideas has not been limited - we are empowered to pursue them with adequate software support.

[7] Developed by Act-3D company
[8] Developed by Epic Games Inc

REFERENCES

[1] Autodesk Inc., AutoCAD, https://www.autodesk.com/products/autocad/overview?plc=ACDIST&term=1-YEAR&support=ADVANCED&quantity=1, Retrieved in October 2020.

[2] Graphisoft, Archicad, https://graphisoft.com/solutions/products/archicad, Retrieved in October 2020.

[3] MicroStation,https://www.bentley.com/en/products/brands/microstation, Retrieved in October 2020

[4] Artlantis, https://artlantis.com/en/, Retrieved in October 2020

[5] Autodesk, https://www.autodesk.com/products, Retrieved in October 2020

[6] SketchUp, https://www.sketchup.com/, Retrieved in October 2020.

[7] Act-3D, https://lumion.com/, Retrieved in October 2020

[8] Epic Games Inc., https://www.unrealengine.com/en-US/twinmotion, Retrieved in October 2020

REGULAR PAPERS

Developing a low-cost Smart Terrarium in the Context of Home Automation Applications

Traian Stanciu, Răzvan Bogdan and Marius Marcu*
* Politehnica University of Timișoara
Computers and Information Technology Department, Timișoara, Romania
traian.stanciu@student.upt.ro, razvan.bogdan@cs.upt.ro, marius.marcu@cs.upt.ro

Abstract – The concept of Internet-of-Things (IoT) requires the connection of heterogeneous devices with the goal of offering improved services for the users. One of the main directions in this domain is the development of Home Automation (HA) applications. This implies monitoring and controlling of different home appliances by taking into consideration IoT-specific communication technologies. Our paper presents the development of a Smart Terrarium in the context of HA user-centric applications.

I. Introduction

The "Smart City – Smart Home" paradigm is one of the central technological developments in the present society. This trend is based on the concept of Internet-of-Things (IoT) in which different classical objects (called *things*) are communicating between themselves in order to transmit data coming from a large number of sensors. The computing power is partially realized onto all these devices, therefore the *edge computing* approach coming into place.

The IoT-based smart home implementations have become a noticeable area of research [1]. The state-of-the-art scientific literature, but also industrial approaches focus on two main directions: smart home applications development and solving issues into the standards, communication protocols, but also human-communication interfaces [2].

This paper is proposing a low-cost architecture in order to implement a Smart Terrarium (ST). Our project lies in the category of Home Automation (HA) applications, being part of the smart home paradigm. The problems we are trying to address are as follow:

- How can a user grow plants in an environment that is not well suited for them because of bad lighting or humidity?
- How can the user monitor the health of the plants with real time sensor data such as temperature, air humidity, soil humidity, soil Ph and so on?
- How could a ST system ensure that the user's plants get the required amount of water?
- How could a ST system ensure that the plants remain healthy?

Home automation is a largely fragmented industry with standards being under development. Even so, the industry is rapidly expanding with the market size projected to be around $114 billion by 2025 [3]. Nowadays there are different home automation systems already implemented. These systems help people to control their houses easily and to secure them. Basically the user can control the lights, climate and the home security system by an application on mobile or tablet. Such kind of systems can be expensive. First, we should buy necessary equipment for manage the lights in the house, for example the Philips Hue starter pack costs about $200 and additional bulbs are $60 each. When we are talking about the climate control the most used and popular device is the Nest Learning Thermostat, its price being around $250 and a smart air conditioner is about $280, the most popular is Aros Smart Air Conditioner. At the security system the market offers Dropcam Pro Security Camera, which is one of the best-selling products in the smart home security category, its price being about $200. The total amount for such a system would be around 1000$. Therefore our aim is to build a low-cost solution for the ST.

The rest of the paper is structured as follows: section II will present previous work. Section III will discuss the proposed architecture and obtained results, while section IV will highlight conclusions and future work.

II. Previous Work

The problem of low-cost solutions for smart home applications has been previously addressed into the scientific literature. In [4] is presented an architecture based on STM32F407 microcontroller which is reducing the energy from a smart home with 40%. The core board of this project is a Mikromedia 7 Board and this has been implemented in Timisoara, Romania. In [5], the user interface of a home automation solution is offered via a mobile applications. This project is using Google assistance for voice command. The home appliance devices are commanded by a Node MCU V3 ESP8266 board. Different types of testing techniques have been used in order to demonstrate the correct functioning of this solution. The low-cost scenario in home automation has been addressed in [6] based on power line communication, by using Cypress CY8CPLC10. This is an integrated power line communication chip which embeds both a PHY modem and a network protocol stack. A detailed implementation of a smart home is presented in [6] by presenting a scale model that represents a smart home. In

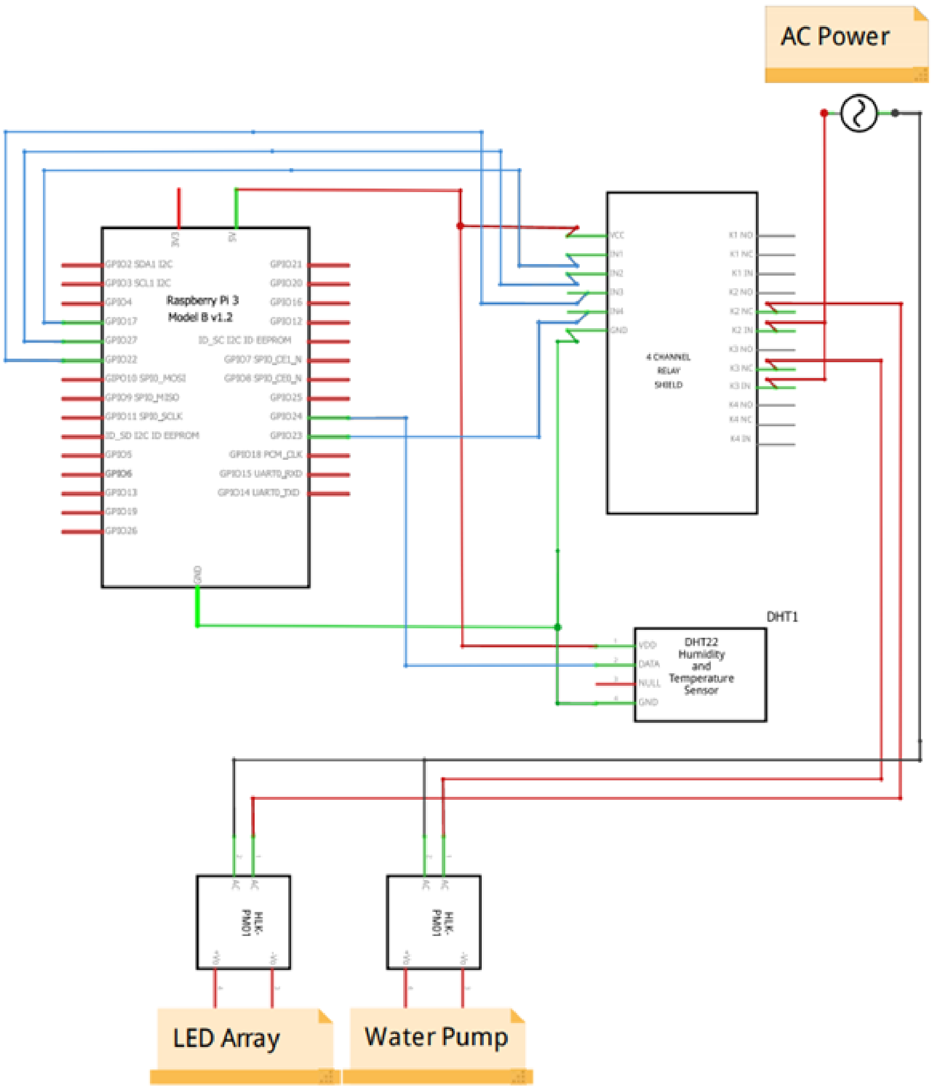


Figure 1. Hardware architecture of the Smart Terrarium

[7] we can find a home automation application for people in developing countries, while in [8] the problem of mobile user interface has been addressed in the context of a smart task scheduling technique.

III. PROPOSED SOLUTION

A. *Hardware Implementation*

The hardware implementation of the Smart Terrarium solution is based on RaspberryPi 3 model B version 1.2, being presented in Figure 1. The 4 channel relay module is fundamental to our application as it will allow us to programmatically control four different high current appliances. The module's 4 inputs will be connected to the Raspberry Pi's GPIO pins thus allowing us to control whether the relays are turned on or off depending on the signal value that we send from the Raspberry Pi's connected pins. The module offers two different functioning modes for the relay outputs:

- Normally Closed (NC) - The normally closed functioning mode does not require an input signal in order to close the circuit, thus anything connected to the NC port will automatically receive power whenever there is power in the circuit. This functioning mode is ideal for supplying power back to our Raspberry Pi, as it does not require input signal to function.
- Normally Open (NO) - The normally open functioning mode requires a high input signal in order to close the circuit and thus provide power. We will be using this functioning mode to supply power to the LEDs, water pump and cooling system.

Taking advantage of the NC, NO functioning modes we will design our circuit in order to have one supply for the microcontroller and 3 other for the equipment we will be controlling.

For the terrarium's watering system we will be using a 3-6V DC powered submersible micro water pump. The pump in question features: DC Voltage: 2.5-6V; Maximum lift: 40-110cm; Outside diameter of water outlet: 7.5mm; Inside diameter of water outlet: 4.7mm; Diameter: approx. 24mm; Length: approx. 45mm; Height: approx. 33mm. We will require a 220V AC to 5V DC to power our pump from our 4 channel relay, as we are running 220V AC. We will also require an irrigation system in order to distribute water properly through our terrarium and we will be using a dripping micro irrigation system. The pump will reside in a water filled container and when powered on push water through to our irrigation system.

In order to provide the required light for growing plants we will need a full spectrum lamp that can provide the appropriate Red / Blue colors as well as UV and IR colors. For our projects we will be using a LED lamp that features: 67 pieces red (620-630nm, 655-660nm); 15 pieces blue (440-450nm, 450-460nm); 8 pieces Warm White (3500k - 4000k); 8 pieces White (6000k); 1 pieces IR (730nm); 1 pieces UV (380-410nm). The lamp will also require a 220V AC to 50-60V DC converter in order to be powered.

For our temperature and air humidity sensor we will be using an AM2302 sensor from Adafruit. The AM2302 is a wired version of the DHT22, in a large plastic body. It is a basic, low-cost digital temperature and humidity sensor. It uses a capacitive humidity sensor and a thermistor to measure the surrounding air, and spits out a digital signal on the data pin (no analog input pins needed). It's fairly simple to use, but requires careful timing to grab data. The only real downside of this sensor is the fact that the user can only get new data from it once every 2 seconds, so sensor readings can be up to 2 seconds old.

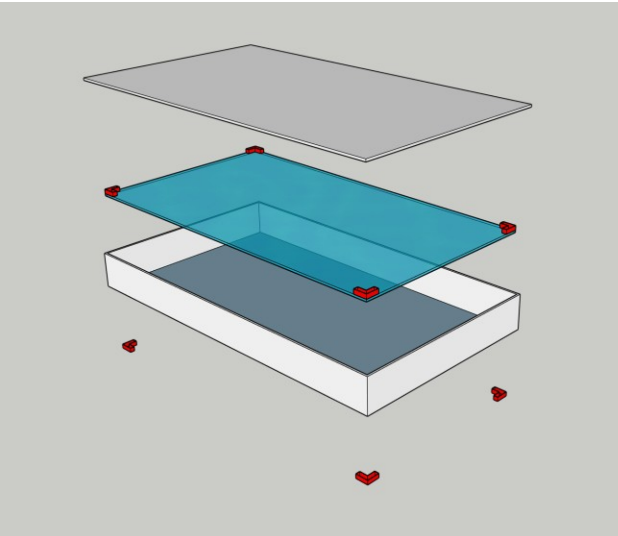


Figure 2. Terrarium Components Box

Our environment will require a (mostly) enclosed container that will be able to hold all our electronics components in a compartment separate from the one which will hold our soil and plants. A custom made acrylic sheet terrarium will be used for this project, although a glass terrarium could work as well. The main body of the terrarium is built from 5 transparent acrylic sheets made out of 5mm thick acrylic sheets with a size of W:800mm x L:450mm and H:450mm. Additionally we will have a custom made box that will act as our cover for the terrarium, which will hold all of our electrical and electronics components, as well as having proper airways for cooling.

B. Software Architecture

The client-server application over LAN of our project is presented in Figure 3.

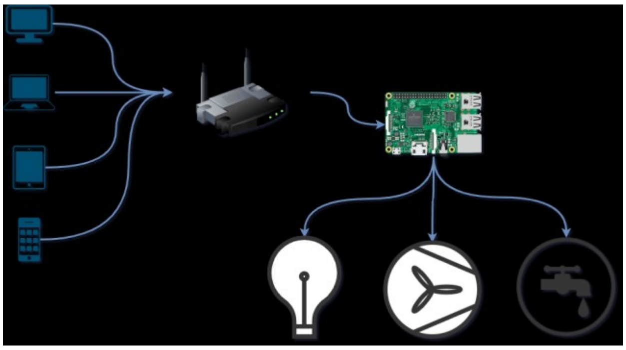


Figure 3. Client - Server application over LAN

In this project we will be using Node.js to turn our Raspberry Pi into a simple but powerful Web Server that will be accessible to any of the devices connected to the same network as the Raspberry Pi. The intent is to implement a simple RESTful API that will allow a client device to interact with the Raspberry Pi's inputs and outputs. To power up our Node server we will also be using the Express framework for NodeJS, which will add a thin layer of fundamental web application features and provide us with all the needed tools to create a robust API. An example flow that we are trying to achieve would look like this:

- Raspberry Pi is connected to the network with an IP address such as 192.168.1.131
- NodeJS is installed on the Raspberry and listening on port 3000
- A request is received on port 3000: POST http://192.168.1.131:3000/set/lights/on
- NodeJS will request the execution of a Python script that turns on the input connected to one of the relays where the lights power source is connected
- Lights are turned on.

For our client application we require a cross platform solution that will be able to interact with our RaspberryPi web server and perform simple HTTP requests. For this reason we have chosen to work with React Native. React Native is a framework for cross-platform mobile app development for iOS and Android, developed by Facebook and first released in 2015. To help us quickly get started and have a cross platform compatible mobile application we will specifically be using the Expo framework. The mobile application use cases diagram are presented in Figure 4. As we can see in the diagram there are 6 main activities that the user can perform while interacting with our application.

- *Configure WiFi activity:* This activity can only happen when the Smart Terrarium is first configured, or when the WiFi network has changed. In this activity the user is required to connect to the WiFi access point created by the Smart Terrarium and then provide it with a SSID and passkey, so that the Smart Terrarium can in turn connect to the local area network.

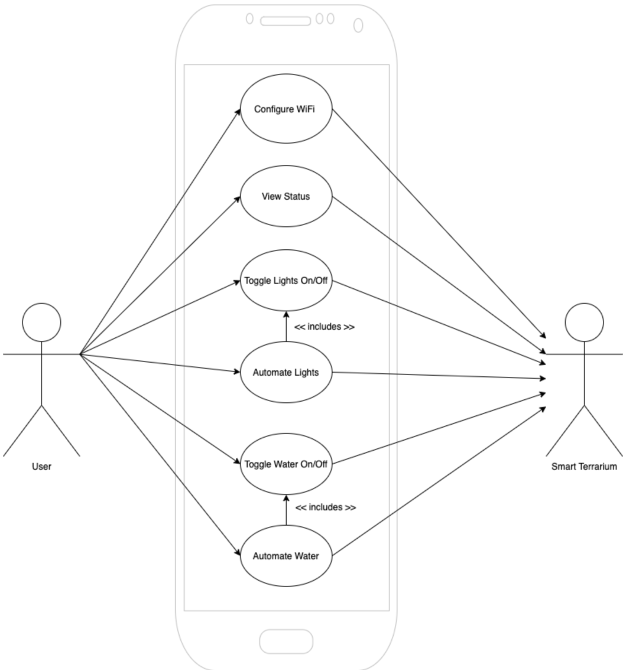


Figure 4. Application Use Cases Diagram

- *Toggle Lights/Water On/Off:* The toggle activity is the most basic activity that the application provides and it offers the user the ability to turn either the lights or the water on or off by simply touching a button from within their mobile application. In other words the mobile device becomes a power switch for the terrarium lights or water.
- *Schedule Lights/Water:* The Schedule activity provides full automation of the lights & water systems of the terrarium. This setting will enable the setting of (daily) time ranges in which the lights/water of the terrarium will turn on, and of course turn themselves back off when the time range has finished.
- *View temperature & humidity:* The user wants to see the temperature & humidity data of the Smart Terrarium so the use only needs to open the application. The latest data is shown and automatically refreshed every 60s with new data. Additionally the user can force an update of data by pressing the "sync now" button.

A very important feature of the application is that of the automation process in the entire project. We could, for example, want to schedule a daily watering task, or tell our Smart Terrarium how much light a day the plants should receive. A client could make a request to turn the lights on every day starting at 8:00PM for a duration of 12 hours. Or we could have a daily watering task that tells our terrarium to start the watering pump each day at 2:00PM for a total of 10 minutes.

In order to do that we will need a data store to be able to memorize our scheduled tasks and ensure they are not lost on a power outage, and additionally we will need a background process that will continuously check any scheduled tasks against the current time and run them when appropriate.

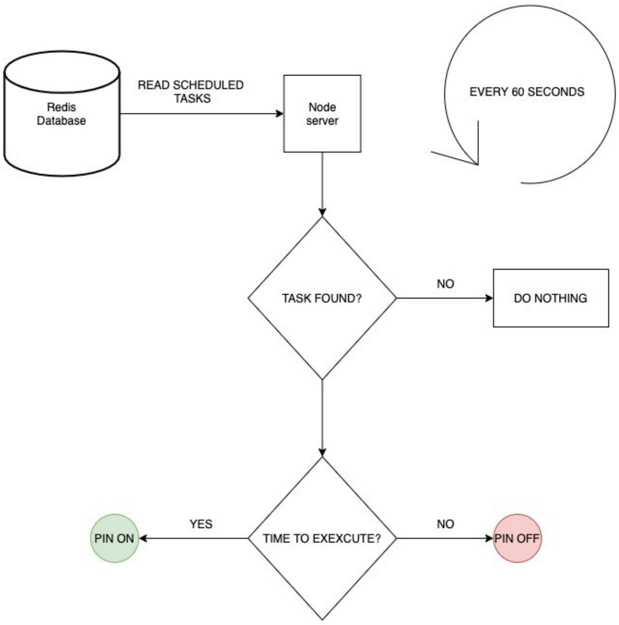


Figure 5. Executing Scheduled Tasks Diagram

In order to execute the scheduled automated tasks, we will be using the javascript *setInterval()* method which is also supported by NodeJS. This function will accept a delay and a callback as parameters and will execute the callback every delay milliseconds. Thus we will create a function that will check the current time, compare it to any scheduled tasks time stored in Redis and execute the task at the appropriate time.

IV. OBTAINED RESULTS

One of the main directions of this solution is to obtain a low-cost Smart Terrarium. This has been obtained by the price of the used components, which is approximately a total of 100$. The mobile application developed for the project is presented in Figure 6.

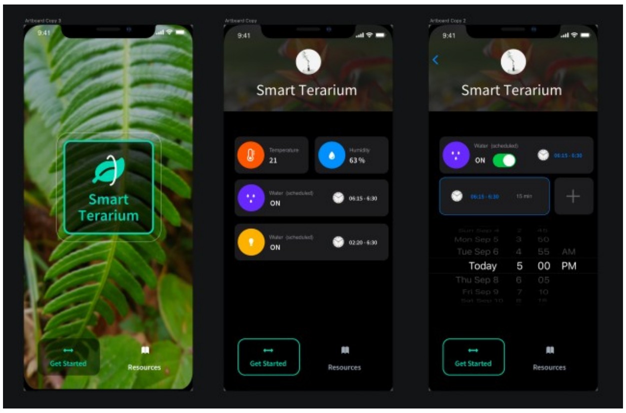


Figure 6. Mobile application interface

The test cases which were developed for the system, shows that the project has achieved the following statistics, which indicates a very high coverage:

- 87,5% code coverage for all statements

- 92.31% code coverage for all branches
- 100% code coverage for all functions
- 100% code coverage for all lines.

The goal of the unit test developed is to cover as much of the code as possible - this way we can be certain when deploying new features that all our present features are still functioning correctly. We are aiming for an above 80% coverage of all our code.

V. CONCLUSION AND FUTURE WORK

The present article underlines a Smart Terrarium solution in the context of Home Automation applications. The Smart Terrarium project is an ongoing project, being highly scalable as well as flexible, and can be turned into a smart grow box or expanded to a larger scale growing environment.

The project is scalable and allows for a large number of expansions and optimizations from a hardware and software point of view. A number of additional sensors and components could be added to the terrarium in order to increase its versatility such as:

- A sensor for measuring soil humidity
- A sensor to measure soil acidity
- A sensor to measure the quality of light received (lumen)
- A sensor to measure the amount of UV light received
- A video camera with which we could stream the contents of our terrarium and have constant supervision.

Additionally as far as the LED lighting goes a new type of LED has recently been developed that could have a very positive impact on the project: the driverless COB LED. COB, which stands for chip on board, is a circuit which contains multiple diodes (9 or more) with only 1 circuit and 2 contacts. COB LEDs are far more energy efficient and provide a far better lumen per watt ratio, and additionally the driverless versions allow us to connect the LED directly to 220AC power, without the need of an intermediate transformer.

REFERENCES

[1] K. Agarwal, A. Agarwal and G. Misra, "Review and Performance Analysis on Wireless Smart Home and Home Automation using IoT," Proc. of The IEEE Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2019.

[2] S. Mehrotra, R. Dhande, "Smart cities and smart homes: From realization to reality," Proc. of The International Conference on Green Computing and Internet of Things (ICGCIoT), 2015.

[3] Fortune Business Insights, Report ID: FBI100074, Home Automation Market Size Predictions and Regional Forecast 2018-2025, [online], https://www.fortunebusinessinsights.com/industry-reports/home-automation-market-100074, accessed on 28/06/2020.

[4] S. Matlak, R. Bogdan, Reducing Energy Consumption in Home Automation based on STM32F407 Microcontroller, 4th Telecommunications Forum TELFOR, Belgrade, Serbia, 2016.

[5] Poonphon Suesaowaluk, Home Automation System Based Mobile Application, Proc. of The IEEE 2nd World Symposium on Artificial Intelligence (WSAI), Guangzhou, China, 2020.

[6] M. Muttillo, V. Muttillo, L. Pomante, L. Pantoli, A Low Cost and Flexible Power Line Communication Sensory System for Home Automation, Proc. of IEEE International Workshop on Metrology for Industry 4.0 & IoT, Rome, Italy, 2020.

[7] D.-I. Gota, A. Puscasiu, A. Fanca, L. Miclea, H. Valean, Smart home automation system using Arduino microcontrollers, Proc. of The IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), Cluj-Napoca, Romania, 2020.

[8] Anand Vignesh, V. Vel, D. Purusothaman, V. Vijayaraghavan, "Low-Cost Non-Intrusive Residential Energy Monitoring System", Proc. of The IEEE Conference on Technologies for Sustainability (SusTech), 2014.

[9] K. Baraka, M. Ghobril, S. Malek, R. Kanj, A. Kayssi, "Low cost Arduino/Android-based Energy-Efficient Home Automation System with Smart Task Scheduling", Proc. of The Fifth International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), 2013.

Performance Analysis of Video Call Mobile Applications

Cs. Szabó* and V. Ramesh**
* Technical University of Košice/Department of Computers and Informatics, Košice, Slovakia
** Technical University of Košice/Faculty of Electrical Engineering and Informatics, Košice, Slovakia
csaba.szabo@tuke.sk, varun.ramesh@student.tuke.sk

Abstract - This research paper is the result of investigations of the authors under a cooperation during a final thesis preparation period. One can find an analysis of the two most popular video calling applications, WhatsApp and Messenger, herein on an Android device, mainly focusing on their power consumption, CPU utilization and memory usage of the device during a video call based on real life scenarios considering the distance, duration and other observable attributes of the video call according to the relation of the person spoken with and the number of applications running in the background during a video call. Or, without any background applications running. The analysis found that WhatsApp was better performing at handling CPU utilization and battery consumption, but it used relatively more system memory compared to that of Messenger.

I. Introduction

The reason why we chose to analyze video calling applications is because video calling has become the most prevalent form of communication. This can be seen in these trying times of mass pandemic. The scenarios were chosen according to how we normally communicate with our friends or family, which we assume to be the same with everyone.

Mobile devices get the energy required for their operation from batteries, in the case of many smartphones nowadays battery size is severely restricted due to weight and size restrictions. This implies that the energy efficiency of these devices should be taken into consideration. A core requirement of efficient management of energy is a good understanding of where and how the energy is used, how much of the system's energy is consumed by which parts of the system and under what circumstances like on an ongoing video call vs. on normal usage. The processor is the next main thing needed in a smartphone to process and carry out all the functions used in a smartphone, like listening to music, watching videos, web browsing, communicating, taking pictures and videos, etc. The focus on this research is to compare battery, memory usage and CPU utilization of the device under test during a video call on selected Android applications that is WhatsApp and Facebook's Messenger. Our approach is to measure the average power consumption, average CPU utilization and average system memory usage of a modern mobile device, the Samsung S10 Plus mobile phone, under a wide range of realistic usage scenarios. The scenarios in this analysis are based on how people communicate on a video call with friends, family, and for business related reasons. The scenarios were chosen considering how long people communicate on a video call, how far apart the persons on the video call are and the time at which the video call was made. For example, a person calling his parents tends to speak less on a video call than to his friends while a business video call lasts according to the work requirement.

This work has been supported by the KEGA grant No. 053TUKE-4/2019: "Learning Software Engineering via Continues Challenges and Competitions".

II. Related work

There are many video calling applications for Android that provide an interface for more than just video calling. Apps like HouseParty allow you to play games and share your screen with your friends on a video call, with Messenger also implementing this feature recently shows the popularity of these applications. There are also video calling applications which many people use only for gaming or community building and discussions on applications like Discord which is really popular for gaming where users can create their own rooms called servers and add their friends for communication on the background during gameplay and it also lets you stream while you are playing, it also lets you create bots using the discord.js package where users can code the bots to control the users in a room based on the rules of it automatically and paid users also get better quality video streaming and many other features. Another video calling application that is popular nowadays is Google Duo which has a feature where you can see the caller before you pick up if they have activated the feature and also it is great for video conferencing as it allows 12 participants at once for an unlimited time, while another application Zoom allows 100 participants at once but for a limited period of 40 Minutes for free users.

People have characterized these applications for various uses based on their functionality. E.g. WhatsApp, Messenger, Google Duo etc. are mainly used for daily communication, group studies or for business related work, while apps like HouseParty and Discord are used for entertainment and fun, and apps like Zoom are mainly used for business conferences and large meetings. The applications used for this analysis are WhatsApp and Messenger, although there are many other video calling applications the reason, we chose WhatsApp and

Messenger was because they are our personal favorites, also being common among all friends and family. They have been downloaded over a billion times just on play store with WhatsApp having more than 5 billion downloads and Messenger having over 1 billion downloads of course other video calling applications like Google Duo, Skype and IMO have over a billion downloads, but the most used applications are WhatsApp and Messenger as they are not only good for video calling but for audio calling, text messaging, sharing files, pictures and videos too.

The Performance analysis tool used in this analysis is the Qualcomm Snapdragon profiler. It can be downloaded for free by anyone by creating an account on Qualcomm's developer website. Snapdragon profiler has many features for analyzing many data metrics of the device in real-time. It also has trace capture mode which can be used to visualize kernel and system events to analyze low-level system events, Snapdragon profiler's interface itself is very simple and easy to use as Qualcomm provides a very good and instructive user guide [15] which is why we chose this tool for performance analysis of WhatsApp and Messenger. There are also other profilers like Android profiler which is integrated into the Android studio and it has similar features to that of the Snapdragon profiler. Android profiler also has features like Identifying CPU hotspots, profiling the apps layout with the hierarchy [19] but the documentation and user guides of this profiler were not detailed and hence was hard to understand. There are also mobile applications for profiling applications like the Trepn Profiler which can be downloaded for free from the play store, it can analyze many data metrics of the device in real-time and overlay the data on any app or the home screen, it also has different modes like Preset mode which has preset profiles for easier use and Advanced mode which can be used to customize overlays and also select data points[20]. But the overlay option clutters your mobile screen and interrupts your video call which is quite disturbing during a video call and also on mobile devices if you have to start or pause profiling you have to switch applications during a scenario which might impact the data metrics.

In the paper [3], authors state that communication has changed a lot for the new generations and are proposing an all in one application with features like in app games, shopping, listening to music and many more. Present IM applications have advantages over SMS like providing a confirmation if the recipient has read the message and apps like snapchat have a feature that allows users to send temporary messages which will disappear after a certain time if it is not saved by either the sender or the receiver, Telegram another application also provides high security by encryption and self-destructing messages. Along with these features, apps can be designed to be able to handle and solve the major issues of the Health in the country. The author thinks that WhatsApp family apps are widely successful but there could be more improvements in these apps, and they provide the list of features along with the explanation that the new proposed app will have.

The paper [4] attempts to identify the different learning preferences by the students by analyzing the WhatsApp conversations among them and also tries to answer whether students prefer to take part in a conversation group for learning, and how many students would stay in the group just to read conversations by the people in the group. This paper shows the advantages of Instant messengers like WhatsApp and Messenger that despite the factors such as geographical differences hinder group work, students are able to interact with their classmates and lecturers. It also shows the disadvantages of online discussions where it lacks the flow of dialogue, facial expressions and hand gestures that is usual in a face-to-face interaction and some students preferred face-to-face format over online discussions. I. Smit concluded that IM application like WhatsApp can be used as a powerful tool to help students in their studies assessments.

In the paper [5], the group of authors investigates group-based communication and how it has changed the way people communicate. This paper analyses the implications and usage of group-based communications the authors conducted a survey based on group communications on the campus of University of Würzburg, Germany. It compared the usage of WhatsApp to SMS and the survey showed that WhatsApp is preferred over SMS and students used it at least once in a day and on average they had 10 group chats this shows that group chat feature is popular among every WhatsApp user. In this paper they also showed that how group-based communication can change the activity patterns of many users and the evolution of the communication towards group-based communication from one to one communication and how it has been progressing. WhatsApp has become a main application for group communications worldwide, this analysis helps to better cope with the network demands by designing the novel traffic management by modelling and simulating the communication in groups.

Paper [6] analyzes the importance of IM messengers like WhatsApp to improve business opportunities for street vendors by conducting in-depth interview with the street vendors to gain knowledge about how social media applications helps to grow their businesses. The analysis found that group-based communication and awareness through WhatsApp helped in the growth of their business and also was helpful in getting feedback from their customers which helped in boosting their confidence. They could reach a large number of people all at once just by spending a small budget on internet connections compared to the high cost of commercial advertisements. It also discusses the performance and popularity of WhatsApp throughout the world.

Authors of [8] try to implement a security system with the interface of a doorbell and WhatsApp application via a mobile phone connected through Bluetooth technology. Through this setup an authorized person can be notified of a visitor to his place on the web through the camera of the mobile phone from anywhere, which can greatly decrease the crimes done with old people at home, can keep track of visitors in the absence of family members and it is also cost effective. This paper shows the vital role of embedded system in Security systems.

Authors of [14] analyze the performance of the WhatsApp and skype and their comparison in terms of the data usage/consumption when sending the messages. For analyzing the setup an experimental test bed to measure the VoIP call quality of both Skype and WhatsApp and their data consumption performance. Through their findings through the experimental scenario they concluded that WhatsApp is a more suitable and reliable app than Skype in terms of making VoIP calls and sending messages.

The paper [13] presents a web-based tool named “WhatsAnalyzer” which can be used to collect and analyze the chat history of the mobile messaging application WhatsApp. To access the chat data, chat histories are extracted in WhatsApp and then sent by email to WhatsAnalyzer. While authors of [10] investigate the cybercriminals with the help of sniffing techniques and network forensic and purposes a packet filter framework to WhatsApp communication pattern to identify criminals. They propose a rule extraction method in sniffing packets to retrieve relevant attributes. This paper will be helpful in modern call record analysis and assist in LEAs to prosecute cybercriminals and bring them to justice.

Authors of [11] analyze the performance of the WhatsApp and presents a case study of quantitative data with disruptive statistics regarding the perception of WhatsApp usage of students in two design-based modules. According to the surveys conducted, the author concluded that students' perception of WhatsApp's advantages is collaboration between students and lecturers, better communication, enhanced motivation, and better academic performance, and the role of WhatsApp usage towards the design-based modules. The Authors concluded that WhatsApp is an extremely powerful tool in education that helps in better collaboration.

In the paper [1] they discuss the importance of a secure private instant messenger. The authors of this paper discuss how secure data transmission is extremely important and how it can only be ensured through a secure and private Instant messenger. Most of the instant messengers do not encrypt network traffic, which increases the risk of infected files, eavesdropping, identity theft, account hijacking and the loss of confidential matter which users is not dangerous, but in this age where the hackers are getting extremely sophisticated, it is extremely risky for safe and smooth performance of Instant messengers. This paper also discusses the architecture of a secure instant messenger. The National Institute of Standards and Technology has created several cryptographic hash functions and one of them is a secure hash algorithm (SHA). SHA is used for text encryption which ensures a secure information transmission between sender and recipient. This paper suggests that secure IM is essential for better communication and through the hashing algorithm we can make it more secure than any other algorithm it also shows the example of implementing a hash function and through its results it is proved that conversations can be secured through hash algorithms and its applications in LAN conversations.

Paper [2] is the analysis of the traffic of two instant messaging systems AOL Instant Messenger and MSN/Windows Live Messenger. They found out that chat messages constitute only a small percentage of IM traffic and most of the traffic is extraneous. Their analysis sheds light on IM system designs and optimization which provides a scientific basis for instant messaging workload generation. For example, out of the twenty IM users interviewed they found out that IMs are usually used for negotiation in co-workers availability who may then switch to a different media for complex discussions and from the study of IM usage from sixteen teenagers they found different social behaviors between high school and college students. According to this paper’s authors understanding the instant messaging traffic characteristics is very important to understand the workload generation and system design.

The authors of [7] conducted a digital forensic investigation and provided the results to help the investigator or analysts to read and reconstruct the chronology of messages that are created by the users. For this investigation they choose the six different scenarios and based on them presented the digital forensic investigation of private and normal chats with Line, KakaoTalk, and Telegram social messenger application on the android devices. They presented the essential correlation between the database files, captured packets and files, and logs, and all of them can be used for digital evidence. Through their list of guidelines, analysts will be able to detect and catch cybercriminals related to these three messengers.

D. R. Vukovic and I. M. Dujlovic mention in [9] the use of bots in Messenger, how they will greatly help in the growth of business. According to this paper, there are only about 2% of bots in the finance sector. Bots can be used to assist the user while shopping, reserving a flight ticket by showing items related to the brought product. The author implemented a bot through the free development automated tool for the bank and named it pseudo bank bot to present the advantages and disadvantages of using Facebook bots for business. The author stated that in the future Facebook bots can be of extreme importance in the finance sector and helps in connecting with users easily which can reduce the cost of customer support.

III. CASE STUDY

Our approach to analyzing video calling on two different applications on an Android device is to take power consumption, CPU utilization, system memory usage parameters and to carry out five different kinds of scenarios on each application with each scenario being very similar on both the applications, so as to compare the results between them. The scenarios were conducted on a number of runs to calculate the average data which is put into the final results table and the type of scenarios were chosen after collecting some data from our friends like, how many background applications are usually open on their device during a video call, the duration of the video call depending on the relation with the person on the video call and the distance of the video call.

A. Scenarios

Scenario 1 on both the applications is a very short distance video call of about 1.5 km with no other

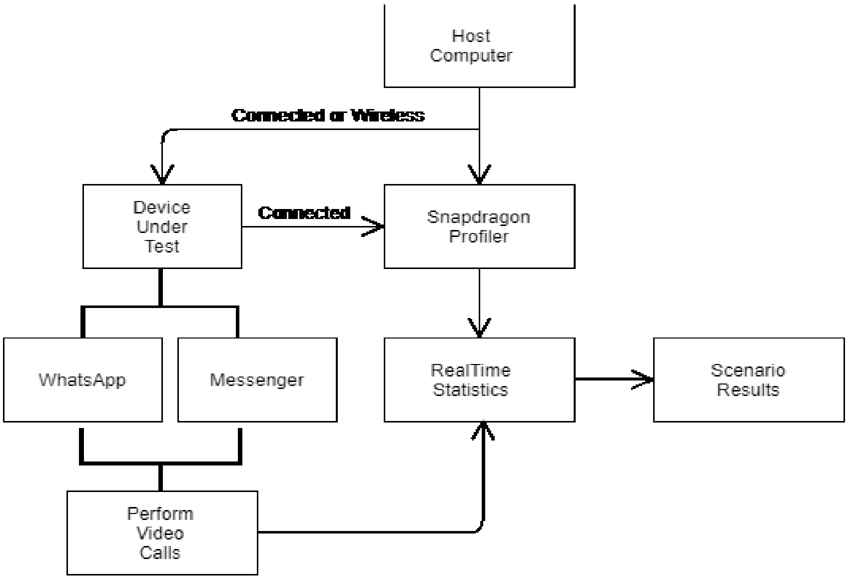


FIGURE 1. THE GENERIC MODEL OF THE MEASUREMENT METHOD

background applications running conducted for 4 min and 52 sec.

Scenario 2 is a long-distance video call of about 6,481 km to Bangalore, India with no other background applications running conducted for 4 min and 52 sec. This video call was conducted for the same length of scenario 1 to check the impact of distance on the data utilization of the device.

Scenario 3 video call was conducted for 6 min and 28 sec with 4 other background applications (YouTube, Call of Duty, PUBG/PlayerUnknown's Battlegrounds and Google Maps) running in the background.

Scenario 4 video call was conducted for 15 min and 30 sec, more than double the length of scenario 3, to check the impact of time on the data utilization of the device with 4 background applications similar to scenario 3 running in the background.

Scenario 5 is a group video call conducted for 25 min and 10 sec with two other people in the group with 3 other background applications (Facebook, Instagram and Snapchat) running in the background.

B. Method

To analyze the data on the device during video calls on the chosen applications on an Android device the device had to be connected to a profiler, in this case we used the Qualcomm profiler [15] because it was very simple and easy to connect. After connecting the device to the profiler and launching the applications on the profiler the initial data on the chosen parameters i.e. battery percentage, memory consumption and CPU utilization were recorded on launching the applications before starting the video call (see Fig. 1). Then, the parameters were recorded again during the video call to check how the data changed when the application was idle and during a video call.

For our analysis, we had to set up the measurement details as presented in Fig. 2. There are three elements to the setup: the device under test, a profiler, and a host computer. After enabling USB debugging under developer settings of the device and downloading ADB tools and Snapdragon profiler, the device can be either connected to the profiler through the host computer via the USB cable or wirelessly

C. Devices

The physical device in our study was a Samsung Galaxy S10+ smartphone, its details follow.

- CPU: Octa-core (2x2.73 GHz Mongoose M4 & 2x2.31 GHz Cortex-A75 & 4x1.95 GHz Cortex-A55) - EMEA/LATAM
- RAM: 8GB
- OS: Android 9.0 (pie) One UI
- GPU: Mali-G76 MP12 - EMEA/LATAM

D. Data collection

The scenarios were carried out on both applications using Qualcomm Snapdragon profiler to get the maximum CPU utilization and maximum memory usage by the application in different scenarios. For all the scenarios the

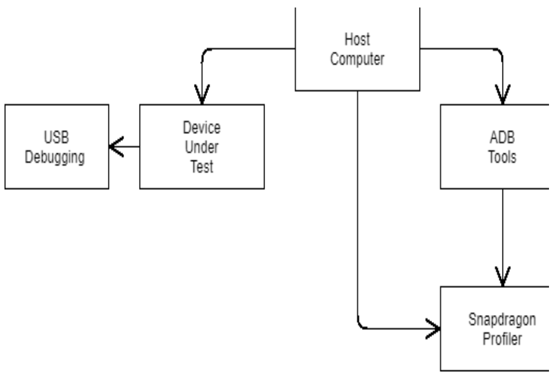


FIGURE 2. DETAILED EXPERIMENTAL SETUP

device was wirelessly connected to the Snapdragon profiler through the host computer. The battery consumed was monitored on the battery usage by apps section of the device in settings and also by manually checking the battery percentage. Four runs were conducted similarly on every scenario and its data was put into the table. Every scenario got calculated the average data and the average data was put into the final results table used for comparison.

IV. RESULTS AND DISCUSSION

For Scenario 1 on Messenger application a very short distance video call of about 1.5 Km was made for 4 minutes and 52 seconds with no other background applications running. When the application was first launched, the CPU utilization was 5.13% and the memory usage by the application was 418.04 MB. During the video call, the CPU utilization increased to a maximum of 46.15% while the memory usage increased to a maximum of 3.07 GB. The battery usage by the application was 2.4%. Table 1 shows Scenario 1 conducted similarly on a number of runs to calculate the average data. On run 3 a check was done to see if there were any changes in the data by turning the camera off and muting audio at separate intervals and there were no noticeable changes due to it. In the case of WhatsApp, when the application was launched the CPU utilization was 4.5% and the memory usage by the application was 3.0 GB. During the video call the CPU utilization increased to a maximum of 26.32% while the memory usage increased to a maximum of 3.87 GB and the battery usage by the application was 1.6%. Table 2 shows Scenario 1 conducted on WhatsApp similarly on a number of runs to calculate the average data. On run 3 a check was done to see if there were any changes in the data by turning the camera off and muting audio at separate intervals. Muting and unmuting audio did not have any noticeable changes in the data but when the camera was turned off the memory usage decreased from 4.07 GB to 4.00 GB while there was no noticeable change in the CPU utilization. When the camera was turned back on the memory usage increased from 4.00 GB to 4.05 GB. It is a significant difference to the results of Messenger measurements in this scenario.

For scenario 2 on Messenger application a very long-distance video call of about 6,481 Km to Bangalore, India was made for 4 minutes and 52 seconds with no other background applications running. When the application was launched the CPU utilization was 2.43% and the memory usage by the application was 15.23 MB, during the video call the CPU utilization increased to a maximum of 45.00% while the memory usage increased to a maximum of 4.25 GB and the battery usage by the application was 3.6%. Table 3 shows Scenario 2 conducted similarly on a number of runs to calculate the average data. When the WhatsApp application was launched the CPU utilization was 2.5% and the memory usage by the application was 3.56 GB, during the video call the CPU utilization increased to a maximum of 28.20% while the memory usage increased to a maximum of 3.75 GB and the battery usage by the application was 3.4%. When the video call is answered on WhatsApp application, the CPU utilization sharply increased from 7.32% to 30.23% while the memory usage increased form 3.83 GB to 4.11 GB. Table 4 shows Scenario 2 conducted similarly on a number of runs to calculate the average data.

TABLE 1. SCENARIO 1 ON MESSENGER

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	3.07 GB	2.4%	46.15 %
Run 2	3.22 GB	2.5 %	51.28 %
Run 3	4.01 GB	2.1 %	48.28 %
Run 4	3.87 GB	2.6 %	43.18 %

TABLE 2. SCENARIO 1 ON WHATSAPP

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	3.87 GB	1.3 %	26.32 %
Run 2	3.98 GB	1.5 %	25 %
Run 3	4.13 GB	1.7 %	25 %
Run 4	4.22 GB	1.4 %	23.5 %

TABLE 3. SCENARIO 2 ON MESSENGER

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	4.25 GB	2.3 %	45 %
Run 2	4.26 GB	2.3 %	44.19 %
Run 3	4.29 GB	2.4 %	42.77 %
Run 4	3.97 GB	2.2 %	47.89 %

TABLE 4. SCENARIO 2 ON WHATSAPP

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	4.0 GB	1.2 %	30 %
Run 2	3.91 GB	1.4 %	37.5 %
Run 3	4.09 GB	1.3 %	26.5 %
Run 4	3.88 GB	1.6 %	25.4 %

TABLE 5. SCENARIO 3 ON MESSENGER

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	4.28 GB	4 %	56 %
Run 2	4.29 GB	3.7 %	51.34 %
Run 3	4.29 GB	3.8 %	48.72 %
Run 4	4.22 GB	3.5 %	53.28 %

TABLE 6. SCENARIO 3 ON WHATSAPP

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	3.96 GB	2.7 %	30 %
Run 2	4.18 GB	2.4 %	31.7 %
Run 3	4.01 GB	2.2 %	33.46 %
Run 4	2.97 GB	-	-
Run 5	3.92 GB	2.5 %	35.14 %

For scenario 3 on Messenger application a video call for 6 minutes and 28 seconds was made with 4 other background applications (YouTube, Call of Duty, PUBG and Google Maps) running. When the application was launched the CPU utilization was 2.05% and the memory usage by the application was 7.72 MB, during the video call the CPU utilization increased to a maximum of 56% while the memory usage increased to a maximum of 4.28 GB and the battery usage by the application was 4%. Table 5 shows Scenario 3 conducted similarly on a number of runs to calculate the average data. On run 2 a check was done to see if there was any major difference in the data due to the contrast in the camera by switching the lights on and off on separate intervals and there were no noticeable changes due to it. When the WhatsApp application was launched the CPU utilization was 4.89% and the memory usage by the application was 2.94 GB, during the video call the CPU utilization increased to a maximum of 30% while the memory usage increased to a maximum of 3.96 GB and the battery usage by the application was 2.7%. Table 6 shows the results from scenario 3 on WhatsApp application conducted similarly on a number of runs to calculate the average data. On Run 1 there was a loss of network which interrupted the normal CPU utilization, after reconnecting automatically there was a sharp increase in CPU utilization from 13% to 29%. On run 2 a check was done to see if there was any major difference in the data due to the contrast in the camera by switching the lights on and off on separate intervals and there were no noticeable changes due to it. Run 4 was interrupted by a normal call so the data was not considered for calculating average.

For scenario 4 on Messenger application a video call was made for 15 minutes and 30 seconds with 4 other background applications (YouTube, Call of Duty, PUBG and Google Maps) running. When the application was launched the CPU utilization was 5.12% and the memory usage by the application was 54.64 MB, during the video call the CPU utilization increased to a maximum of 41.67% while the memory usage increased to a maximum of 4.28 GB and the battery usage by the application was 4.2%. Table 7 shows Scenario 4 conducted similarly on a number of runs to calculate the average data. When the WhatsApp application was launched the CPU utilization was 4.88% and the memory usage by the application was 3.14 GB, during the video call the CPU utilization increased to a maximum of 34.15% while the memory usage increased to a maximum of 3.87 GB and the battery usage by the application was 4.3%. Table 8 shows the results from scenario 4 on WhatsApp application conducted similarly on a number of runs to calculate the average data.

TABLE 7. SCENARIO 4 ON MESSENGER

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	4.28 GB	5.7 %	41.67 %
Run 2	4.27 GB	5.2 %	59.52 %
Run 3	4.22 GB	4.9 %	54.36 %
Run 4	4.18 GB	5.5 %	53.97 %

TABLE 8. SCENARIO 4 ON WHATSAPP

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	3.87 GB	4.3 %	34.15 %
Run 2	3.92 GB	4.6 %	31.7 %
Run 3	3.99 GB	4.2 %	31.7 %
Run 4	3.96 GB	4.4 %	36.4 %

TABLE 9. SCENARIO 5 ON MESSENGER

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	4.29 GB	7.2 %	50 %
Run 2	4.28 GB	6.9 %	55.25 %
Run 3	4.20 GB	7.4 %	53.45 %
Run 4	4.27 GB	7.1 %	51.28 %

TABLE 10. SCENARIO 5 ON WHATSAPP

Runs	Maximum Memory Usage	Battery Usage	Maximum CPU Utilization
Run 1	4.25 GB	6.7 %	30.77 %
Run 2	4.29 GB	5.9 %	36.10 %
Run 3	4.17 GB	6.2 %	33.28 %
Run 4	4.22 GB	6.4 %	35.27 %

TABLE 11. SCENARIO AVERAGE VALUES ON MESSENGER

Scenario	Average Memory Usage	Average Battery Used	Average CPU Utilization
Scenario 1	3.54 GB	2.4 %	47.22 %
Scenario 2	4.19 GB	2.3 %	44.96 %
Scenario 3	4.27 GB	3.75 %	52.34 %
Scenario 4	4.24 GB	5.33 %	52.4 %
Scenario 5	4.26 GB	7.15 %	52.5 %

TABLE 12. SCENARIO AVERAGE VALUES ON WHATSAPP

Scenario	Average Memory Usage	Average Battery Used	Average CPU Utilization
Scenario 1	4.05 GB	1.48 %	24.96 %
Scenario 2	3.97 GB	1.38 %	29.85 %
Scenario 3	4.01 GB	2.45 %	32.58 %
Scenario 4	3.94 GB	4.38 %	33.49 %
Scenario 5	4.23 GB	6.3 %	33.86 %

For scenario 5 on Messenger application a group video call with two other people was made for 25 minutes and 10 seconds with 3 other less intensive background applications compared to Scenario 3 and 4 running (Facebook, Instagram, Snapchat). When the application was launched the CPU utilization was 2.56% and the memory usage by the application was 31.88 MB during the video call the CPU utilization increased to a maximum of 50% while the memory usage increased to a maximum of 4.29 GB and the battery usage by the application was 7.2%. Table 9 shows Scenario 5 conducted similarly on a number of runs to calculate the average data. When the WhatsApp application was launched the CPU utilization was 2.4% and the memory usage by the application was 2.25 GB, during the video call the CPU utilization increased to a maximum of 30.77% while the memory usage increased to a maximum of 4.25 GB and the battery usage by the application was 5.9%. Table 10 shows the results from scenario 5 on WhatsApp application conducted similarly on a number of runs to calculate the average data.

A. Discussion

The tables 11-12 show the average memory usage, average battery used and the average CPU Utilization by the applications WhatsApp and Messenger for each scenario conducted on four similar runs.

First, we discuss results of the Messenger application. Scenario 1 and Scenario 2 are based on distance, while scenario 1 is a short-distance video call and scenario 2 is a long-distance video call conducted for the same time for about 4 min and 52 sec to compare the results and there were no background applications open during this scenario. According to the above results short-distance video call had just 8% less average memory usage and 0.1% more average battery usage while the average CPU utilization was also 2.26% higher in short distance video call which shows that distance has no impact on the performance of the application. Scenario 3 and 4 are based on the length of the video call. Scenario 3 was conducted for 6 min and 28 sec while scenario 4 was conducted for 15 min and 30 sec and there were four background applications (PUBG, Call of Duty, YouTube, Google maps) running in the background. According to the above results the average memory usage was 0.38% lower in a lengthier video call while the average CPU utilization was 0.06% higher in a lengthier video call, both of which can be considered negligible. The average battery consumption was 1.58% higher in a lengthier video call which shows that the average memory usage and CPU utilization were not affected by the length of the video call except for the average battery consumption which is higher in a lengthier video call due to the time difference between scenario 3 and scenario 4. Scenario 5 is a group video call with two other people in the group and was conducted for 25 min and 10 second with three other less intensive background applications (Facebook, Instagram and Snapchat) running in the background compared to that of scenario 3 and 4. If we compare the average memory usage, CPU utilization and the battery consumption with that of scenario 4 it shows that scenario 5 had just 0.3% higher average memory usage and 0.1% higher average CPU utilization than scenario 4 but the average battery consumption was 1.82% higher due to the length of the video call. This scenario showed that even though it was a group video call there were no significant changes in the average memory usage and CPU utilization of the device except for the average battery consumption which is higher due to the length of the video call.

In the case of WhatsApp, Scenario 1 and Scenario 2 showed that short-distance video call had 1% more average memory usage and 0.1% more average battery usage while the average CPU utilization was 4.89% higher in long-distance video call which shows that distance had no impact on the performance of the application. During Scenario 3 and 4, average memory usage was 0.88% lower in a lengthier video call while the CPU utilization was 0.91% higher in a lengthier video call, both of which can be considered negligible but the average battery consumption was 1.93% higher in a lengthier video call, which shows that the average memory usage and CPU utilization were not affected by the length of the video call except for the average battery consumption which is higher in a lengthier video call due to the time difference between scenario 3 and scenario 4. If we compare the average memory usage, CPU utilization and battery consumption with that of scenario 4, it shows that scenario 5 had just 3.63% higher average memory usage and 0.37% higher average CPU utilization than scenario 4 but the average battery consumption was 1.92% higher due to the length of the video call. This scenario showed that even though it was a group video call there were no significant changes in the average memory usage and CPU utilization of the device except for the average battery consumption which is higher due to the length of the video call.

V. CONCLUSION

The average readings recorded from different scenarios conducted on 4 runs on WhatsApp and Messenger applications from the Snapdragon profiler are compared below.

In scenario 1 the average memory usage by Messenger is 6.38% lower than WhatsApp application while the average CPU utilization by WhatsApp is 22.26% lower and the average battery consumption by WhatsApp is 0.92% less than that of Messenger.

In scenario 2 the average memory usage by Messenger is 2.75% higher than WhatsApp application while the average CPU utilization by WhatsApp is 15.11% lower and the average battery consumption by WhatsApp is 0.92% less than that of Messenger.

In scenario 3 the average memory usage by Messenger is 3.75% higher than WhatsApp application while the average CPU utilization by WhatsApp is 19.76% lower and the average battery consumption by WhatsApp is 1.3% less than that of Messenger.

In scenario 4 the average memory usage by Messenger is 3.75% higher than WhatsApp application while the average CPU utilization by WhatsApp is 18.91% lower and the average battery consumption by WhatsApp is 0.95% less than that of Messenger.

In scenario 5 the average memory usage by Messenger is 0.37% higher than WhatsApp application while the average CPU utilization by WhatsApp is 18.64% lower and the average battery consumption by WhatsApp is 0.85% less than that of Messenger.

According to the above comparison between the applications from the average data calculated the average memory usage on both the applications varied on different scenarios with the difference being not so significant. Even though the average battery consumption is lower on WhatsApp application in all the scenarios it is also not significantly lower than that of Messenger application which shows that these instant messaging applications have a similar approach to memory usage and battery consumption of the device while the CPU utilization by WhatsApp application on all the scenarios was on an average 19% lower compared to that of Messenger application which is significantly lower and shows that the CPU utilization was better handled by WhatsApp. This analysis showed that the memory usage and the battery usage of the device from both the applications were more or less similar, but the CPU utilization was better handled by WhatsApp application. On scenario 3 run 4 of WhatsApp application during video call there was an incoming normal call which interrupted the scenario, so the run was not considered for calculating the average data.

We can list factors affecting the data metrics of the device during a video call as shown below:

On Messenger and WhatsApp application a sharp increase in CPU utilization and Memory usage was noticed when the video call was answered. The CPU utilization sharply increased from 12 % to 45 % and the memory usage increased from 3.07 GB to 4.25 GB. On WhatsApp application the CPU utilization sharply increased from 7.32% to 30.23% and the Memory usage increased from 3.83 GB to 4.11 GB

On WhatsApp application there was a change in memory usage when the camera was turned on and off during separate intervals. When the camera was turned off the Memory usage decreased from 4.07 GB to 4.00 GB and when the camera was turned back on, the Memory usage increased from 4.00 GB to 4.05 GB while there were no noticeable changes in the CPU utilization. Muting and unmuting the audio also did not have any noticeable changes. These factors did not show any noticeable changes on the data on Messenger application.

On WhatsApp application when there was a loss of network it interrupted the CPU utilization. After reconnecting to the network, the CPU utilization sharply increased from 13% to 29%, while there was no noticeable change in the Memory usage.

Factors not affecting the data metrics of the device during a video call are mentioned below.

Distance had no impact on the memory usage, CPU utilization or the battery consumption of the device on both the applications.

There was also no change in the memory usage, CPU utilization or the battery consumption on both the applications when the lights were turned on and off on separate intervals, to check if the contrast in the camera had any effect on the data.

ACKNOWLEDGMENT

This paper is part of dissemination of results related to the Erasmus+ Key Action 2 (Strategic partnership for higher education) project No. 2017-1-SK01-KA203-035402: "Focusing Education on Composability, Comprehensibility and Correctness of Working Software". The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

REFERENCES

[1] M. K. Yusof and A. F. A. Abidin, "A secure private instant messenger," The 17th Asia Pacific Conference on Communications, Sabah, 2011, pp. 821-825, doi: 10.1109/APCC.2011.6152921.

[2] Z. Xiao, L. Guo and J. Tracey, "Understanding Instant Messaging Traffic Characteristics," 27th International Conference on Distributed Computing Systems (ICDCS '07), Toronto, ON, 2007, pp. 51-51, doi: 10.1109/ICDCS.2007.149.

[3] H. Bajaj and R. Jindal, "Thinking beyond WhatsApp," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2015, pp. 1443-1447.

[4] I. Smit, "WhatsApp with learning preferences?" 2015 IEEE Frontiers in Education Conference (FIE), El Paso, TX, 2015, pp. 1-6, doi: 10.1109/FIE.2015.7344366.

[5] M. Seufert, T. Hoßfeld, A. Schwind, V. Burger and P. Tran-Gia, "Group-based communication in WhatsApp," 2016 IFIP Networking Conference (IFIP Networking) and Workshops, Vienna, 2016, pp. 536-541, doi: 10.1109/IFIPNetworking.2016.7497256.

[6] N. N. Amal, D. Kusumawati and F. Ansoriyah, "Join my group: Evidences of WhatsApp use to run a better breakfast street food business," 2016 22nd Asia-Pacific Conference on Communications (APCC), Yogyakarta, 2016, pp. 119-123, doi: 10.1109/APCC.2016.7581474.

[7] G. B. Satrya, P. T. Daely and S. Y. Shin, "Android forensics analysis: Private chat on social messenger," 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), Vienna, 2016, pp. 430-435, doi:10.1109/ICUFN.2016.7537064.

[8] A. Saini, A. Marwah, P. Bansal and N. Ruhil, "Photo of a visitor using WhatsApp," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 3251-3253.

[9] D. R. Vukovic and I. M. Dujlovic, "Facebook Messenger bots and their application for business," 2016 24th Telecommunications Forum (TELFOR), Belgrade, 2016, pp. 1-4, doi: 10.1109/TELFOR.2016.7818926.

[10] F. Tsai, E. Chang and D. Kao, "WhatsApp network forensics: Discovering the communication payloads behind cybercriminals," 2018 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon-si Gangwon-do, Korea (South), 2018, pp. 1-1, doi: 10.23919/ICACT.2018.8323881.

[11] P. E. Hertzog and A. J. Swart, "The use of WhatsApp in design-based modules," 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, 2018, pp. 445-450, doi: 10.1109/EDUCON.2018.8363264.

[12] Á. J. Sánchez, M. R. F. Sánchez, E. C. V. Ruiz, M. A. Sánchez, V. M. Dueñas and J. M. Lavín, "WhatsApp and Facebook as a New Political Agora?," 2018 International Conference on eDemocracy & eGovernment (ICEDEG), Ambato, 2018, pp. 366-370, doi: 10.1109/ICEDEG.2018.8372340.

[13] A. Schwind and M. Seufert, "WhatsAnalyzer: A Tool for Collecting and Analyzing WhatsApp Mobile Messaging Communication Data," 2018 30th International Teletraffic Congress (ITC 30), Vienna, 2018, pp. 85-88, doi: 10.1109/ITC30.2018.00020.

[14] N. Patel, S. Patel and W. L. Tan, "Performance Comparison of WhatsApp versus Skype on Smart Phones," 2018 28th International Telecommunication Networks and Applications Conference (ITNAC), Sydney, NSW, 2018, pp. 1-3, doi: 10.1109/ATNAC.2018.8615445.

[15] "Qualcomm profiler user guide" https://developer.qualcomm.com/download/sdprofiler/sdprofiler-user-guide.pdf

[16] "Samsung S10 plus specifications"
https://www.gsmarena.com/samsung_galaxy_s10+-9535.php

[17] "WhatsApp Encryption Overview"
https://scontent.whatsapp.net/v/t61.2286834/68135620_760356657751682_6212997528851833559_n.pdf/WhatsAppSecurityWhitepaper.pdf?_nc_sid=41cc27&_nc_ohc=9tBPAXBRomwAX9g4Zrc&_nc_ht=scontent.whatsapp.net&oh=990db51c2b0ead3cefc6de82ba8bdf46&oe=5ED02853

[18] "Messenger secret conversations"
https://fbnewsroomus.files.wordpress.com/2016/07/messenger-secret-conversations-technical-whitepaper.pdf

[19] "Android Studio documentation"
https://developer.android.com/studio/profile

[20] "Trepn profiler user guide"
https://developer.qualcomm.com/download/trepn/trepn-user-guide.pdf

Distributed Retail Systems

A. Asani*, T. Dimovski* and I. Hristoski**
* "St. Kliment Ohridski" University – Bitola / Faculty of ICTs – Bitola, N. Macedonia
** "St. Kliment Ohridski" University – Bitola / Faculty of Economics – Prilep, N. Macedonia
arijan.asani.bt@gmail.com, {tome.dimovski, ilija.hristoski}@uklo.edu.mk

Abstract - In the era of modern computing, along with the huge advancements achieved in the field of computer science, most of today's processes are digitalized. There is an increasing demand for reliability, availability, stability, and scalability. When it comes to retail business, retail shop software plays an important role, which points out the need for designing more sophisticated systems. One of the most popular approaches in addressing ever-increasing system requirements is the development of distributed systems. By definition, a distributed system represents a group of computers that process data/tasks together and appears as a single entity to the end-user of the system. In this paper, we are presenting the design and the development of a distributed system that is intended to be used in the retail sector and elaborate on the functionality of each of the individual entities comprising the model. The described distributed retail system represents a solid foundation for further improvements and upgrading.

I. Introduction

We live in a world where technology is evolving at an accelerated pace and is becoming an inevitable part of our daily lives. Information and communication technologies are increasingly present in almost all spheres of our society. The information systems, and especially distributed ones, are widespread today, playing a key role in the economies and societies across the globe. In the retail sector, the use of modern software solutions is necessary, as there is an increasing need to meet progressively complex customer requirements. Faced with the new challenges, there is an emphasized need for the design, development, and deployment of distributed retail systems.

Distributed systems are systems that are composed of multiple entities, usually spread out across different physical locations, which are mutually synchronized with each other and present to the end-user a picture of being a single entity. The implementation of such types of systems addresses the majority of the customer's needs, such as the need for increased system availability and dependability, better scalability and stability, greater performance improvements, as well as easier data integration and data analysis.

Having minded the previous, the main goal is to present the concepts behind a newly designed and implemented distributed retail system, which has great potentials to grow into a fully functional software solution intended for use by micro- and small-sized retailers.

The rest of the paper is structured as follows. Section II focuses on some of the most relevant research made on this topic recently. In Section III, a brief overview of several well-known retail systems is made. The major capabilities of contemporary retail systems are given in Section IV. Section V discusses the proposed model of a distributed retail system and elaborates on the functionality of each of the individual entities of the model. The last section highlights the conclusions and future work.

II. Related Research

Most of the recent research on this topic has been focused on the exploration of newer technologies that can be leveraged to improve retail business efficiency and performance, as well as to maximize the return on investment (ROI). In this context, some authors elaborate theoretically on the proposed frameworks and consequently describe the implementation of such systems. For instance, Adewumi *et al*. describe the design and the development of *SkyRetail*, a cloud-based retail management system [1]. In his white paper, Dion explores the impact of the deployment of technology in a retail business [2]. More specifically, he looks at the effects on sales, profitability, and productivity of the use of Point-of-Sale, Inventory Control, and Customer Profiling Software in small- to mid-sized retail stores. According to his findings, "the deployment of technology in a retail store leads to higher sales, reduced expenses, and increased gross margins, which have ultimately produced an increase in the overall profitability of those stores who not only have deployed the technology but have also learned how to use it." In his thesis, Cote elaborates on a project that highlights a systematic method for searching, identifying, evaluating, selecting, and recommending a Point-of-Sale (POS) system and Inventory Management (IM) system for a small business, based upon its specific industry needs [3]. According to his findings, information obtained from POS systems improves marketing by helping salespersons make better judgments and ultimately practice smarter selling. Data reports highlight specific needs and eliminate guessing and bias that employees develop throughout their work. Suriyantphupha & Bourlakis conduct further research on the usage and the role of IT in a traditional retail supply chain [4]. The result of the structured literature review they conducted on this topic reveals that the most widespread technologies are radio frequency identification (RFID), point-of-sale (POS), and other inter-organization systems providing visibility in the value chain, reduce labor, lower operation time, and minimize operating costs. However, such innovative technologies are still limited to major retailers due to implementation costs and the compatibility within existing systems and trading partners' systems. Lal *et al*. focus on the study of the effectiveness of POS data in managing the supply chain [5]. By carrying out a comparative analysis of questionaries' data regarding retail stores in a region in India, they conclude that the most notable benefits of adopting POS systems in the retail sector are improved inventory management, increased flexibility of response to customer demands, and reduction in costs and times. Hu *et al*. leverage two new technologies, blockchain, and edge computing, to propose, design, and develop a novel robust retail POS system, which can be used in 'weak' Internet environments [6]. In his technical report, Khaneja describes an efficient Point-of-Sales system that can generate and maintain transaction receipts, inventory reports, and sales records in the big retail businesses, using UML diagrams and the COMET methodology, consisted of three steps for software development, requirement modeling, analysis modeling, and design modeling [7]. Plomp *et al*. investigate the extent to which Point-of-Sale (POS) systems support retail chain digitization, i.e. inter-organizational processes as being an exchange of orders-related and sales-related information, and consequently, they develop a two-dimensional maturity model for categorizing existing POS systems by their level of backward and forward chain digitization support [8]. Taylor elaborates on mobile payment technologies in the retail sector and reviews potential benefits and risks [9]. By linking research from diverse fields, her paper aims to elucidate the potential impacts of mobile technologies on retail theft and internal technological and process issues, before offering possible solutions.

III. A Brief Overview of Some Existing Retail Systems

In this section, an analysis of the general characteristics of some of the most advanced retail systems available nowadays is conducted. All of these solutions, which are used by millions of users and businesses worldwide every day, are built upon and run distributed systems.

Shopify [10], is one of the most widely used retail systems, developed by Tobias Lütke in Ottawa, Ontario, Canada, in 2004. Over time, the software was continually updated and upgraded, so it nowadays includes more advanced features tailored to the needs of software users. Shopify introduced a handful of new functionalities, such as payments without the need for third-party authentication providers. It is available for several platforms including the most popular ones, such as Android and iOS. Different types of technologies have been used in the development of Shopify, primarily the Ruby on Rails framework. This framework is based on the MVC (Model View Controller) approach to application development, where data transfer uses standards such as JSON and XML. Frontend technologies such as HTML, CSS, and JavaScript have also been used. Shopify uses MySQL as a relational database management system. After several years, the development team started using the PHP language as the primary backend language that has provided a great integration with MySQL databases. The mobile platform applications are developed in Swift/Java.

LightSpeed [11] is an all-in-one cloud-based POS solution, entirely based on the principles of distributed systems, founded by Dax DaSilva in 2005, with its headquarters in Montreal, Quebec, Canada. The technical aspects of the development are similar to those of other retail POS systems. The development tools that are used for the development of this software are the following ones: the TypeScript programming language is used for the frontend. It is a syntactic superset of JavaScript, so the code written in this programming language is also valid as a JavaScript file, i.e. the same files written in TypeScript and JavaScript are mutually compatible; the PHP programming language is used as a backend programming language; the database server runs MySQL RDBMS; Python programming language is used to analyze business data. As with most modern distributed solutions, Amazon Web Services (AWS) is being used as a cloud providing platform.

ShopKeep [12] is one of the most widely used POS systems, especially by small-sized businesses (e.g. retail shops, coffee shops, restaurants, and bars) throughout the United States and Canada. It was developed by Bill Walton in 2008. What is specific about this software is that it is intended for mobile clients, with a special emphasis on tablets (iPads and Android tablets). Distributed software consists of several entities that are synchronized with each other, so they give the end-user a picture as if it were a single whole. The central software is placed on the cloud platform, while the sales part (i.e. the terminal) is the tablet device itself. ShopKeep provides a real-time overview of the entire operation and if the user owns multiple outlets, the data will be grouped so that he can access them at any time. The system allows merchants to ring up sales, print or email receipts, pop a cash drawer, accept credit cards, and print remotely to the kitchen right from an iPad or Android tablet. The web-based BackOffice allows inventory, employee, and customer management, as well as advanced analytics and reporting. The smartphone dashboard app allows merchants to view real-time store sales remotely. SQLite databases are used and installed on tablets, while data on the main server is stored in the PostgreSQL database. The central software and database are set up on the Amazon Web Services (AWS). The MVC framework is Ruby on Rails.

Microsoft Dynamics Retail Management System (RMS) and POS [13] is an application developed by Microsoft that offers small- and mid-size retailers a complete Point-of-Sale (POS) solution that can be adapted to meet unique requirements. It provides centralized control for multi-store retailers and integrates with Microsoft Office applications. It also offers benefits in ease of use, automation, efficiency, flexible reporting, and scalability. It can be deployed for any form of a retail enterprise, from pharmaceuticals to grocery stores. The system, however, is not cloud-based and also targets only businesses that run the Windows operating system. Therefore, it is not a cross-platform solution.

Epicor Cloud Retail Software [14] is a SaaS-based retail solution, developed by Epicor Solutions. It serves small- to mid-sized retailers who want to leverage efficiently their insufficient IT resources. Epicor delivers a model that significantly reduces capital investments, implementation challenges, and ongoing requirements of managing IT. With Epicor, retailers can integrate their sales channels, order management, POS systems, inventory, and other operations to access the right information at the appropriate time. Epicor cloud retail software supports merchandising, store operations, CRM, audit and operations management, and planning. The system, however, is difficult to customize to suit well a particular organization.

IV. Capabilities of Modern Retail Systems

The major functions each retail POS systems is expected to support are the following ones:

- *Transaction execution*: Today, almost every retail POS system has an integrated scanner that automates the ringing up of sales by reading the bar code on each item a customer is purchasing. The Stock Keeping Unit (SKU) number and price are then entered directly into the POS software, which keeps a running total of the order and calculates the final amount due. Some retail POS systems include a touchscreen component that is used, in conjunction with or in stores with very limited inventory, as an alternative to scanning. Instead of scanning items (or manually entering prices, as with an electronic cash register), store associates tap a corresponding icon on the touchscreen to 'build' the transaction. Unlike electronic cash registers, many retail POS systems also feature integrated payment processing capabilities. With these systems, retailers need not use a separate credit card reader to read the magnetic stripe on credit cards. Rather, this part of the transaction can be handled right on the POS terminal;

- *Inventory control*: In many instances, retail POS systems incorporate software modules that allow retailers to better manage their inventory: each time a customer purchases a particular item, information about it (e.g., its SKU number and the quantity bought) is automatically transmitted from the POS software to the inventory management module. The quantity purchased is then subtracted from the inventory 'tally', without any human intervention. Some retail POS systems generate an alert when inventory levels reach user-defined minimums; some also generate re-orders based on this information;
- *Labor management*: Over the past few years, vendors have introduced configurations in which the POS system features labor management functionality through time and attendance and/or scheduling module. With the former, employees clock in and out right on the POS terminal, making it easy to keep tabs on attendance and limiting employees' ability to punch in and out for each other or work beyond their scheduled hours. An interface with the POS system permits retailers to compare store traffic with sales patterns, and then adjust schedules based on historical information and real needs rather than guesswork;
- *Reporting*: A good POS system gives retailers the option to generate a myriad of basic and advanced reports and, in some cases, transfer them directly to other systems. Standard reports created by the retail POS system break out customer purchase histories and merchandise sold (by SKU), as well as indicate the cost of goods sold, gross sales, low inventory counts, existing inventory counts, customer purchase history, and item-specific sales. Some POS systems allow users to customize their reports and create new ones;
- *Marketing*: Retail POS systems track customers' purchase histories, along with their contact information. With this information in hand, retailers can identify the audience for a variety of marketing campaigns, ranging from the somewhat general (e.g. all female customers) to the very specific (for example, all customers who reside in a specific geographic area). A comprehensive POS package will allow for e-mail marketing, as well as direct mail reports so that marketing e-mails may be composed on and sent directly from the POS system. It is also helpful in face-to-face communication with customers. Most POS systems accommodate the addition of miscellaneous notes about each customer (such as his or her birthday and other specific details), that can be leveraged in marketing campaigns.

V. Distributed Retail System

The main objective of this paper is to present a model of a distributed retail system and the functionality of each of the individual entities of the model [15].

A. Model of a Distributed Retail System

In this section, we are presenting a model of a distributed retail system that we have used during the development and implementation phase. The model, presented in Fig. 1, consists of two types of nodes:

- Client nodes
- Central node

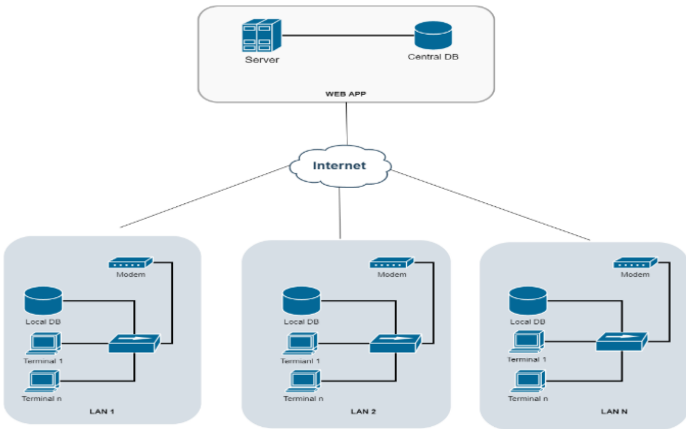


Figure 1. Model of a distributed system

The *client node* is composed of two main components.

The first component is the *terminal*. The terminal can be any kind of a desktop or a laptop computer, on which the *client software* is installed, which is part of the distributed sales system. Each terminal has a local database installed on it, which contains all the data needed for work/sales. The data in the local database is synchronized with the data in the central database, which is located on the central node of the distributed system. Additionally, each terminal may have a cash register and a bar-code reader. The client node can consist of *N* terminals, depending on the number of payment points required.

The second component of the client node is the *computer network*, which enables network communication and data exchange between the terminal(s) and the central node of the distributed retail system.

The *central node* of the distributed retail system consists of an *application server* and a *database management server*. The application server is responsible for the synchronization of the local databases from the client nodes with the central database in the central node, as well as for the functioning and coordination of the distributed retail system. A central database is set up on the database management server, which synchronizes and stores the database data from the client nodes [16].

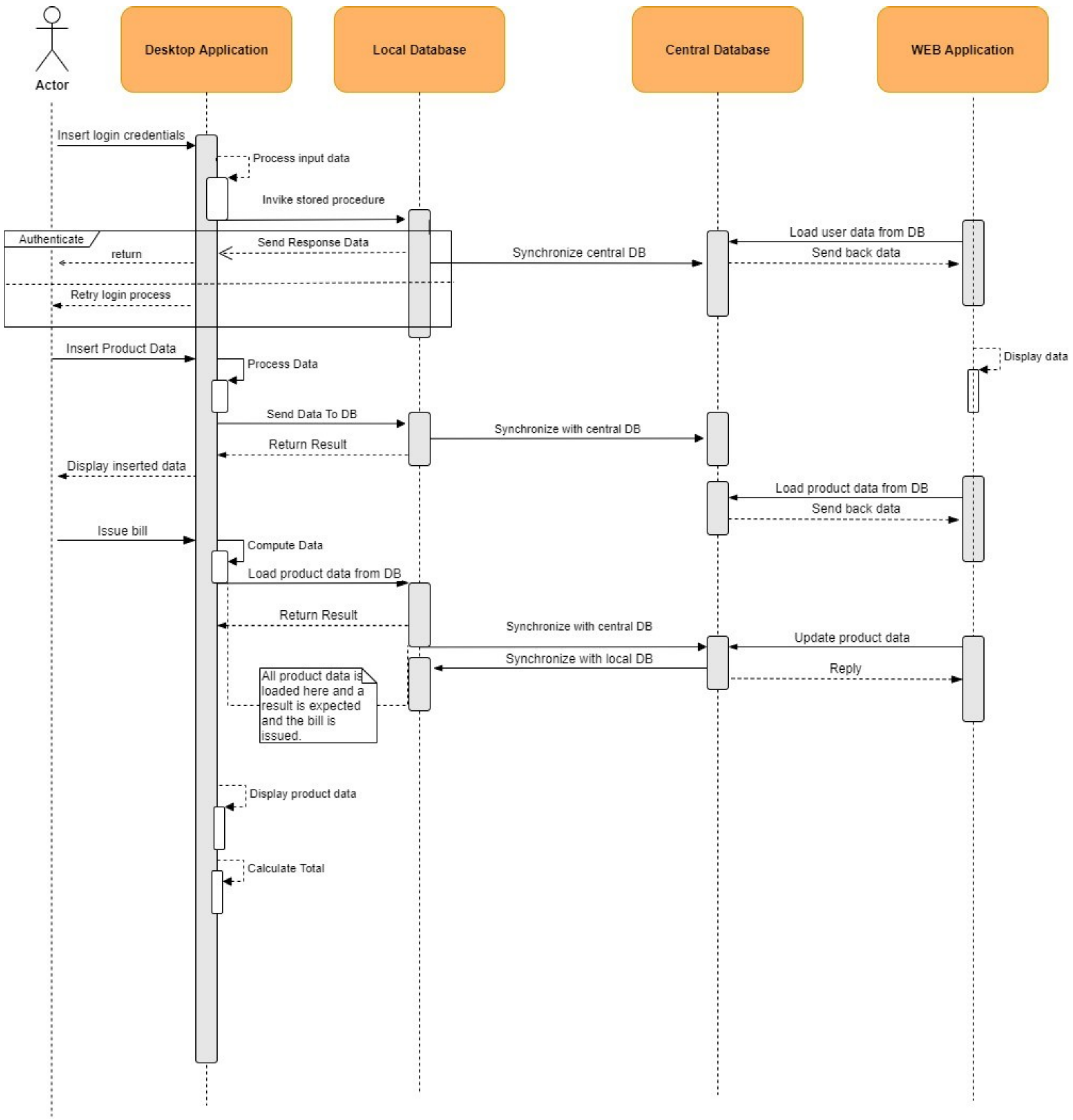


Figure 2. Sequence UML diagram

B. Functionality of the System Entities

This section presents the entities that have been developed and represent an integral part of the distributed retail system. Using the Sequence UML diagram, shown in Fig. 2, the interaction and communication between the individual model entities of the distributed retail system are shown. The distributed retail system entities are interconnected and synchronized. The entities, i.e. the separate parts of the designed solution, can be physically located in different locations, but at the same time, they are functioning as if they represent a single monolithic system.

The distributed retail system consists of the following entities:

- Desktop application
- Web application
- Local database
- Central database

The *desktop application*, schematically depicted in Fig. 3, is the client software that runs on the terminal.

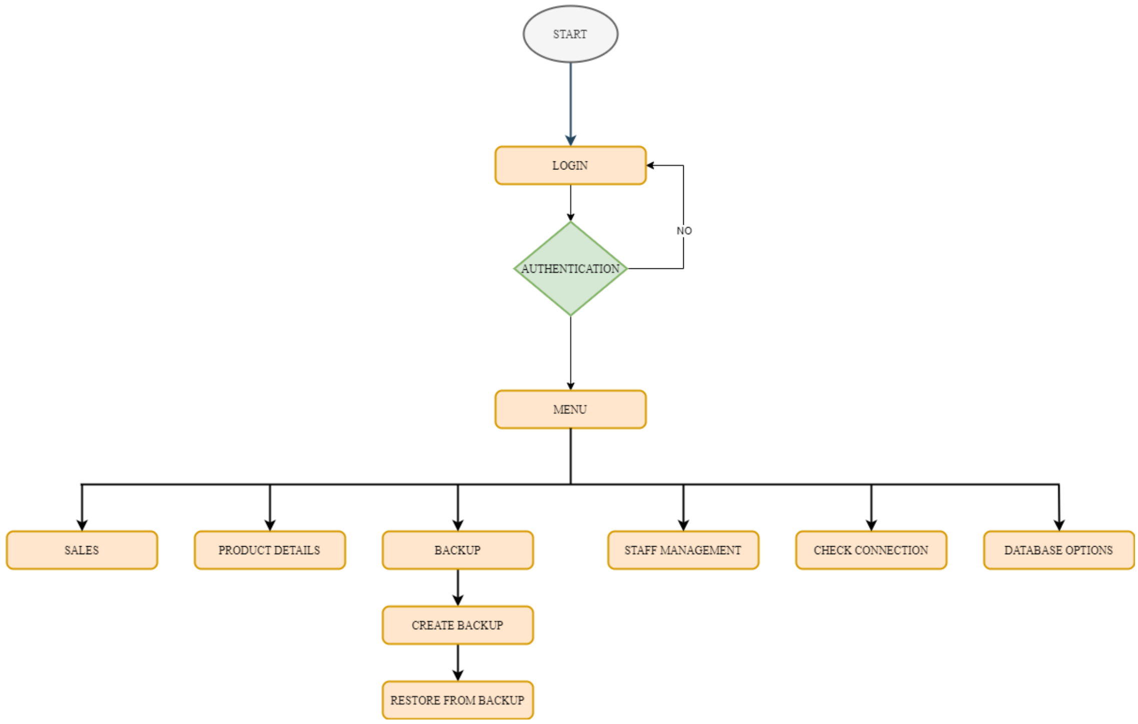


Figure 3. UI flowchart

User Login is a process in which a previously registered user logs into the desktop application. The desktop application performs authentication. If data entered by the user is correct, the working session begins. All processes covered by the login process are synchronized with the central database. In the proposed distributed retail system, two roles are defined, a user and an administrator. The UML Use Case diagram, shown in Fig. 4, provides a clear picture of each of the roles and privileges of each role in the system. The user role has reduced privileges as compared to the administrator role.

Terminal Sales is a process that enables the user to issue all sorts of receipts for items used for sales. Also, this process keeps track of the number of times each item has been sold.

Connection Checker is a module that checks the connection between client nodes and the central node, which is critical to their successful synchronization.

Through the web application, which is a part of the distributed retail system installed on the central node, items used for sale are registered, and the information is distributed to the desktop applications installed on the client nodes. The web application monitors database synchronization, generates product sales reports, registers users in the distributed system, and configures retail system parameters.

Central and local databases are one of the most important components because they contain the data and reflect the business model. Each terminal has a local database installed on it. The central database is located on the central node of the distributed retail system. By entering or changing any data in any local database, synchronization with the central database is done automatically by database replication. In our distributed retail system, we are using asynchronous replication, i.e. the local database copies the data to the central database after the data is already written to the local database. The replication process occurs following a previously defined schedule.

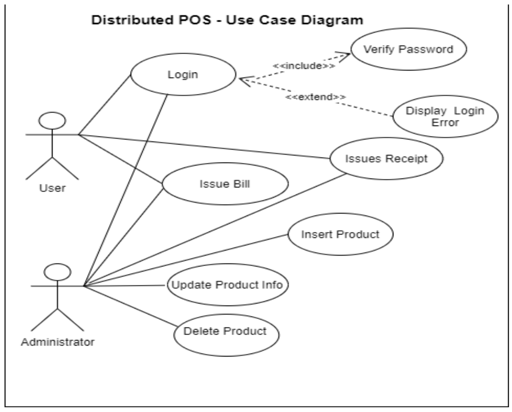


Figure 4. Use Case Diagram

C. *Development Technologies*

This section briefly lists the selected technologies used in the development of the developed distributed retail system.

Frontend technologies such as HTML, CSS, and JavaScript were used to create the web application. We used the Bootstrap development framework which allows creating interfaces that are compatible with all devices. Both .NET and C# programming language were used to create the backend part of the web application, as well as to create the desktop application. MySQL server was used as a relational database management system for running central and local databases [17–21].

VI. Conclusion

Retail shop systems dramatically changed over the past decade. Today's modern solutions provide users with the ability to control their business from literally any point, using even their mobile devices. The contemporary cloud-based and distributed solutions deliver efficiency and, at the same time, save money and time. Distributed retail systems can operate and be efficient even if certain parts of the entire system fail. The future work vis-à-vis the hereby described software framework includes addressing its scalability and portability issues in terms of their improvement, as well as the deployment of the whole system to a cloud-based service. One of the projected directions for further development of the proposed solution is also its adaptation for mobile platforms. All aspects of the proposed retail management software framework need to be tested and validated in the real operating surrounding, so the usability study of the system will also be conducted as a part of the future work. As a way of improving the core features of the proposed software framework, integrating third-party tools, such as a payment gateway to process card payments, and implementing SMS and/or e-mail services to deliver instant messages or reports to the users, will be considered, as well.

REFERENCES

[1] A. Adewumi, S. Ogbuchi, and S. Misra, "A Cloud-Based Retail Management System," in R. Intan, C. H. Chi, H. Palit, L. Santoso, Eds. Intelligence in the Era of Big Data. ICSIIT 2015. Communications in Computer and Information Science (CCIS), vol 516, pp. 446–456, Springer, Berlin, Heidelberg, March 2015.

[2] J. E. Dion, "The Effects of POS Implementation and Retail Technology on Sales and Profitability for Small to Mid Sized Retailers," Fargo, ND, USA, Dionco Inc., White Paper, April 2003. [Online]. Available: http://www.dionco.com/public/articles/POS_Study_White_Paper.pdf

[3] M. Cote, "The Power of Point of Sale: Improving Growth, Profit, and Customer Service in a Retail Business," B.S. thesis, Bryant University, Smithfield, RI, USA, April 2015. [Online]. Available: https://digitalcommons.bryant.edu/cgi/viewcontent.cgi?article=1016&context=honors_marketing

[4] P. Suriyantphupha and M. Bourlakis, "Information Technology in a Traditional Retail Supply Chain: A Structured Literature Review," Projectics / Proyéctica / Projectique, vol. 22, no. 1, pp. 89–102, 2019.

[5] M. K. Lal, A. C. Shukla, and B. E. Tarangini, "Study of Effectiveness of POS Data in Managing Supply Chain," Industrial Engineering Journal, vol. 11, issue 10, pp. 1–9, October 2018.

[6] B. Hu, H. Xie, Y. Ma, J. Wang, and L.-J. Zhang, "A Robust Retail POS System Based on Blockchain and Edge Computing," EDGE 2018, LNCS 10973, pp. 99–110, 2018.

[7] D. Khaneja, "Point of Sales Terminal (POST) Using UML," Project Report, Carleton University, Ottawa, ON, Canada, December 2017. [Online]. Available: https://www.researchgate.net/publication/323073643_POINT_OF_SALES_TERMINAL_POST_using_UML

[8] M. G. A. Plomp, G. van Rijn, and R. S. Batenburg, "Chain digitisation support by point-of-sale systems: an analysis of the Dutch product software market," International Journal of Information Technology and Management, vol. 11, no. 4, pp. 257–272, 2012.

[9] E. Taylor, "Mobile Payment Technologies in Retail; A Review of Potential Benefits and Risks," International Journal of Retail and Distribution Management, vol. 44, no. 2, pp. 159−177, 2016.

[10] K. Shatrov, "E-Commerce at Scale: Inside Shopify's Tech Stack - Stackshare.io", Shopify Engineering, August 2018. [Online]. Available: https://engineering.shopify.com/blogs/engineering/e-commerce-at-scale-inside-shopifys-tech-stack

[11] LightSpeed, "Reinvent your business with the omnichannel retail POS system," 2020. [Online]. Available: https://www.lightspeedhq.com/pos/retail/

[12] ShopKeep, "Do More With ShopKeep's Retail POS System," 2020. [Online]. Available: https://www.shopkeep.com/business-types/retail-pos

[13] Microsoft, "Microsoft Dynamics RMS and POS," 2020. [Online]. Available: https://mbs.microsoft.com/customersource/Global/RMSPOS

[14] Epicor, "Epicor Retail Cloud: Retail Growth Simplified," 2020. [Online]. Available: https://www.epicor.com/en-us/retail-software/epicor-retail-cloud/

[15] A. S. Tanenbaum and M. van Steen, *Distributed Systems: Principles and Paradigms*, 2nd ed. London, England: Pearson, 2006.

[16] H. Garcia-Molina, J. D. Ullman, J. Widom, *Database Systems: The Complete Book*, 2nd ed. London, England: Pearson, 2008.

[17] J. P. McManus and C. Kinsman, *C# Developer's Guide to ASP.NET, XML, and ADO.NET*. Boston, MA, USA: Addison-Wesley Professional, 2002.

[18] D. Avison and G. Fitzgerald, G., *Information systems development: Methodologies, Techniques and Tools*, 4th ed. New York, NY, USA: McGraw-Hill Education, 2006.

[19] T. Barnaby, *Distributed .NET Programming in C#*, New York, NY, USA: APress Media, LLC, 2002.

[20] J. Bender and J. McWherter, *Professional Test Driven Development with C#: Developing Real World Applications with TDD*, 1st ed. Indianapolis, IN, USA: Wrox/Wiley Publishing, Inc., 2011.

[21] B. Schwartz, P. Zaitsev, and V. Tkachenko, *High Performance MySQL: Optimization, Backups, and Replication*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2012

Evaluation of the Ontology Visualization using Key Performance Indicators

B. Grujovski, S. Savoska, A. Bocevska and N. Blazheska Tabakovska
Faculty of Information and Communication Technologies, Bitola, Republic of North Macedonia
"St. Klement Ohridski University" –Bitola
blagojce.grujovski@gmail.com, {snezana.savoska, andrijana.bocevska, natasa.tabakovska}@uklo.edu.mk

Abstract - Semantic Web supports methods that go beyond the traditional web application in a way that it can facilitate machines to understand the meaning of information on the Internet. Ontology is a package of data together with their relationship, structure, and constraints. Ontology makes information a meaningful knowledge which can not only convey semantic meanings but can be interpreted and understood by machines as well. Ontologies are a key technology for the semantic web, as they are responsible for providing this context. Visualization is an important task related to ontologies. Visualization of ontologies is needed for showing their content and relations between their elements. In this paper we generated visual representation of three ontologies using Protégé as one of the most popular tools for ontology visualization. Protégé can be extended with pluggable components to add new functionalities. The evaluation of the ontology visualizations are implemented according Key Performance Indicators (KPIs) such as: interaction possibility, insight in data, interoperability, visibility.

I. Introduction

Ontology is a conceptualization of a domain into a human understandable, machine-readable format [1]. Ontologies are semantic data models that define the types of things that exist in our domain and the properties that can be used to describe them [2]. Basically ontologies are defined model that organizes structured and unstructured information through entities, their properties and the way they relate to one another. There are many reasons why any user could benefit [3]:

- Managing content more effectively;
- Maximizing findability and discoverability of information;
- Increasing and reusing of less seen and unknown information; and
- Improving search engine operations

Before the creation of an ontology, it is necessary to understand its components. An ontology consists of Class, Relationship and Attribute; the components of an ontology allow us to fully define a domain of knowledge context through the entities defined in classes, relationships between classes, and data associated with classes. These three elements allow us to use abilities greater than traditional knowledge organization techniques such as folder structures and taxonomies [4].

Broadly ontologies can be divided into two main categories [5]:

- Lightweight Ontologies: Include class hierarchy or taxonomy of classes, subclasses, values and attributes.
- Heavyweight Ontologies: Represent the domains in depth-wise manner by including axioms and constraints.

Protégé is a free knowledge-modeling tool developed at Stanford University. Ontologies and knowledge bases can be edited interactively within Protégé and accessed with a graphical user interface. Protégé can be extended with pluggable components to add new functionalities and services. There is an increasing number of plug-ins offering a variety of additional features, such as extra ontology management tools, querying and reasoning engines, visualization tools *etc*. There are various forms such as Resource Description Framework (RDF), Ontology Web Language (OWL) and XML Shema in which Protégé ontology can be exported [6].

In this paper we developed an ontology using data that is part of the COST Action CA15211, Atmospheric Electricity Network: coupling with the Earth System, climate and biological systems (Electronet[1]), and we used two other ontologies that are in the same range and are free to download, the Evidence & Conclusion Ontology (ECO[2]) and Exposure Ontology (ExO[3]). These three ontologies are visualized with the same plug-ins in Protégé and some other online tools and programs.

II. Related Woks

In the simplest case [7], an ontology describes a hierarchy of concepts related by subsumption relationships; in more sophisticated models, suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation. Broader explanation and term definitions can be found in [7]. In this book, the different ontology types are explained as well as the processes to establish

[1] https://dataspace.atmospheric-electricity-net.eu/
[2] https://www.ebi.ac.uk/ols/ontologies/eco
[3] https://bioportal.bioontology.org/ontologies/EXO

relations between them with the objective of selecting those that are more suitable for the information retrieval context. A study of families and types of ontology models is shown to compare them and detect common characteristics and differences. This analysis has the objective of providing the context in which the terminological ontologies are placed and showing how each model is related to the rest. In [8] a classification of types of ontologies according to the degree of formalism and semantics provided in their specification is proposed. They range from simple lists, passing by subject sets, to complex reasoning models.

A complementary classification is described in [9]. This model focuses on explaining the different types of ontologies from the semantic interoperability point of view and focusing on the ability to express hierarchical relations. The categories range from taxonomies, which are able to express few semantics (subclassification relationship) and only provide syntactic interoperability, to logical theory models, which thanks to their strong semantics provide the most complete form of semantic interoperability. In this categorization, terminological models are those that provide syntactic and structural interoperability, and the axiomatic ones those that provide semantic interoperability.

There exist many different ontologies, built for many different types of applications, and they vary in both the amount of detail they express and the generality of their use [10]. A key feature of ontologies is that, through formal, real-world semantics and consensual terminologies, they interweave human and machine understanding [11]. This important property of ontologies facilitates the sharing and reuse of ontologies among humans, as well as among machines. A major reason for the recent increasing interest in ontologies is the development of the Semantic Web [12], which can be seen as knowledge management on a global scale. Tim Berners-Lee, inventor of the current World Wide Web and director of the World Wide Web Consortium (W3C), envisions the Semantic Web as the next generation of the current Web.

The exponential growth of information available on the Internet makes it increasingly necessary to introduce intelligent agents to facilitate the processing of information and knowledge [13]. Other researches include: Ontology visualization Protégé tools – a review [14] and older approaches of ontology visualization and visualization tools: a survey of the state of the art [15].

III. Ontology Visualization

Although every ontology visualization tool is different and unique and no general formalization exists of the way ontologies are displayed, there can be identified some commonalities between them, forming a so-called ontology visualization method. This classification is based on three main criteria: dimensions used by the visualization method (2D, 2.5D, 3D), graphical elements (node-link) and the layout method used on the elements on the screen (force-directed, treemaps, radial) [15].

A. Ontology visualization tools in Protégé and other web tools

Visualization can contribute a lot, and there are many benefits and advantages like better cognitive memory and resource processing, easier information searching, better pattern recognition, enabling of perception inference operations like close watch mechanism, information coding in manipulation medium, large information processing, easy focus on details and work process monitoring and abstract presenting of a situation by following and annotating information. There are many challenges related to the size of the ontology, the technical aspects and the need for using different techniques for interaction with different use cases.

The continuous daily increasing number of data creates the information flood which brings the need of more effective and efficient methods for handling data. Knowledge is distributed through different media, digital and analog TV, via Internet or physically, newspapers or books. This big volume of information and the fact that information is reproduced easily through the mentioned media, is a great information resource for the human. But with the big volume of information and knowledge, the well known problem of its inefficient use and chaotically structured data that we can not understand has appeared. The visualization of information is brought up as an intuitive alternative and need for comprehension of the data. Visualization in fact means reduction of the time needed for analyzing and kind of “translation” of complex semantic relationships into an interactive visual map that is more readable. Ontologies are a great way of presenting the relationships between data while visualizing them for better understanding and use. In this part we visualize three ontology examples: Electronet, ECO and ExO. All three ontologies are related to ecology data and influence on living beings. The ontologies are visualized using the tools from Protégé and other online tools (WOWL, The Brain, OLS-graphview).

Protégé is designed so everyone can use it and easily create an ontology using OWL. The demand for the ontologies written in OWL is rapidly growing. In this paper the focus is on the ontologies created with OWL. OWL is constructed on top of RDF-Schema, and therefore it shares many RDF properties and it can easily be translated into RDF. Its name suggests that it was designed to be a language for representation of ontologies for the Semantic Web. OWL can be used to code information about its domain, like metainformation about entities, called annotations [16].

Electronet is an ontology that was developed in order to quantify the atmospheric electromagnetic fields and their influence on living organisms and to show data and its relations in a way that they will be more understandable and readable, providing reusability and research of data as shown in Fig. 1. Every tool has different way of visualization and different way of insight in data (zooming, analyzing *etc*.).

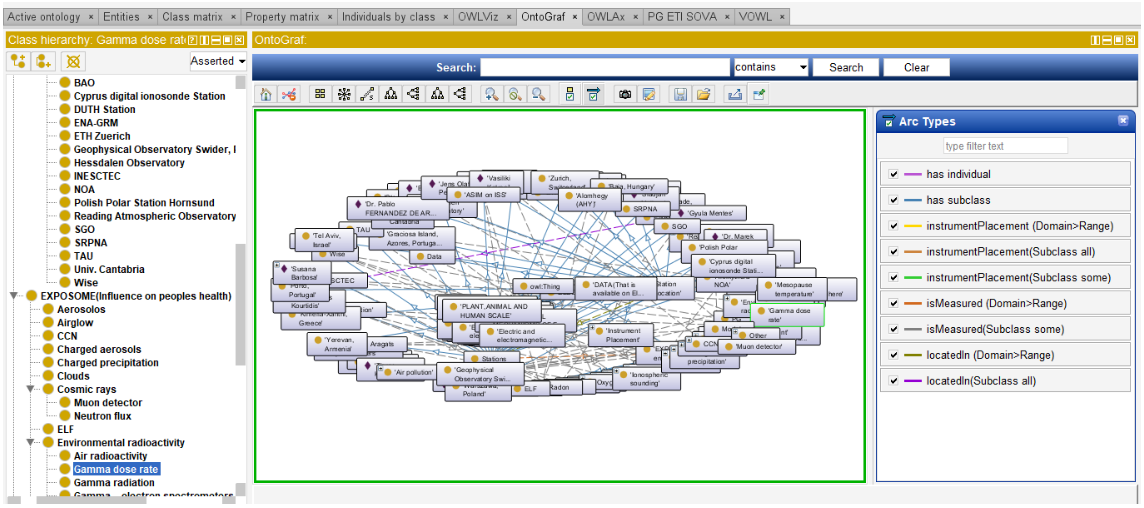


Figure 1. Electronet visualization with the OntoGraf plug-in of Protege

The Evidence & Conclusion Ontology describes types of scientific evidence within the biological research domain that arise from laboratory experiments, computational methods, literature curation, or other sources. Researchers use evidence to support conclusions that arise out of scientific research. Visualization with the tool OWLViz can help the user see the hierarchy in which the ontology is displayed and distinguish classes and subclasses, and their relations, as shown in Fig. 2.

Figure 2. Visualization of ECO with the OWLViz plug-in

The Exposure Ontology was developed to respond to the needs expressed by the National Institute of Environmental Health Sciences (NIEHS) and partner agencies for inclusion of exposure data when prioritizing research and performing toxicity testing. It also addresses the need for centralization of exposure data in a broader biological context and provides "real-world" exposure context for data in The Comparative Toxicogenomics Database (CTD). The resulting resource enables new opportunities for understanding and prioritizing human health effects from exposure and their underlying etiologies, and coordinates data key to enhance the capacity for toxicity prediction and risk assessment. Its visualization is shown in Fig. 3.

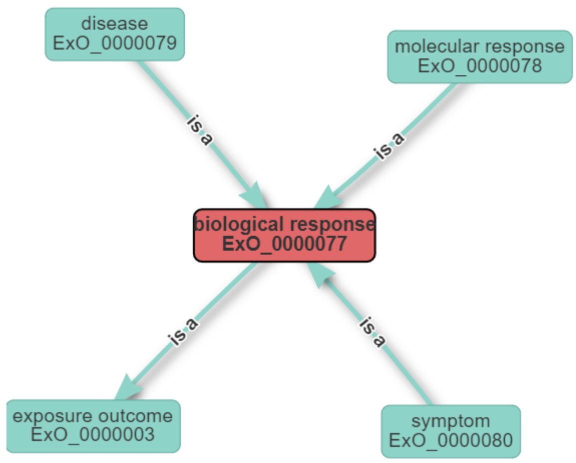


Figure 3. Visualization of ExO biological response class using OLS-graphview

The visualization capabilities are great and while it is easy to create an ontology thanks to the Protégé interface in the background there is code being generated for every class, subclass, data property that we add. This code is written in OWL with the RDF Scheme, Fig. 4, but it also can be translated into OWL with the XML Scheme, Fig.5.

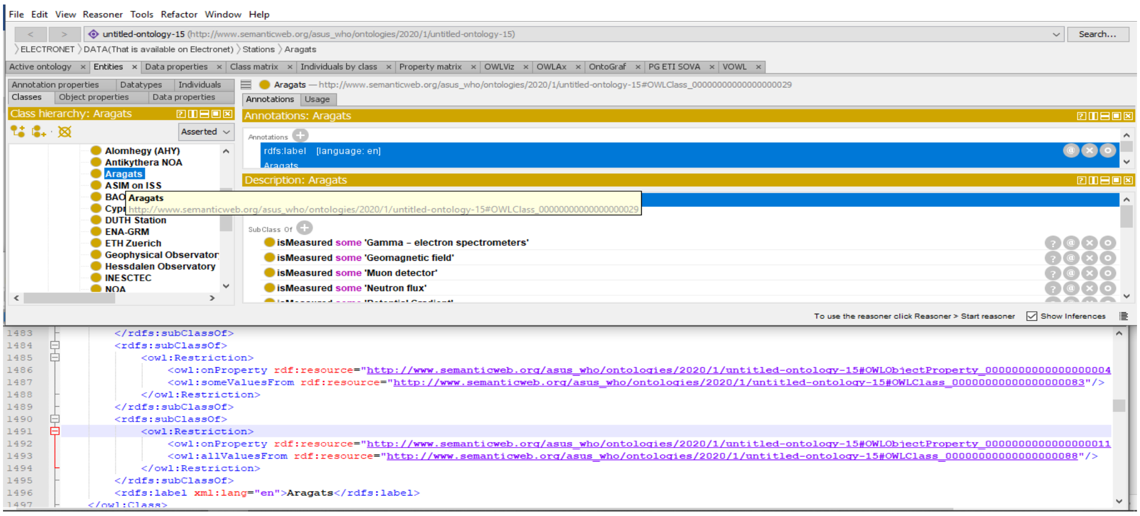


Figure 4. UI view of a class and a code generated in the OWL RDF Schema

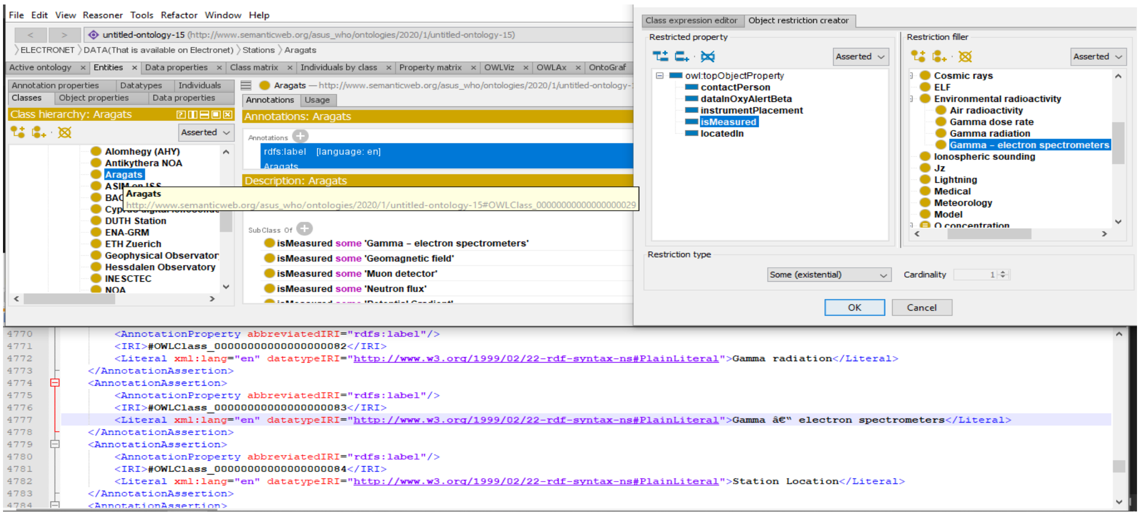


Figure 5. UI view of a class and a code generated in the OWL XML Schema

IV. KEY PERFORMANCE INDICATORS EVALUATION OF THE ONTOLOGY VISUALIZATION

In business and the scientific world, the term KPI refers to measuring the key indicators of performance. According to Oxfords[4] dictionary KPI is a quantifiable measure used to evaluate the success of an organization, employee, *etc.* in order to evaluate the performance of the wanted activities for success. Adequate, clear and relevant information that is obtained and is carefully defined is important information for the course of performance management activities and is more likely to be appropriate so it can be absorbed and acted upon by relevant persons [17].

KPIs are indicators that measure performance according to the key activities and initiatives that are being taken. They follow the intuitive idea that good work will lead to good results. KPIs have clearly defined thresholds for all measured indicators. They can be related to a specific time, shorter or longer, as they are commonly used to observe continuous tasks that may last a shorter or longer period of time. The current importance of KPIs has been extended to the world of facts and science to show the retrieval of evidence as coefficients of evidence reliability derived from existing data with the methods and techniques used to obtain evidence, quantifying the impact of obtaining scientific evidence [18].

KPIs are also important for semantic analysis because business rules and policies are derived from KPI results. However, the information they provide for company purposes does not always have to be accurate, as KPIs measure only a subset of factors that affect the outcome of a Key Result Indicator (KRI) [19].

Many classifications and taxonomic methods used in the visualization of data and information usually start from the data itself and the used techniques. There are many different approaches to create a taxonomy of the three types of visualization (information, data and scientific), according to the preferences of users and developers. It is very difficult to deal with the daily outflow of information, which is subject to analysis by managers and other users. That is why a proactive policy

[4] https://www.lexico.com/definition/key_performance_indicator

is needed to prepare visual reports and effective presentations [20]. KPIs are needed in assessing the impact of ontology visualization on users, their understanding of ontology data and their insight in it, as well as in performing basic statistical or mathematical analyzes.

When creating a KPI for the visualization we need to evaluate the possibility to easily obtain the information and easily access the data, not only by professionals but also by a non-professional user, to assess the visibility of the data, whether it is possible to work with multiple screens, whether there is a possibility to select objects and whether the links between them will be visible and how far can we go through the selected data. The level at which connections can be understood and the ability to discover connections between entities are also important, as well as data aggregation, their clustering, classification or association. These allow hidden links between data to be detected. Data visualization is presented on the so-called visualization wall, which provides opportunities for selection and zooming of a particular cluster, data detection and the ability to analyze them.

Fig. 6 illustrates the use of KPIs needed to evaluate the visualization of ontologies, explaining the indicators and their codes that will be used to evaluate the KPIs according to their usefulness and relevance.

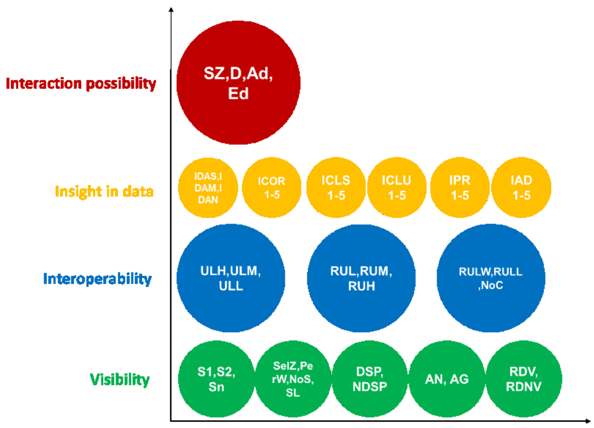


Figure 6. Visual representation of KPI of the second dimension [21]

The selected indicators aim to help assess the success of the ontology visualization. The Visibility indicator is presented as the ability to use one or more visual displays to present data and visualize it, Fig. 6. Another attribute is the possibility to select the navigation object through the made dimensions and calculations, the level of detail of the presented data and the possibilities to display the relations between the data on the diagrams. The Interpretability indicator can be divided into three main components: data comprehension level, link comprehension level and aggregation detection. The insight in data indicator is related to the ability to analyze and detect mathematical and other relationships between data such as statistical indicators, correlations, classifications, rankings, associations. Interaction possibility enables selection, navigation, zooming, filtering of data, selection of analysis attributes and other possibilities of interaction with data.

A. Validation of the visualization of the ontologies made for the needs of the paper

Table I was composed to validate the designed ontology visualizations in order to show how the KPI indicators for an ontology visualization are used in practice.

From this table, it can be seen which visualizations have better KPIs and which are more suitable for different purposes. The value of these KPI estimates can be very useful if correlated with software tools that involve the application of artificial intelligence in estimating the types of ontology visualizations appropriate for a particular user purpose.

The numbers in range from 1-5 are written as a grade to rate our experience and our perception of the used tools in Protégé and online ones. We graded the tools in accordance to the KPIs that have been created; 1 being the lowest grade in our opinion or the tool we recommend less probably. The score goes up to 5, graded as the tool we recommend as the best used for the appropriate KPI.

Related to the visibility, the criterion *Screens* is rated 1 for all tools, because they only support one screen at a time. The grade 4 for *Objects* and *Aggregation* visibility refers to very good object visibility or very good possibilities for gathering and consolidating data, which can still be further improved. For tools with grade 5 for *Relation* visibility relationships are shown and can be distinguished excellently.

The criterion *Data understanding level* for the KPI *Interoperability* is rated 3 when the grouping and arrangement is not that good because of the UI quality of the web page. Relations that are visible but not understandable and have code labels are rated with 3 for *Relationship understanding* level. Some tools are graded with 2 and 3 for the criterion *Aggregation*, based on how though can be the gathering and processing of data for beginners.

For *Insight* in data, some tools are rated 2, due to their lack of operations support, needed for the criterion *Statistics* and mathematics analysis.

Related to *Interaction* possibility, when some selection can be done, the criterion *Selection* is rated 3. The grade 2 for deleting, adding and editing for some tools is given as a result of the interaction, which is provided, but can be improved, because there are other tools that do these operations very good.

TABLE I. EVALUATION AND VALIDATION OF ONTOLOGY VISUALIZATIONS ACCORDING TO THE DEVELOPED KPIS

	Visibility					Interoperability			Insight in data						Interaction possibility			
Ontology Visualization	Screens	Objects	Data Sel.	Agreg. Viz.	Relation Viz.	Data Und. level	Rel. Und. level	Agreg.	Stat. Mat. Analys.	Correlation	Class	Cluster Analysis	Pattern recog.	Assoc.	SZ	D	Ad	Ed
OWLViz in Protégé ELECTRONET	1	4	4	4	5	4	4	3	2	3	4	4	3	3	3	2	2	2
OntoGraf in Protégé ELECTRONET	1	4	5	5	5	5	5	3	4	4	4	4	3	3	4	4	3	4
VOWL in Protégé ELECTRONET	1	4	4	3	4	3	4	3	2	3	4	5	2	3	3	2	2	2
VOWL на Web in Protégé ELECTRONET	1	4	4	4	4	3	4	3	3	4	4	5	3	3	3	2	2	2
OWLAx in Protégé ELECTRONET	1	3	4	3	3	4	4	4	4	4	4	3	3	4	4	4	4	4
The Brain application ELECTRONET	1	5	5	5	5	5	5	4	4	4	4	4	4	4	5	5	5	5
OWLViz in Protégé ECO	1	4	4	4	5	4	4	3	2	3	4	4	3	3	3	2	2	2
OntoGraf in Protégé ECO	1	4	5	5	5	5	5	4	4	4	4	4	3	4	4	4	3	4
OLS-graphview on the Web ECO	1	4	4	3	4	3	3	2	2	4	4	4	3	2	3	2	2	2
OWLViz in Protégé ExO	1	4	4	4	5	4	4	3	2	3	4	4	3	3	3	2	2	2
OntoGraf in Protégé ExO	1	4	5	5	5	5	5	4	4	4	4	4	3	4	4	4	3	4
OLS-graphview on the Web ExO	1	4	4	4	5	4	4	3	2	3	4	4	3	3	3	2	2	2

The average values of the rated KPIs shown in Table I can be used to determine which visualisation tools are more suitable and recommended in our opinion for cetrain ontology visualisations. For example if we add up all the grades for The Brain application with an exception of the first attribute (*Screens*) and we calculate the average grade we get 4.6, which means this application is almost excellent for Electronet ontology visualization. Similarly, OntoGraf is the one most suitable and recommended for ECO and ExO ontology visualisations. Its average value is 4.2 for both ontology visualisations.

V. CONCLUSION

The daily influx of data collection from many sources that have big data features and the amount of data flowing into many distributed systems makes it difficult to process and extract the most important information. Today, when computers are an integral part of our daily lives, we have a way to alleviate the problem of processing and reusing data for analysis in other applications by using ontologies. Ontologies have emerged as a powerful method for sharing and understanding data, primarily because of their ability to link an unlimited amount of metadata with concepts and connections between them, but also with the process of gathering knowledge. Connecting and sorting knowledge leads to its better understanding and faster retrieval of necessary information and creation of domain knowledge.

This paper aims to describe ontologies and their connection to the semantic web as technical and conceptual terms and to show possible ways of their visualization. Ontologies are a great way to show the connections between data and at the same time visualize them in order to gain their reuse and understanding. For the obtained ontology visualizations with the tools and the applications that are used, validation and evaluation was done in Table I according to the previously proposed methodology for ontology visualization. The evaluation was translated into KPI performance indicators of ontology visualizations that are subject to the parameters grade. Scores from 1 to 5 are used for all defined KPIs. Representativeness assessment is performed according to pre-defined KPIs, shown in Figure 6. Tools that we recommend to be used for appropriate KPIs are rated with grade 5.

The development of the semantic web is another reason why ontologies are so popular. They intertwine human and computer understanding of symbols, creating a connection that the human eye and computers can jointly understand and use in practice. In future works the visualization of ontologies can be broaden to other great programs and applications that are being developed every year and can help ontology engineers and knowledge managers to comprehend the daily data income and visualize it to extract as much possible information easily.

REFERENCES

[1] N. Guarino and P. Giaretta, “Ontologies and Knowledge Bases: Towards a Terminological Clarification” in Toward Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, Amsterdam: IOS Press, 1995.

[2] B. Schrader, “What’s the Difference Between an Ontology and a Knowledge Graph?”, [Online]. Available: https://enterprise-knowledge.com/cms/assets/uploads/2020/01/Ontologies-vs.-Knowledge-Graphs.pdf

[3] J. Hilger, “What is an Ontology and Why Do I Want One?” [Online]. Available: https://enterprise-knowledge.com/what-is-an-ontology/

[4] B. White, "From Taxonomy to Ontology", [Online]. Available: https://enterprise-knowledge.com/from-taxonomy-to-ontology/

[5] S. Sosnovsky and D. Dicheva, "Ontological technologies for user modeling," International Journal of Metadata, Semantics and Ontologies (IJMSO), Vol. 5, No. 1, 2010.

[6] M. A. Musen, R. W. Fergerson, W. E. Grosso, N. F. Noy, M. Crubézy and J. H. Gennari, "Component-based support for building knowledge-acquisition systems," Proceedings of the intelligent information processing (IIP2000) conference international federation for processing (IFIP), world computer congress (WCC’2000)., pp. 18-22, 2000.

[7] J. Lacasta, J. Nogueras-Iso and F. J. Zarazaga-Soria, “Terminological Ontologies - Design, Management and Practical Applications, Springer, 2010.

[8] O Lassila and D. MacGuinness, “The Role of Frame-Based Representations on the Semantic Web”, Technical Report KSL-01-02, Knowledge Systems Laboratory, Standford University, Standford, California, 2001.

[9] A. Sigel, "From traditional Knowledge Organization Systems towards ontologies on the web”, 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.

[10] D. Fensel, H. Lausen, J. Bruijn, M. Stollberg, D. Roman and A. Polleres, Enabling Semantic Web Services; The Web Service Modeling Ontology, Springer, 2007.

[11] D. Fensel, "Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce", Springer, 2nd edition Berlin/Heidelberg, 2003.

[12] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web, A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities", Scientific American, vol. 5, no. 284, pp. 34-43, 2001.

[13] G. Paquette, "Visual Knowledge Modeling for Semantic Web Technologies: Models and Ontologies", Information Science Reference (an imprint of IGI Global), 2010.

[14] R. Sivakumar and P.V. Arivoli, "Ontology visualization Protégé tools – a review," International Journal of Advanced Information Technology (IJAIT), vol. 1, no. 4, 2011.

[15] M. Dudáš, S. Lohmann, V. Svátek and D. Pavlov, "Ontology visualization methods and tools: a survey of the state of the art," The Knowledge Engineering Review, vol. 33, pp. 1-39, 2018.

[16] I. Horrocks, P. F. Patel-Schneider and F. Harmelen, "From SHIQ and RDF to OWL: The Making of a Web Ontology Language", Journal of Web Semantics, vol. 1, pp. 7-29, 2003.

[17] A. Wille, "What is a KPI? ", [Online]. Available: https://www.klipfolio.com/resources/articles/what-is-a-key-performance-indicator

[18] G. Binefa, M. Garcia, N. Milà, E. Fernández, F. Rodríguez-Moranta, N. Gonzalo, L. Benito, A. Clopés, J. Guardiola and V. Moreno, "Colorectal Cancer Screening Programme in Spain: Results of Key Performance Indicators After Five Rounds (2000–2012)," Sci Rep 6, 19532, 2016.

[19] M. M. Roldán-Garcíaa, J. García-Nietoa, A. Matéb, J. Trujillob and J. F.Aldana-Montesa, "Ontology-driven approach for KPI meta-modelling, selection and reasoning," International Journal of Information Management, 2019.

[20] S. Savoska and S. Loskovska, "Taxonomy of User Intention and Benefits of Visualization for Financial and Accounting Data Analysis", ICT Innovation, 2013.

[21] S. K. Card, J. D. Mackinlay and B. Scheiderman, "Readings in Information Visualization: Using Vision to Think.," Morgan Kaufmann Publishers, San Francisco, 1999.

Intelligent Technique for Human Authentication using Finger vein

Mona A. Ahmed[1], and Abdel-Badeeh M. Salem[2]
[1, 2]Computer Science Department ,Faculty of Computer and Information Sciences ,Ain Shams University, Cairo, Egypt
eng.monaa_cs@hotmail.com, absalem@cis.asu.edu.eg

Abstract - In this paper, we propose a new intelligent technique to authenticate human using Finger vein (FHV) pattern. We developed an image analysis technique to extract region of interest (ROI) from FHV image. After extracting ROI we design a sequence of preprocessing steps to improve finger vein images using Median filter, Wiener filter and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance finger vein image. Our smart technique is based on the following intelligent algorithms, namely; principal component analysis (PCA) algorithm for feature extraction and k-Nearest Neighbors (K-NN) classifier for matching operation .This technique has been applied on the Shandong University Machine Learning and Applications - Homologous Multi-modal Traits (SDUMLA-HMT) database. The experimental results show that the result of Correct Recognition Rate (CRR) is 94.6 %

Key Words: Biometric, dorsal hand vein, computational intelligence, feature extraction, PCA, K-NN, machine learning.

1. INTRODUCTION

Recently, authentication is considered the most important objective to be satisfied whether it is physical world or the internet of things (IoT) world. Different approached and techniques exist to authenticate the user such as passwords, smart cards, and pins. Modern approaches authentication include biometrics like voice, finger prints, retina, iris, facial expressions, signatures, face and vein pattern. Among all the authentication techniques present, biometrics is considered as the most reliable authenticators since they are unique to every individual and hard to get [1].

The technology of Vein Patterns (VP) as a type of biometric authentication was first intended in 1992.VP is the network of blood carriers below a person's skin layers. VP structure distinct and distinguishable patterns across various people and they remain the same irrespective of age. The patterns of blood veins are unique to each person, even among twins. There are internal and external biometric systems. External include face, Iris, finger print based systems. Palm vein, finger vein, dorsal veins structure the internal biometric frameworks. Veins are intra-skin elements, consequently this feature makes the frameworks exceptionally secure, and they are not influenced by state of the external skin [2].Generally, biometric system works in two modes namely :(i) verification mode in which biometrics can be used to verify a person's identity and (ii) identification mode in which biometrics can be used to determine a person's identity, even without that individual's information [3].The ideal FHV authentication system consists of the following components; image acquisition from the database and pre-processing, finding of region of interest, extraction of dorsal hand vein pattern features and recognition.

The rest of the paper is organized as follows. In Section 2, the related works in this field are introduced. We briefly explain the general architecture of the finger vein system in section 3. The explanation of canny edge detector is introduced in section 4.Section 5presentedthe explanation of median filter. Sections 6 introduce how we used Wiener filter. The explanation of Contrast Limited Adaptive Histogram Equalization filter is presented in section 7. The explanation of PCA algorithm is presented in section 8. In section 9 the explanation of matching algorithm K-NN is presented. Section 10 illustrates the steps of extract Region of interest (ROI) and the proposed algorithm of preprocessing image presented .The discussion of results is introduced in section11.Finally; conclusions and future work are presented in section 12.

2. RELATED WORK

Recently finger vein biometric has attracted increasing interest from many researchers and thus considerable development is seen in the past decade. D.Mulyono and H.Shi Jinn introduced preliminary process to enhance the image quality worsened by light effect and noise produced by the web camera [4]. T.Pham et al [5] proposed a FV recognition system using a near infrared (NIR) image sensor. The experimental results obtained with three databases showed that their system can be operated in real applications with high accuracy. X.Meng proposed technique based on that regular deformation, which corresponds to posture change, can only exist in genuine vein patterns [6].G.Ayappan and A.Shankar proposed new algorithm with high performance and optimum accuracy. The proposed image preprocessing method comprises of 8 sub-blocks. Seen from the experimental result the method is very effective as a biometric personal identification system [7].M.M.Ali [8] presented a robust method for FV recognition based on the discrete wavelet transform. The simulation results show that this method is robust and fast for feature extraction and classification .A.S.Bader and A.Sagheer [9]used computer vision algorithms (FAST and

Harris) proved that use of two algorithms together produce a reliable system of finger vein identification. B.Janney et al proposed method aims to implement the secured patient database in hospitals using a FV authentication system. In preprocessing stage, an image filtered by median filtering. Extract the features of an image using DWT method. In matched stage selects feature of an image using Manhattan distance. This distance is used for selecting test image feature and database image feature and calculates distance between test image and database image [10].H.Qin et al [11] proposed a new approach which can extract two different types of FV features and achieves a most promising performance. They proposed a localized matching method to accommodate the potential local and global variations at same time. This paper investigates an approach for FV identification combining SIFT features, shape and orientation of FV. The experimental results suggest the superiority of the proposed scheme.

R.Das et al propose a convolutional-neural-network-based FV identification system and investigate the capabilities of the designed network over four publicly-available databases. The reported extensive set of experiments show that the accuracy achievable with the proposed approach can go beyond 95% correct identification rate for all the four considered publicly-available databases [12].K.Kaur and S.Bhushan proposed a newly developed method. An enhanced Human Identification algorithm is developed using FV which based on Automatic Trimap Generation, Repeated Line Tracking, Gabor and SVM. This algorithm is fast and more accurate with respect to other finger vein identification technique and also takes less time as comparison to other technique [13].Madhusudhan M V et al [14] presented a novel approach to authenticate an individual the algorithm resulted in 99.06% of accuracy in person authentication, we can conclude that the CNN model designed is found to be reasonably good.

Table 1Analysis of dorsal hand vein recognition techniques

Author(s)	Method of preprocessing	Method of feature extraction	Method of matching	Database size	Accuracy
Mulyonoet al (2008)	Median filter	CCL	Rm threshold	1000 images	100%
Pham et al (2015)	Gabor filter	LBP	HD	1200 images 1980 images 3816 images	96.35%
Meng (2018)	Morphological operations	SIFT	proposed method	1872 images 3816 images	99.68% 99.70%
Ayappan et al (2017)	Median filter, Gaussian low pass filter& Morphological operations	CN	MHD	150 images	100%
Ali et al (2017)	DWT	GLCM	Euclidean distance	3816 images	92.4%
Bader et al (2018)	Median filter & Zhang-thinning	FAST &Harris	Manhattan distance	636 images	99.71%
Janney et al (2017)	Median filter	DWT	Manhattan distance	-	-
Qin et al (2013)	Morphological operations	SIFT	SVM	4260 images 7120 images	95.04%
Das et al (2019)	histogram equalization & Morphological operations	Gabor filters	CNN	3132 images 5904 images 3816 images 1440 images	95%
Kauret al (2015)	Repeated Line tracking	Gabor filters	SVM	-	-
Madhusudhan et al (2020)	-	Dropoutfilters	CNN	3816 images	99.06%

3. PROPOSED PARADIGM

In our study, we present a proposed intelligent paradigm to authenticate personal based on finger veins. This paradigm is used to enhance the accuracy of finger vein authentication. Figure 1 shows the methodology of the authentication model using finger veins biometrics. The main characteristics of our model are;

(a) Using the Shandong University Machine Learning and Applications - Homologous Multi-modal Traits (SDUMLA-HMT) Database image which has been capture under a NIR infrared radiation.

(b) To detect ROI we used canny detector to extract useful structural information from finger veins images.

(c) Using Median, Wiener and CLAHE filter to improve finger vein image.

(d) Using PCA method for feature extraction.

(e) Using K-NN classifier, matching was done between SDUMLA-HMTFV Database image and the extracted finger vein image.

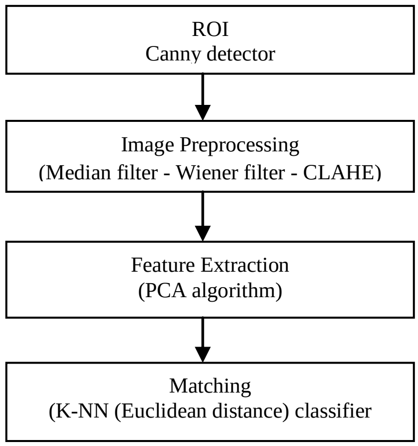


Figure 1The methodology of the authentication model using finger veins biometrics

4. CANNY EDGE DETECTOR (CED)

The edge detection is an important process in many of the image processing algorithms. Significant property of the edge detection is the detection of the specific edges along with the great orientation of the object in the image [15]. In our technique the CED is used due to the following features:

1. Signal to noise ratio of the gradient is maximized, leading to High quality of edge detection.
2. Good localization, localized edges should be with maximal accuracy to real ones
3. Minimal responses, the detector should give a single edge and eliminate false edge caused by noises [16].

The steps of canny algorithm are given as follows:

1. Image smoothing by convolution with Gaussian filter as a result noises are removed.
2. Magnitude and angle of gradient calculation for each pixel of smoothed image in the horizontal and vertical directions
3. Thinning image by application of non- maximum suppression of image’s pixels, that suppress all pixels except the ones with local maxims, as a result thin edges are preserved. This progression removes pixels which are not considered as a major aspect of an edge. Only thin lines will remain, these contain pixels which are viewed as a major aspect of an edge.
4. Hysteresis- The final step, two thresholds is utilized named as lower and upper.
   a. If the gradient value of a pixel is higher than the upper threshold, then the pixel is considered as an edge pixel.
   b. If the gradient value of a pixel is less than the lower threshold, then the pixel is rejected.
   c. If the gradient value of the pixel is between the lower and upper thresholds, then the pixel will be accepted only if it is connected to a pixel that is above the upper threshold.

5. THE MEDIAN FILTER

The digital image processing could be defined as the processing of the images using computer. These images may be corrupted with noise during its acquisition and transmission and also due to blurring artifacts. Most of the applications need to be applied on the original images rather than the noisy ones. Many types of noise could corrupt the images and several filtering techniques could be implemented to eliminate or reduce these noises [17]. Median filter is classified as nonlinear filter; it is computing the median of the corruption block, then the median value will be the new value for the center pixel of the current block, repeat the step until maintain all the corruption image region[18].

The steps of median filter algorithm are given as follows:

1- Place a window over pixels.
2- Sort the pixels value ascending or descending
3- Compute the median
4- The median value will be the new value of the center pixel of the window.
5- Repeat the above process for all corrupting image area.

6. WIENER FILTER

In our technique, we use the WF to calculate the statistical estimate of an unknown signal using a related signal as an input and filtering that referred to signal to produce the estimate as an output. The WF executes an ideal tradeoff between inverse filtering and noise smoothing. It removes the additive noise and modifies the obscuring at the same time. The WF is ideal as far as the mean square error [19].The goal of the WF is to compute a statistical estimate of an unknown signal using a related signal as an input and filtering that referred to signal to produce the estimate as an output WF is portrayed by the following:

1. Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation
2. Requirement: the filter must be physically realizable / causal .
3. Performance criterion: minimum mean-square error (MMSE)

7. CLAHE

The complexity improvement one of the significant sorts of processing technology for images. It can effectively improve the visual quality of an image for human authentication. Along with this, preprocessing strategy is significant for automatic pattern recognition, to get fundamental features images and also for other applications. To enhance the contrast of an image many contrast improvement technology has been presented. Histogram equalization is one of the essential strategies [20]. In this method contrast adjustment can be done using image histogram in image processing. One of the local histogram equalization based image upgrade technique is contrast limit adaptive histogram equalization (CLAHE). In CLAHE strategy which cuts the histogram above the clip limit furthermore dispersed to the some other histogram of varies regions which will have histogram value below the clip limit. CLAHE uses contrast intensification constraining methodology that is applied for each neighboring pixel which then forms a transformation function in order to reduce the noise issue. Following is the overview of the algorithm for this function:

1. Estimate a grid size dependent on the greatest element of the image.
2. If a window size isn't indicated picked the grid size as the default window size.
3. Identify grid points on the image, beginning from upper-left corner. Every grid point is isolated by grid size pixels.
4. For every grid point compute the histogram of the region around it, having region equivalent to window size and focused at grid point.
5. If a clipping level is determined clips the histogram estimated above to that level and then uses the new histogram to estimate the cumulative distribution function (CDF).
6. After compute the mappings for every grid point, repeat steps 6 to 8 for every pixel in the input image.
7. For every pixel locate the four nearest neighboring grid points that surround that pixel.
8. Using the intensity value of the pixel as an index, discover its mapping at the four grid points depended on their CDF s.
9. Interpolate among these values to get the mapping at the present pixel area. Map this intensity to the range [min:max) and put it in the output image.

8. Feature Extraction Algorithm

In our technique, we use principal component analysis (PCA) algorithm to extract features from images. A matching procedure is then applied to locate the best match from the data base to authenticate the individual. It changes various conceivably associated factors into fewer new factors, called as principal components [21]. Since a digital image can be viewed as a two (ormore) dimensional function of pixel values and represented as a 2D or 3D data array. More technical details can be found in [22].

<u>The steps of algorithm:</u>

1- Suppose data matrix is B of size m x n. Estimate mean μ_i for each dimension.
2- Subtract the mean from every column to fetch A
3- Calculate covariance matrix C of size n x n which C= A^T A
4- Compute the eigenvalues and eigenvectors (E, V) of the covariance matrix C
5- Project the data step by step onto the principle components$\vec{v_1}$,$\vec{v_2}$,……, etc
6- Select n eigenvectors that compare to the biggest n eigenvalues to be the new basis.

9. Matching Algorithm

In our technique, we use the K-NN classifier. The nearest neighbor classifier works depended on a simple nonparametric decision. Every query image Iq is inspected depended on the distance of its features from the features of other images in the database. The nearest neighbor is the image which has the minimum distance from the query image in the feature space. The distance between two features can be compute depended on one of the distance functions such as, city block distance d_1, and Euclidean distance d_2 or cosine distance d_{cos} [23].

$$d_1(x,y) = \sum_{i=1}^{N} |x_i - y_i| \tag{1}$$

$$d_2(x,y) = \sqrt{\sum_{i=1}^{N} |x_i - y_i|} \tag{2}$$

$$d_{cos}(x,y) = 1 - \frac{\vec{x}.\vec{y}}{|x|.|y|} \tag{3}$$

K nearest neighbor algorithm utilizes K nearest samples to the query image. Every one of these samples belongs to a known class Ci. The query image Iq is arranged to the class CM which has the most of events among the K samples. The presentation of the KNN classifiers highly related to value of the k, the number of the samples and their topological distribution over the feature space.

10. The Proposed Algorithm Preprocessing Finger Vein Image

In our study, we used the SDUMLA-HMT Database [24].This database includes a multimodal data (face, finger vein, iris, finger print and gait) from 106individuals. The SDUMLA-HMT is the first available open access database. The acquisition system is conceived by Joint Lab of Wuhan University for Intelligent Computing and Intelligent Systems. Each subject contains his/her code as index, middle and ring finger of both hands. The set of 6 fingers is repeated 6 times to obtain 6 finger vein images. Accordingly, finger vein database includes 3,816 images. All images are encoded in “bmp” format by a resolution of (320 × 240). At the end, finger vein database takes up around 0.85 GBytes approximately.

The proposed algorithm of ROI extraction and preprocessing of finger vein image includes 6 tasks, as show in figure 2.

1- Convert image to binary.
2- Detect finger edge from the binary image using canny operator.
3- A sub-image is detected and extracted as the ROI of hand vein image.
4- Apply the median filter to reduce the black noise between vein pattern lines
5- Apply the wiener filter to remove effect of high frequency noises.
6- Apply the CLAHE filter to enhance contrast of vein image.

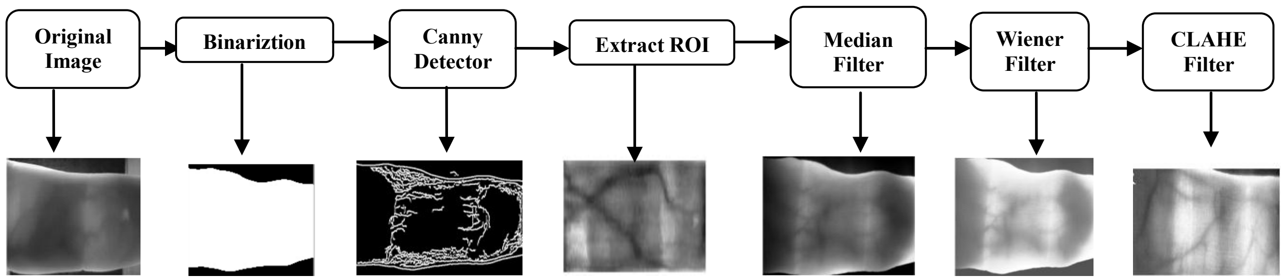


Figure 2 The steps to detect ROI and preprocessing image

11. Results and Discussion

Finger vein recognition includes training and recognition phases. In training phase, features of the training samples are calculated and stored in a database template. In the recognition phase, features of the input vein are determined and then matched by using K-NN matching classifier. After this, these features are compared with the stored template to obtain the recognition result. We do our experiment by divided the database to 5 Cases as table 2 shows:

Table 2 Data base for 5 cases

Case No.	Training	Testing
1	One image for every person (106 images)	Five images for every person (506 images)
2	Two images for every person (206 images)	Four images for every person (406 images)
3	Three images for every person (306 images)	Three images for every person (306 images)
4	Four images for every person(406 images)	Two images for every person (206 images)
5	Five images for every person (506 images)	One image for every person (106 images)

By applying the PCA algorithm with K-NN (Euclidean distance) the result is 100% for all training cases and testing result of every case showed in table 3 and figure 3. We have two potential results, the first result is where the user that is unauthorized which means that his/her template is not found in the database, and the other result is the user is authorized, i.e. a template similar to his/her is found in the database. The experimental results show that the result of recognition Correct Recognition Rate (CRR) is 94.6 %.Hence this method can be successfully used for recognition.

Table 3 The testing results for each case

Case No.	CRR
1	93
2	94
3	95
4	95
5	96

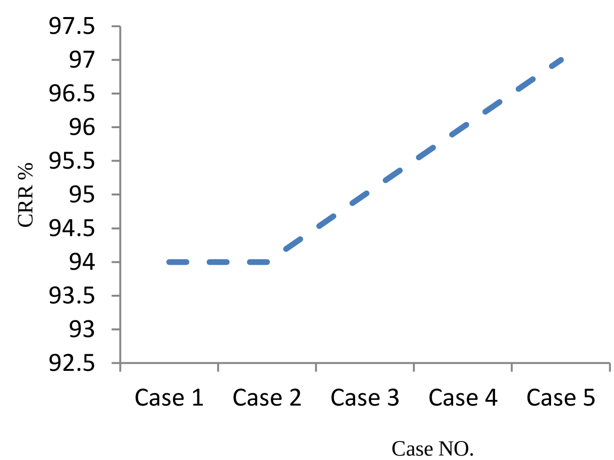


Figure 3 Result of cases

12. Conclusion and Future Work

In this paper, we have developed a new practical and intelligent technique for biometric recognition based on finger vein. The technique consists of the following steps: Image acquisition, determining the region of interest and pre-processing, extracting the finger vein pattern features and recognition. We proposed an original method based on the principal component analysis (PCA) algorithm to extract features and using K-NN (Euclidean distance) matching classifier in matching. The developed technique achieves a very high performance proved by perfectly extract region of interest (ROI) from 633 image from 636finger vein images used in our experiments with 99.5% correct recognition rate. In addition, this smart technique has many advantages and characteristics of flexibility of the former approaches; such as it can overcome the problem of rotation and shift, accurate, simple, practical and fast. In this paper, a complete biometric system based finger veins has been developed. We proposed an original method based on the PCA algorithm to extract features and using KNN classifier in matching. The experimental results show that the result of recognition CRR is 94.60%.Hence this method can be successfully used for recognition. The vein pattern identification can proceeded in a perfect way using the method proposed in this paper which is accurate, simple, practical and fast.

In our opinion, this developed improvement increases the usefulness and usability of this efficient technique, especially as regards its application in all security tasks and domains. Future work may involve applying additional/ alternative pattern recognition algorithms or turning it into a multimodal system where other additional biometrics traits are considered and making the system more invariant to illumination conditions.

REFERENCES

[1] Mubeen Fatima, “Securing the Biometric template: A Survey”, International Journal of Computer Applications, Vol.160, NO. 2, pp.1-8, February 2017.

[2] Sumit Kulkarni , Manali Pandit, “Biometric Recognition System based on Dorsal Hand Veins”, International Journal of Innovative Research in Science, Engineering and Technology(IJIRSET), Vol. 5, Issue 9,pp. 18899-18905, September 2016.

[3] Laxmi Menaria, Kalpana Jain, “ A Survey on Biometric Template Protection”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT),Vol. 2 ,Issue 2,pp.995-999, 2017.

[4] David Mulyono and Horng Shi Jinn,“ A Study of Finger Vein Biometric for Personal Identification”, International Symposium on Biometrics and Security Technologies conferences, April 2008.

[5] T. D. Pham, Y. H. Park, D. T. Nguyen, S. Y. Kwon and K. R. Park,“ Nonintrusive Finger-Vein Recognition System Using NIR Image Sensor and Accuracy Analyses According to Various Factors”, US National Library of Medicine National Institutes of Health, Vol.13,No.7,pp.16866-16894, Jul 2015 .

[6] X. Meng, Xiaoming XI, G. YANG and Y. YIN, “Finger vein recognition based on deformation Information”, SCIENCE CHINA Information Sciences ,Vol. 61, May 2018.

[7] G Ayappan and A Shankar, “Finger Vein biometric Authentication System”, International Journal of Trend in Research and Development, Vol. 4, No.2,pp. 51-56, Apr 2017.

[8] M. M. Ali, .K. Nusrat, A. A. I. Mohamed, M. A. Hagal,H. H. A. Almezogi, J. Rahebi and A. Hançerlioğullari, “Finger Vein Recognition with Gray Level Co-Occurrence Matrix based on Discreet Wavelet Transform”, International Journal of Computer Science And Technology ,Vol. 8, No2, pp.108-112,June 2017.

[9] A. S. BADER and A.M.SAGHEER, “Finger vein identification based on corner Detection”, Journal of Theoretical and Applied Information Technology, Vol.96, No 9,pp.2697-2705, May 2018.

[10] B. Janney.J, S. Divakaran and G.Umashankar, “Finger vein recognition system for authentication of patient data in hospital”,International Journal of Pharma and Bio Sciences ,Vol.8,No.2,pp.(B)-5 -10, Apr 2017.

[11] H.Qin, L.Qin,L. Xue, X.He, C.Yu and X. Liang, US National Library of Medicine National Institutes of Health, Vol.13,No.11,pp.15048-15067, Nov 2013.

[12] R. Das, E. Piciucco, E. Maiorana and P. Campisi, “Convolutional Neural Network for Finger-Vein-based Biometric Identification”, IEEE Transaction for Information Forensics and Security,vol. 14, no. 2, pp. 360-373, Feb. 2019

[13] K. Kaur and S. Bhushan, “Finger Vein Enhancement Using Segmentation and SVM”, International Journal of Computer Science and Information Technologies, Vol. 6, No5, pp.4549-4554, 2015.

[14] Madhusudhan M V, Udayarani V and ChetanaHegde, “Finger Vein Based Authentication using Deep Learning Techniques”, International Journal of Recent Technology and Engineering (IJRTE) ,Vol .8,pp.5403-5408, 2020.

[15] AMRUTA L KABADE , Dr.V.G. Sangam, “Canny edge detection algorithm”, International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 5, Issue 5,pp.1292-1295, May 2016.

[16] Jamil A. M. Saif, “Tuning of Canny Image Edge Detection”, International Journal of Computer Science and Information Security (IJCSIS),Vol. 15, No. 11,pp.184-187, November 2017.

[17] Dr. Ahmad Odat,Dr. Mohammed Otair, and FadiShehadeh, “Image Denoising by Comprehensive Median Filter ”, International Journal of Applied Engineering Research ,Vol. 10, No. 15,pp 36016-36022,2015.

[18] B. Barathi, “Effective Filtering Algorithms for Enhancing Mammogram Images”, International Journal of Advance Research in Computer Science and Management Studies. Vol. 2, No. 3, pp. 287-292, 2014.

[19] Mr. AkashKethwas, Dr. BhavanaJharia, “Comparison Study on Image Denoising Through Wiener Filter”, International Journal of Engineering Research & Technology (IJERT),Vol. 3 Issue 8,pp.962-966, August – 2014

[20] Smitha N, Ujwala B S, Dr. Manjunatha P, Chethan L S,“ Contrast limited Adaptive Histogram Equalization and Discrete Wavelet Transform Method Used for Image Enhancement”, International Journal for Research Trends and Innovation(IJRTI), Vol 2, Issue 8,pp. 142-146,2017

[21] Nathan Halko, Per-Gunnar Martinsson, YoelShkolnisky and Mark Tygert, “ An algorithm for the principal component analysis of large data sets” ,Siam journal on scientific computing ,vol. 33, No. 5 , pp.1-12, October 2011.

[22] AlešHladnik, “Image Compression and Face Recognition: Two Image Processing Applications of Principal Component Analysis”, International Circular of Graphic Education and Research, No. 6, 2013.

[23] Muhammad NaufalMansor,SazaliYaacob, HariharanMuthusamy, ShafrizaNishaBasah,Shahrul Hi-fi Syam bin Ahmad Jamil,MohdLutfiMohdKhidir,MuhammadNazriRejab,KuMohdYusri Ku Ibrahim,Addzrull Hi-fi Syam bin Ahmad Jamil,Jamaluddin Ahmad ,Ahmad KadriJunoh, “ PCA-Based Feature Extraction and k-NN algorithm for Early Jaundice Detection”, International Journal of Soft Computing And Software Engineering (JSCSE),Vol.1,No.1,pp25-29, 2011.

[24] http://www.mla.sdu.edu.cn/sdumla-hmt.html access 2018.

Internet of Things Technology Purposeful For Monitoring Road Traffic Air Pollution

Gordana Jotanovic*, Jelena Stojanov**, Dragan Perakovic***, Zeljko Stojanov**, Goran Jausevac*, Mirko Stojicic*
*University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Herzegovina
**University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
***University of Zagreb, Faculty of Transport and Traffic Sciences, Zagreb, Croatia
gjotanovic@yahoo.com, jelena.stojanov@uns.ac.rs, dperakovic@fpz.unizg.hr, zeljko.stojanov@uns.ac.rs, gjausevac@gmail.com, mirkostojcic1@hotmail.com

Abstract - The paper presents a preliminary study of measuring road traffic noise, air quality, and vehicle frequency by using Internet of Things (IoT) technology in order to determine the need for continuous monitoring of these parameters for urban areas where the population does not exceed 100,000. The study is presented for the town of Doboj, Bosnia and Herzegovina, and includes architecture of the system for monitoring air pollution in road traffic based on the IoT technology. It is used to collect, process, and record the values of road traffic noise, air quality index (PM2.5, PM10, CO2, CO, temperature, and humidity) and traffic flow frequency. In addition, it would inform citizens and relevant services on real-time values of air pollution in the city of Doboj, Bosnia and Herzegovina. The system aims to improve the quality of human life and health in urban areas.

I. Introduction

Fine dust air pollution produced by the discharge of pollutants into the atmosphere causes many side effects [1, 2]. These effects are: climate change caused by global warming, various types of diseases caused by the violation of the human body, and the growing economic loss in every year [3]. Particulates of the dust (PM2.5, PM10) deeply penetrate the respiratory system of the human body. In addition, particulates are the cause of major diseases such as bronchitis, asthma, lung disease, and lung cancer [4, 5].

The heating season and specific weather conditions have led to the fact that we can see and feel the air pollution in our cities. Only then, the alarms go off.

However, one type of pollution is present in cities throughout the year, even when we do not notice it directly, and it relates to all of us who own and drive motor vehicles. Less or more, all citizens who breathe polluted air are victims of such a situation. Pollution caused by burning coal, in thermal power plants or private houses, as well as pollution by road traffic, are the topics of this paper.

Road traffic is not the "number one" polluter in Bosnia and Herzegovina, but it is too present in urban areas to be ignored. Air pollution becomes the topic only when our senses react - the sense of sight and the sense of smell. The reason for this approach is in the perception as a subjective experience of certain external events, its specific character, and difficulties in linking the causes with the effects on human health [6].

According to the International Council on Clean Transport (ICCT), road traffic accounts for about 20% of global air pollution. Also, majority of this pollution is concentrated in urban areas, where the largest number of people lives. Therefore, the "specific weight" of the harmfulness of motor vehicles' exhaust gases is much higher in urban areas. The countries of Western Europe have recognized this problem and are working with combined methods to reduce or completely neutralize it.

According to a 2015 survey of approximately 13,000 premature deaths in Germany can be attributed to ambient PM2.5 and ozone from exhaust pipeline emissions. The deaths attributed to ambient PM 2.5 and ozone from all sources were 43,000, meaning that transportation accounted for just under one-third (31.4%) of all deaths from PM2.5 and ozone pollution that year. Germany had the 4th highest number of premature deaths due to PM2.5 transportation and ozone pollution among all countries in 2015, behind China, India, and the United States. Germany had the highest mortality rate attributable to transport emissions from any major economy, with 17 premature deaths per 100,000 population, which is more than three times the global average and almost 50% above the EU average. Road diesel vehicles provided two-thirds of the health burden in Germany. Among the 100 major urban areas around the world assessed by the study, 3 of the 6 cities with the highest mortality rate attributable to transport emissions in 2015 in Germany. These 6 cities were Milan, Turin, Stuttgart, Kiev, Cologne, and Berlin. The social loss of social welfare associated with transport health damage in Germany amounted to 110 billion (US $ 2015) (approximately EUR 97 billion) in 2015, which is about 3% of Germany's gross national income [7, 8, 9, 10].

The precondition for solving such a complex problem is a lot of money, but this does not relieve Bosnia and Herzegovina, as poorly society, of responsibility for passivity in solving the problem. Money insufficiency is an aggravating factor, but it seems that even obstacles are the ignorance of the population, the lack of will to start solving the problem, neglecting the systemic solution and its strict and consistent application.

The paper is structured as follows: Section II presents IoT architecture for purposeful monitoring of road traffic air pollution. Section III describes the data collection process and methodology. Results and discussion are outlined in the section IV, while the conclusion is presented in the section V.

II. IoT Architecture Purposeful For Monitoring Road Traffic Air Pollution

Based on the IoT architecture for monitoring road traffic, large amounts of data can be collected and processed in real-time. Mobile communication has recently been developed very fast. The IoT sensor calculates the pollutant concentration and sent the collected data to the IoT cloud. IoT rapidly grows by connecting variety of devices and providing data on their interaction [11].

The technical system for monitoring air pollution of road traffic is based on the layered architecture that is shown in Figure 1. The architecture of the IoT system consists of four layers: Perception, Network, Middleware, and Application [12].

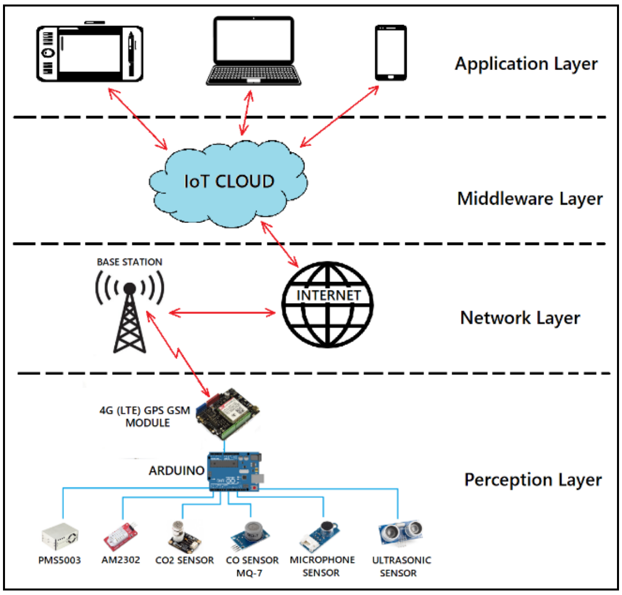


Figure 1. The architecture of the IoT system for monitoring air pollution by road traffic [12]

The Perception layer collects data via sensors. The collected the values relate to road traffic noise, air quality index (PM2.5, PM10, CO2, CO, temperature, and humidity) and traffic flow frequency. Communication equipment that enables data transfer to an IoT node is located in the Network layer [13]. Also, the Perception layer contains a hub device that has the task of collecting data from the sensors and enables their mutual communication as well as sending data to the Network layer.

Mobile 4G, hopefully soon applicable 5G and similar networks enable communication between the Perception layer and the Network layer. Also, the previously mentioned mobile networks enabled data transfer to the Internet.

The Network layer is in the charge of communication within the IoT and is the backbone of the system. It is responsible for communication and data transfer, and secure transfer of information from the Perception layer to the Application layer and reverse. Communication protocols are in this layer. The architecture of the IoT system for monitoring air pollution of road traffic, designed for the urban environments, can be seen in Figure 1. Therefore, the IoT environment for air pollution in road traffic commonly uses long-range wireless technologies such as Lora and LoRaWAN [14].

The Middleware layer covers data storage, data analysis, and data processing. The Middleware layer consists of storage devices and data analysis tools. Also, used as a service for extended storage and related applications, it includes user reporting via mobile and web applications. This layer uses a variety of technologies such as database software, cloud computing, intelligent processing, etc. Also, this layer is responsible for service management. The Middleware layer is based on the concept of cloud computing and contains resources and elements for data processing [15, 16].

The Application layer has the task of integrating the collected and processed information and delivering it together with the applications to users via mobile and web applications. This layer is responsible for the global management of applications implemented by IoT. Safety in such environments must cover all horizontally and vertically oriented layers [16].

III. Material and Methods

This paper presents the results of the study of traffic noise levels, Air Quality Index, and traffic flow frequency in the area of Doboj, Bosnia and Herzegovina. The research deals with the collection, statistical processing, and recording of the pollution values in traffic flow frequency (N), road traffic noise values expressed in (dB), and PM2.5 and PM10 (μg/m3). The research identify the need for continuous monitoring of air pollution, noise, and vehicle frequency in cities with a population of less than 100,000, which includes Doboj in the Republic of Srpska.

A. Measuring Road Traffic Air Pollution

Measurements of air pollution (PM2.5, PM10, road traffic noise, and traffic flow frequency) are performed during 16 days in 2020 in the winter and summer seasons in the city of Doboj, Bosnia and Herzegovina. Measurements consist of 4 phases with four days at two locations marked 1-yellow and 2-yellow, as it can be seen in Figure 2. The first two measurement phases are done in winter and the other two phases in summer. The first two phases of measurements were performed during the operation of the Heating Plant, which is one of the largest air pollutants in the city of Doboj. The other two phases of measurements were performed in the period when the Heating Plant is not active.

The focus of the observation are the locations marked in yellow (1, 2, and 3) in Figure 2.

Figure 2. Key observation points: (1- yellow) Main road M17, (2-yellow) Shopping center "Tropic" and (3-yellow) City heating plant

The location marked with 1-yellow represents the intersection of the Main road M17 from the direction of the settlement of Modrica, where the traffic flow frequency and road traffic noise were measured. 1-yellow represents the location where the highest frequency of traffic is on the territory of the city of Doboj. The observation location (1 yellow) is chosen due to traffic jams in the morning, at the end of the working day, and at night. For this reason, the police often regulate traffic at the designated location.

The measurement of the traffic flow frequency (NoV) was performed at certain intervals and represents the number of vehicles in them. Sixteen measurements were performed in the morning, afternoon, and evening (traffic peaks).

Road traffic noise was measured at the same time intervals as the traffic flow frequency. The measuring point of the road traffic noise is located 1 m from the Main road M17, at a height of 1.5 m [17, 18, 19]. Traffic noise was measured using the Arduino module and microphone sensor. Noise level sensor connects via an analog interface to a microcontroller [20].

The location marked with 2-yellow is a measuring point for placing sensors near the Shopping center "Tropic" in the town of Doboj. The measuring point is located approximately halfway between the Main road M and the City Heating Plant. Also, it should be pointed out that the measuring point is located in the center of the city of Doboj near the Health Center. Measurement data were collected using a PurpleAir sensor for PM2.5 and PM10 pollutant air particles. PurpleAir sensors measure airborne particulate matter (PM). Particulate matter describes solid particles suspended in air. This includes dust, smoke, and other organic and inorganic particles. PurpleAir sensors use laser particle counters to count the number of particles of 2,5 and 10 µm sizes, and use the count data to calculate mass concentrations of PM2.5 and PM10. PurpleAir sensors use laser counters to measure particulate matter in real-time. A laser counter uses a fan to draw a sample of air past a laser beam. Any particles in the air will reflect some light from the laser beam onto a detection plate, like dust shimmering in a sunbeam. The reflection is measured as a pulse by the detection plate, and the length of the pulse determines the size of the particle, while the number of pulses determines the particle count. The PM2.5 and PM10 mass concentrations are calculated from these particle counts.

The marked location 3-yellow is not a measuring point. However, the Doboj heating plant is located at the mentioned location. The heating plant is one of the biggest polluters. In addition to air pollution caused by road traffic, it is also important for the analysis of air quality in the city of Doboj.

B. Data descriptions

Collected data contain information on daily base, for the following air pollutants:

1. Concentrations of PM particles of two types, PM2.5 and PM10, expressed in µg/m3 – data are overtaken from PurpleAir sensors [21].
2. Traffic intensity - the number of vehicles within traffic peaks, during two fifteen-minutes periods in the morning (7:00-7:15 am., 7:25-7:40 am), twice in the noon (2:45-3:00 pm, 3:10-3:25 pm) and twice in the evening (10:00-10:15 pm, 10:25-10:40 pm).
3. Road traffic noise (dB) was measured at the same periods as the traffic flow frequency.

TABLE I. AVERAGED NUMBER OF VEHICLES AND NOISE LEVELS

season	day	PM2.5	PM10	M NoV	M noise (dB)	N NoV	N noise (dB)	E NoV	E noise (dB)
Winter	1	208	189,47	176,5	59	221,5	60,5	49,5	52
	2	196	283,73	72	54,5	209	62,5	60,5	46,5
	3	363	267,81	107,5	44,5	160	47,5	31,5	42,5
	4	267	258,9	100	51,5	159	59	31,5	55
	5	102	39,71	115	52,5	171,5	54,5	96,5	52
	6	152	68,27	207,5	68,5	255	66	72	61,5
	7	171	107,56	294,5	61,5	109,5	55,5	88,5	56,5
	8	173	113,35	177,5	65,5	310	63,5	63,5	60
Summer	1	60	16,57	182	52,5	292	56,5	52	52,5
	2	69	23,2	377	48,5	348	67	85	53
	3	78	28,92	377	41	348	65,5	147,5	62,5
	4	87	33,24	231,5	63,5	302	29	82,5	53
	5	89	34,98	167,5	51,5	238,5	60,5	160	63
	6	60	18,2	305	48,5	248,5	60,5	237,5	30,5
	7	58	17,67	305	67,5	248,5	63	156,5	63
	8	82	32,01	229	61,5	282	64	106,5	41

Data were collected during eight winter days and eight summer days. Average number of vehicles and noise levels are taken for morning, noon and evening. Sorted data are presented in Table 1.

The labels used in the table are: M NoV - Morning average number of vehicles within 15min, M noise - Morning noise, N NoV - Noon average number of vehicles within 15min, N noise - Noon noise, E NoV - Evening average number of vehicles within 15min, E noise - Evening noise.

Since the sample is small (N=16) and data highly deviate, the usual descriptive statistic is pointless.

Two following scatter plots on traffic intensity and traffic noise are separated by the season and confirm common daily activity (see Figures 3 and 4). Deviations are greater in summer, which is explained by increased societal activities and therefore by alternative noise sources. Day period (morning, noon, and evening) has an obvious impact on the monitored variables, especially on the traffic density (number of vehicles). Hence averaging data on a daily level would mean a loss of information.

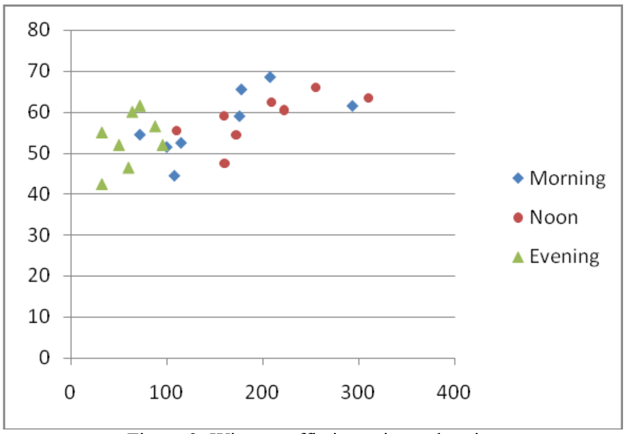


Figure 3. Winter traffic intensity and noise

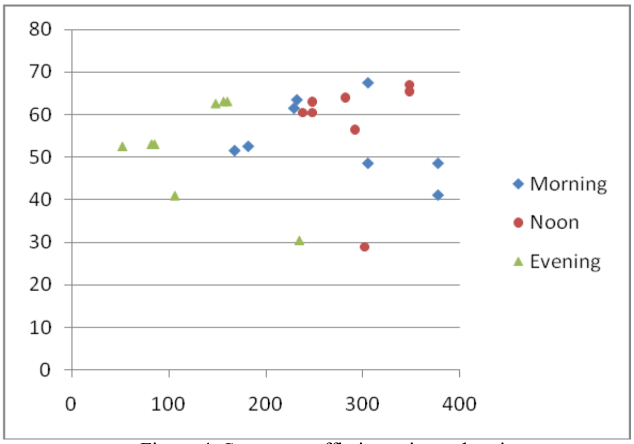


Figure 4. Summer traffic intensity and noise

IV. RESULTS AND DISCUSSION

In general, deviated data have low correlation (variations of one variable need not to be followed by the related variations of the other variable). The correlation coefficient is a measure of the relative consistency of variations of two variables. Sample correlations among the four variables (traffic intensity, noise, concentrations of PM2.5 and PM10) are determined and presented according to day periods (see Tables II, III and IV).

TABLE II. MORNING SAMPLE CORRELATION, IN WINTER AND SUMMER SEASON

sample correlation, morning				
w \ s	NoV	Noise	PM2.5	PM10
NoV		-0,39	-0,31	-0,27
Noise	**0,68**		0,01	0,03
PM2.5	-0,35	**-0,64**		**0,99**
PM10	-0,58	-0,61	**0,79**	

TABLE III. NOON SAMPLE CORRELATION, IN WINTER AND SUMMER SEASON

sample correlation, noon				
w \ s	NoV	Noise	PM2.5	PM10
NoV		0,03	0,12	0,1
Noise	**0,7**		-0,37	-0,31
PM2.5	-0,27	-0,53		**0,99**
PM10	-0,2	-0,22	**0,79**	

TABLE IV. EVENING SAMPLE CORRELATIONS, IN WINTER AND SUMMER SEASON

sample correlation, evening				
w \ s	NoV	Noise	PM2.5	PM10
NoV		-0,29	-0,17	-0,12
Noise	0,4		0,24	0,23
PM2.5	**-0,88**	-0,6		**0,99**
PM10	**-0,83**	**-0,67**	**0,79**	

Emphasized coefficients in the previous tables confirm correlations existence, but having in mind features of the sample, they are not reliably.

Achieved data analysis verifies that the research results are promising, but there is a need for a significantly greater sample. In addition, taking into consideration more factors could improve reliability of results and conclusions. Particularly, temperature and humidity of the air, and wind type and speed impact noise and concentrations of pollution particles.

V. CONCLUSION

This paper aims to investigate the needs of cities with less than 100,000 inhabitants, in particular the city of Doboj, for monitoring road traffic air pollution. For that purpose, an IoT architecture for monitoring of road traffic air pollution is proposed. The mentioned IoT architecture monitors noise levels and other parameters that affect the air quality near roads in urban areas and informs citizens, competent inspections, and services about values that are dangerous to human life.

A parameter that we cannot ignore relates to city heating plants, factories, and other major air pollutants that also affect air quality. In the preliminary research presented in the paper, air pollution from the heating plant in the city of Doboj was taken into account. The collected data were analyzed in the summer and winter seasons, ie. in the season of operation and downtime of the city heating plant, and at different times of the day (morning, day, night). During the period of dormancy of the city heating plant as the biggest polluter on the territory of the city of Doboj, we can refer to the emission of traffic pollution to a greater extent. However, we cannot fully adopt this statement because, in addition to pollutants in the city of Doboj, air pollution can come from other settlements. To look at the problem in its entirety, we would have to take into account meteorological parameters such as humidity, winds, temperature, etc. Therefore, we accept the data collected during the standstill period of the heating plant with the reservation that they are exclusively caused by traffic pollution.

Air pollution emitted by the city heating plant and other major polluters does not tend to decrease depending on the number of users. Therefore, the air pollution that we can influence is traffic air pollution.

One of the goals of this work is to use the proposed IoT architecture to enable a strategy for the continuous monitoring and partial elimination of traffic noise and air pollution in urban areas produced by mobile sources of traffic pollution (cars, trucks, trains, planes, etc.), aimed at protecting human health.

The offered solutions for solving the problem of air pollution can be the redirection of road traffic, and analysis of the possibility of introducing electric vehicles, bicycles in urban areas. The long-term strategic goal should be to create transport that is independent or to a less extent dependent on fossil fuel. The introduction of electric vehicles would have an impact on environmental protection, reduction of exhaust emissions, lower noise levels, etc. which is especially pronounced in densely populated places where the use of such vehicles is recommended. Also, the use of public transport, taxis, electric trains, would affect the lower level of pollution, cheaper transport, and a great reduction in costs. And finally, this research encourages more active work of the competent inspections responsible for the preservation of the human environment.

ACKNOWLEDGMENT

Ministry for scientific and technological development, higher education and information society of the Republic of Srpska supports this research under the project "Smart system based on IoT technology designed for monitoring of traffic air pollution", contract number 19.030/3-2-25-2/1.

REFERENCES

[1] J. Xun. et al. "A study of aerosol changing rule in Beijing during a decade". Proceeding of 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4115-4118. 2016.

[2] Y. Han, B. Park, and J. Jeong. "A Novel Architecture of Air Pollution Measurement Platform Using 5G and Blockchain for Industrial IoT Applications". Procedia Computer Science, vol. 155, pp. 728-733. 2019.

[3] S. Kim. et al. "Development of an IoT-based atmospheric environment monitoring system." Proceedings of 2017 International Conference on Information and Communication Technology Convergence (ICTC), pp.861-863. 2017.

[4] M. C. Turner. et al. "Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of neversmokers." American journal of respiratory and critical care medicine, vol. 184, no.12, pp. 1374-1381. 2011.

[5] A. C. M. Fong. et al. "An automated solution for disease control and management to combat the health impact of indoor air pollution.". Proceedings of IEEE Conference on Industrial Electronics and Applications, pp. 472-476. 2012.

[6] B.Đurić, A. Trifunović, S. Čičević, A. Josić, and I. Stokić. "Merenje emisije buke motornih vozila na teritoriji grada kragujevca". Put i saobraćaj, vol. 64, no. 1, pp. 61-66. 2018.

[7] S. Anenberg, J. Miller, D. Henze, and R. Minjares. A global snapshot of the air pollution-related health impacts of transportation sector emissions in 2010 and 2015. International Council on Clean Transportation: Washington, DC, USA. 2019.

[8] G. P. Chossière, R. Malina, A. Ashok, I. C. Dedoussi, S. D. Eastham, R. L. Speth, and S. R. Barrett. "Public health impacts of excess NOx emissions from Volkswagen diesel passenger vehicles in Germany". Environmental Research Letters, vol. 12, no. 3, 034014, 2017.

[9] M. Krzyżanowski, B. Kuna-Dibbert, and J. Schneider (Eds.). Health effects of transport-related air pollution. WHO Regional Office Europe. 2005.

[10] K. Rehdanz, and D. Maddison. "Local environmental quality and life-satisfaction in Germany". Ecological economics, vol. 64, no. 4, pp. 787-797. 2008.

[11] G. Kortuem. et al. "Smart objects as building blocks for the Internet of things". IEEE Internet Computing, vol. 14, no. 1, pp. 44-51. 2010.

[12] I. Cvitic, M. Vujic, and S. Husnjak. "Classification of Security Risks in the IoT Environment". Annals of DAAAM and Proceedings of the International DAAAM Symposium, pp. 0731–0740, 2016.

[13] I. Cvitić, D. Peraković, M. Periša, and M. Botica. "Novel approach for detection of IoT generated DDoS traffic". Wireless Networks, vol. 25, pp. 1-14. 2019.

[14] D. Dobrilović, M. Malić, D. Malić, and S. Sladojević. "Analyses and optimization of Lee propagation model for LoRa 868 MHz network deployments in urban areas". Journal of Engineering Management and Competitiveness (JEMC), vol. 7, no. 1, pp. 55-62, 2017.

[15] J. Park, H. Kwon, and N. Kang. "IoT–Cloud collaboration to establish a secure connection for lightweight devices". Wireless Networks, vol. 23, no. 3, pp 681–692, 2017.

[16] R. Kopp, J. Howaldt, and J. Schultze. "Why Industry 4 0 needs Workplace Innovation: a critical look at the German debate on advanced manufacturing". European Journal of Workplace Innovation, vol. 2, no. 1, pp. 7-24, 2016.

[17] J. Božić, P. Ilić, and L. S. Bjelić. ECONOMIC ASPECTS OF THE CITY TRAFFIC NOISE: CASE STUDY//EKONOMSKI ASPEKTI BUKE OD GRADSKOG SAOBRAĆAJA: STUDIJA SLUČAJA. EMC REVIEW-ČASOPIS ZA EKONOMIJU, vol. 8, no. 1, pp. 134-149. 2018.

[18] European Environment Agency. Noise in Europe. EEA Report No. 10/2014. Luxembourg: Publications Office of the European Union. 2014.

[19] A. Hassan, and J. B. Alam. Traffic Noise Levels at Different Locations in Dhaka City and Noise Modelling for Construction Equipments. IJERA, vol. 2, no. 3, pp. 1032-1040. 2013.

[20] https://www.banggood.com/Microphone-Noise-Decibel-Sound-Sensor-Measurement-Module-3p-4p-Interface-p-1553625.html?cur_warehouse=CN, last accessed 2020/27/09.

[21] https://www.purpleair.com/map?opt=1/mAQI/a10/cC0#4.5/40.38/14.4, last accessed 2020/29/09.

Web-Based Solution for Student Attendance Tracking: Development and Evaluation

Z. Kotevski, N. B. Tabakovska and T. Dimovski
"St. Kliment Ohridski" University – Bitola / Faculty of Information and Communication Technologies
Bitola, Macedonia
zoran.kotevski@uklo.edu.mk, natasa.tabakovska@uklo.edu.mk, tome.dimovski@uklo.edu.mk

Abstract - In present times, student attendance tracking systems are primarily based on contemporary identification technologies. These technologies comprise RFID, Bluetooth, Barcode and several biometric such as Fingerprint, Voice and Face recognition. Some solutions employ web technologies only, and offer attendance tracking services over the Internet. All the different types of attendance tracking systems are built on IP infrastructures, but exhibit different performance that make them more or less suitable for use in classrooms. This paper, at first, presents a review of existing systems and technologies for attendance tracking, identifying certain issues concerning their performance. For the most part, it appears that biometric systems exhibit insufficient accuracy, while RFID systems inquire higher implementation and maintenance costs. Barcode systems generally suffer from longer registration durations and Bluetooth systems are affected by distance. The focus of the research is put on the development and performance evaluation of a custom student attendance tracking solution, built on current, free and open source web technologies. The goal was to make an effort to surpass some disadvantages identified in other systems. The developed solution, entitled RENAP, is implemented at a University in Macedonia, whose performance assessment revealed that it excels at several aspects, including accuracy, swiftness, affordability, easiness of usage and flexibility.

I. Introduction

Managing attendance records can present quite difficult task for many organizations, and the maintenance of an appropriate attendance tracking imposes as an important factor, especially for educational institutions. The research of Newman-Ford, Fitzgibbon, Lloyd and Thomas [1] presents evidence of a significant correlation between students' attendance and their academic performance. Similar claims by Marr and Lancaster [2] suggest that the students who have poor attendance records generally link to poor retention. Furthermore, attendance tracking appears to be quite valuable to teachers who obtain important information about student's behavior, as well as to the managers as a support to their decision making.

Manual acquisition of attendance records and their long term maintenance became outmoded when many novel identification (ID) technologies have found applications in attendance tracking systems. In present times, the most commonly used ID technologies include Radio Frequency ID (RFID), Bluetooth, Barcode ID, and several biometric, such as Fingerprint, Voice and Face recognition. Some attendance tracking solutions are based on web technologies only and offered as services over the Internet. Regardless of the fact that contemporary attendance tracking systems share a common concept, i.e. the IP infrastructure, due to the ID technologies employed they exhibit quite different performance. For this reason, the first part of this research aims to determine the standard issues concerning their performance, regarding the implemented ID technology. The second elaborates the development of a web-based student attendance tracking system, entitled as RENAP. The main goal behind the effort to develop a new system was to surpass some performance issues exhibited by other technologies, and to provide custom solution according to the specific requirements of an academic establishment. As a final part, the presented evaluation provides useful insights into the RENAP's performance, and confirms that some of the intended goals are attained.

The rest of this paper is organized as follows. Section two presents the related work and includes a discussion about the performance concerns and drawbacks of the various types of existing attendance tracking systems. The advantages and disadvantages are summarized in a table of performance that are further used as an input for the design of the proposed attendance tracking solution. Section three describes the developed attendance tracking system and the technologies employed. Furthermore, this section presents the results of the RENAP's performance evaluation, provided by the analysis of a survey results. Section four gives the concluding remarks, with a brief elaboration of the system's future requirements and improvements.

II. Related Work

The convenience of implementation of RFID technology has led to the present situation where the majority of modern attendance tracking systems are RFID based (Arulogun, Olatunbosun, Fakolujo & Olaniyi [3]; Chand, Meeran & Prabaharan[4]; Chiagozie & Nwaji [5]; Kassim, Mazlan, Zaini & Salleh [6]; Mohammed & Kameswari [7]; Patel, Patel & Gajjar [8]; Silva Filipe & Pereira [9]; Yuru, Delong & Liping [10]; Singhal & Gujral [11]; Saparkhojayev & Guvercin [12]; Tiwari, Tiwari, Ade, Sheikh, Patel & Khan [13]; Wahab, Kadir, Mutalib & Mohsin [14]; Yadav & Nainan [15]). Systems of this type impose certain expenses for equipment, comprising RFID tags for attendees and RFID readers placed in each classroom. The attendance tracking process is usually managed by a software running on an application server that receives events, i.e. information about the tag ID, date, time, classroom location and other data regarding the class settings. Some efforts have combined RFID with other technologies, such as General Packet Radio Service (GPRS) and Web technologies (Tiwari et al. [13]), or Global System for Mobile Communications (GSM) (Yadav & Nainan [15]), in order to augment the system's capabilities. To alleviate the implementation expenses, Kassim et al. [6] and Mohammed and Kameswari [7] have augmented their RFID based student attendance tracking systems with web technologies (IIS, ASP.NET and Microsoft SQL or MySQL server) and have provided the teachers' with real-time access to the required reports. Some efforts even combine RFID with some form of face detection and recognition techniques to double check the number of registered students with the number of students present in the classroom (Saparkhojayev & Guvercin [12]). The experiences of usage are quite common among the different implementations, where the main conclusions are that RFID based attendance tracking systems work effectively and accurately and provide a convenient method for attendance tracking management.

Bhalla, Singla, Gahlot and Gupta [16] have proposed a Bluetooth based attendance tracking system that uses a software application installed on the instructor's mobile telephone, which enables querying students' mobile telephones via Bluetooth connections. The system is reported to be of low cost, low power and considered as a sufficiently robust solution. The main constraints include limitations of Bluetooth technology regarding the operational distance and, for this particular implementation, the restriction of students' registrations by the MAC addresses of their phones requires constant updates because the students are prone to changing their mobile devices quite frequently.

Subramaniam, Hassan and Widyarto [17] have proposed and developed a student attendance tracking system (SAS) based on Barcode ID. The authors report high user satisfaction after the initial implementation, compared to the traditional attendance tracking on paper. Apparently there has been some issues that required consideration in the next revision, but since it was in a construction phase the behavior has been considered as expected.

Some authors have proposed somewhat different approach for student attendance tracking by using biometric technologies. Joshi and Joshi [18], Nawaz, Pervaiz, Korrani and Azhar-Ud-Din [19], Potadar, Marathe, Khose and Kotkar [20], Shafi, Khan, Munir and Baloch [21], Shoewu and Idowu [22], Talaviya, Ramteke and Shete [23], and Verma and Gupta [24] have incorporated fingerprint recognition technology and, where, after a match is determined, the user (student) data is written in a database. For example, in the research of Shoewu and Idowu [22] the data transfer is realized via wired LAN network and written in a Miscrosoft SQL database, while in other systems the data transfer is realized wirelessly using ZigBee module (Joshi & Joshi [18]; Potadar et al. [20]; Talaviya et al. [23]). The main conclusion, as a drawback from these types of attendance tracking systems, confirms that the systems' accuracy is a bit impaired ranging between 94 % and 98.57 %, and the time required for each user to register is about 4 sec. Combinations with other technologies are also common. Integration of GSM modules for delivering reports to certain stakeholders (Potadar et al. [20]; Joshi & Joshi [18]; Verma & Gupta [24]) are also common. Some systems include Short Message Service (SMS) services for provision of various types of information to the concerned parents (Shafi et al. [21]). These systems are generally regarded as reliable, secure, fast and efficient.

Uddin, Rashid and Mostafa [25] and Reda, Panjwani and Cutrell [26] presented an attendance tracking solutions that are based on voice recognition technology, where the system by Reda et al. [26] is additionally supported by GPS location tagging. Uddin et al. [25] state that the percentage of accurate recognition when students pronounce their names is only 46,15%, which represents a serious performance constraint. Another issue of this type of systems is that students' registration is a serial process (sometimes a single student is required to pronounce its name more than once when misinterpretations occur) which leads to unwanted long durations of attendance registrations.

Concerning the face recognition technology, it is also implemented in many attendance tracking systems (Kar, Debbarma, Saha & Pal [27]; Patil & Shukla [28]; Patil, Karhe & Jain [29]; Tamimi, AL-Allaf & Alia [30]). Some usage reports show that such systems provide automatic attendance recording with zero wasted time, but, Kar et al. [27] report that their system's accuracy in detecting and recognizing faces is around 95 % for frontal face orientation, which drops to zero when the angle of the face toward the camera is 72 degrees. The system created by Patil et al. [29] is additionally equipped with GSM module and is claimed to be highly accurate, while the accuracy of the system of Tamimi et al. [30] is reported to be 94.75%.

Using the aforementioned technologies, many commercial systems are developed and offered as web services for a certain subscription price ("AccuClass" [31]; "Attendance Tracking for Colleges and Universities" [32]; "Register Attendees for Events" [33]; "Student Attendance Management System" [34]; "Student Attendance Tracking" [35]).

The system in "Attendance Tracking with Browsers" [36] represents a genuine web-based solution where each student registers her/his attendance using personal mobile device. The system is publicly available, and any institution that requires attendance tracking services can open an account. However, there are two negative sides of such paradigm. First, the system is dependent on Internet connection that may result in blocked registration when Internet connection is not in operation of various reasons. There are also possibilities for students' registration when they are not being present in the classroom, even though the system offers certain securities using GPS location. Second, the customizability of such system is not on an adequate level, since different institutions in different locations over the world have quite different requirements.

The system developed by Akhila, Prathyusha, PavanKumar and Amrutha [37] presents an IP networked solution that employs web technologies and databases for storing the attendance information. Its management interface is developed for the Android OS and installed on the teacher's mobile device. Similar systems are presented in ("Attendance Tracker" [38]; "BePunctual" [39]; "MyAttendanceTracker" [40]). These systems seem to be quite plausible for use in classrooms and appear not to be affected by the issues exhibited with the implementation of certain ID technology.

Thus, it is important to emphasize that all the aforementioned attendance tracking solutions share a common communication foundation, i.e. the IP protocol. But, different implementations regarding the ID technology contribute to different performance of the variety of attendance tracking systems. Table 1 summarizes the primary advantages and disadvantages of the diverse systems.

TABLE I. ADVANTAGES AND DISADVANTAGES OF DIFFERENT IDENTIFICATION TECHNOLOGIES

Implemented Identification Technology	Primary Advantages	Primary Disadvantages
RFID	Accuracy, Speed, Reliability	Implementation and maintenance costs
Bluetooth	Accuracy, Reliability	Operating distance
Barcode	Accuracy, Reliability	Speed, Implementation costs
Fingerprint recognition	Accuracy, Reliability	Speed, Implementation costs (Storing of personal sensitive data)
Voice recognition	Simplicity of usage	Accuracy, Speed, Reliability, Operating distance
Face Recognition	Speed	Accuracy, Reliability
Subscription Web-based solutions	Accuracy	Reliability, Subscription costs, Speed
Web technology	Customizability, Accuracy, Implementation and maintenance costs	Speed

III. WEB-BASED SOLUTION FOR STUDENT ATTENDANCE TRACKING – EVALUATION AND DISCUSSION

Taking into account the advantages and disadvantages of identification technologies presented in Table 1, we have developed a solution for tracking of student attendance to classes using free and open-source web technologies, such as HTML5, CSS3, JavaScript, PHP, MySQL. It is deployed since January 2016 in a local server, thus the stakeholders can access it regardless of the occasional problems with the Internet connection. The solution follows standard system of roles (administrators, teachers, staff and students), while special attention was given to the speed of attendance registration, in order to overcome the disadvantage. In this manner, prior to a lecture, the teacher needs to log-in and start the attendance registration for the particular lecture. Afterwards, the students are enabled to register their attendance using their personal mobile device, laptop or desktop, via a web interface. The system is called RENAP, which is an acronym for register of classes and attendance in Macedonian language.

In order to evaluate the proposed system for student attendance tracking, a survey was conducted that addressed the users' experiences of all educational system stakeholders, i.e. teachers, students and faculty managers. The research instrument consisted of several questionnaires, which were distributed and answered electronically by 84% of the statistical population. According to the analyzed and interpreted collected data, as presented in Fig. 1, RENAP is very easy for student's use (80.4%) and teacher's use (93.3%). The registration process takes a very short time according to 62.5% of the students and 86% of the teachers, and does not disturb classroom activities (73.2-students, 93%-teachers). 60% of the teachers and 59.8% of the student respondents consider that RENAP offers sufficient data safety. On the question whether they frequently noticed security breaches we received affirmative answers from 6.7% of the teachers and 5.4% of the students, while on the question about occasionally noticed security breaches, even 53.4% of teachers and only 12.5% of the students confirmed. The main negative remark on system's security is the possibility of student attendance registration when a student is not present in the classroom. This can be either performed by a student outside the classroom, or a student that has access to more than a single registration device. Despite of detected opportunities for abuse, only a small percentage of students (9.5%) reported that they abused the system and registered a colleague who was not present in a certain lecture. However, 60% of the teachers and only 31.2% of students consider that RENAP system offers a reliable and adequate security level.

In terms of registration and sign-in to the system, respondents generally expressed high levels of satisfaction. 53.3% of the teachers and 48.3% of the student reported that very rarely had any problems with registration, while only 13.4% of the teachers and 20.6% of the students reported frequent registration issues. The most common problem encountered by students was the inability to connect to the Wi-Fi network (93.3%), while 75% of the teachers reported the same issue. Further analyses revealed that the noted registration problems were associated when students in higher densities attempted to access the system at the same time, and/or when the lectures/exercises were held in a particular classroom where the Wi-Fi signal was weak.

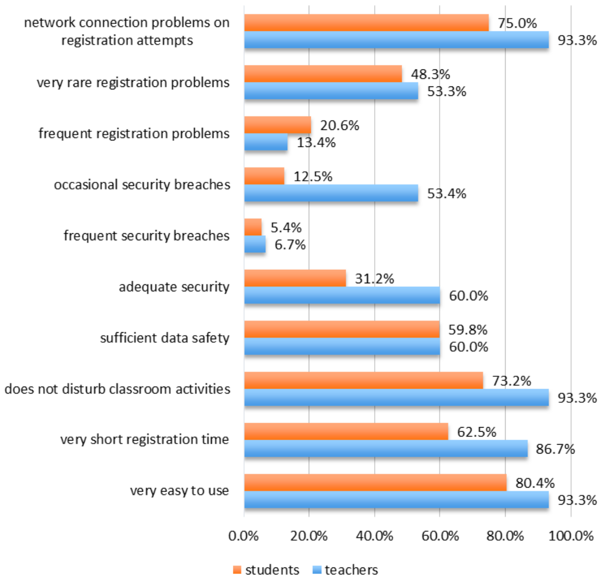


Figure 1 - RENAP's performance evaluated by teachers and students

Regarding the ideas for future addition of other functionalities, given in Fig. 2, students and teachers have suggested several improvements to the system, such as to improve the user interface (62.5%-students; 33.3%-teachers); to incorporate the log-in form for students and enable greater manipulation of personal data in the system (67%-students); to integrate with the university system (58.9%- students; 80%- teachers) and to strengthen the system's security (67%-students; 46.7%- teachers). Other noticeable frequent suggestions were to provide applications for various Android and iOS mobile platforms and many suggestions to improve the network infrastructure.

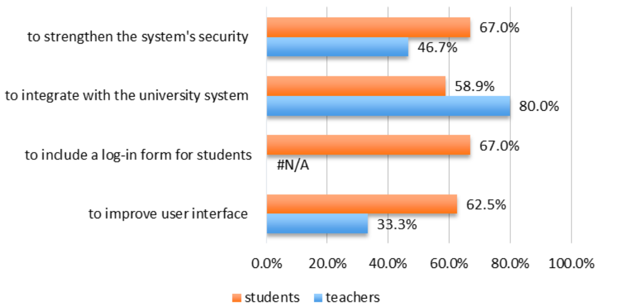


Figure 2 - Suggestions for improvement of RENAP's functionalities

As apparent from the presented evaluation results, the students and teachers appeared quite satisfied of the RENAP system, besides their suggestions for system's improvements. If we compare the results obtained, we can infer that teachers are a bit more satisfied compared to the students, but have almost the same suggestions for improvements. Concerning the satisfaction of the faculty managers, an interview revealed that they are quite satisfied with the system's functionalities as well, and find it very useful in the referred academic environment. The provided possibility to review all the teaching activities is regarded as important and essential for an educational establishment, and that it provides opportunities for better operational and strategic planning, as well as, a substantial support for making decisions.

IV. Conclusion and Future Work

Retaining a track of attendance records is an important aspect for the management of an educational institution. Nowadays, in order to improve and automate attendance tracking processes, many current technologies are employed, among which the most commonly used include: RFID, Bluetooth, Barcode ID, and several biometric technologies, such as Fingerprint recognition, Face recognition and Voice recognition. Yet, there are some that are purely based on web technologies and offered as Internet services. All the aforementioned concepts for student attendance tracking are built on an IP infrastructure, but exhibit different performance that make them more or less suitable for use in classrooms for student tracking purposes, according to the required system's features and functionalities. The analysis of their distinctive properties was an inspiration to develop a custom web-based solution for student attendance tracking that employs the latest and free web technologies. The developed system, titled RENAP, is designed to take into account all the pros of the existing systems and adhere to the defined criteria. RENAP is characterized with short registration times, and uses a simplicity approach where only few data fields need to be filled for attendance registration. It also prevents unintentional erroneous input and provides substantial level of security. The survey conveyed among the interested parties revealed substantial user satisfaction, and several minor deficiencies. The suggestions for future improvements mainly include the requirements from students to enable student's log-in form and provide native applications for the most popular mobile platforms.

References

[1] Newman-Ford, L., Fitzgibbon, K., Lloyd, S. & Thomas, S. (2008). A large‐scale investigation into the relationship between attendance and attainment: a study using an innovative, electronic attendance monitoring system. Studies in Higher Education, 33(6), 699-717.

[2] Marr, L., & Lancaster, G. (2005). Attendance System. Learning and Teaching in Action, 4(1), 21-26.

[3] Arulogun, O.T., Olatunbosun, A., Fakolujo, O.A. & Olaniyi, O.M. (2013). RFID-based students attendance management system. International Journal of Scientific & Engineering Research, 4(2), 1-9.

[4] Chand, M.V.V., Meeran, M.S. & Prabaharan, V. (2015). Secured Attendance Management System Using RFID Technology. International Research Journal of Engineering and Technology, 2(8), 512-516.

[5] Chiagozie, O.G. & Nwaji, O.G. (2012). Radio frequency identification (RFID) based attendance system with automatic door unit. Academic Research International, 2(2), 168-183.

[6] Kassim M., Mazlan H., Zaini N. & Salleh M.K. (2012). Web-based student attendance system using RFID technology. In Proceedings of the 2012 IEEE Control and System Graduate Research Colloquium, (pp. 213-218). Shah Alam: IEEE.

[7] Mohammed, A.A., & Kameswari, J.U. (2013). Web-server based student attendance system using RFID technology. International Journal of Engineering Trends and Technology, 4(5), 1559-1563.

[8] Patel, R., Patel, N., & Gajjar, M. (2012). Online students' attendance monitoring system in classroom using radio frequency identification technology: a proposed system framework. International Journal of Emerging Technology and Advanced Engineering, 2(2), 61-66.

[9] Silva, F., Filipe, V. & Pereira, A. (2008). Automatic Control of Students' Attendance in Classrooms Using RFID. In D. Davcev (Ed.), IEEE Third International Conference on Systems and Networks Communications, (pp. 384-389). Sliema: IEEE Computer Society.

[10] Yuru, Z., Delong, C., & Liping, T. (2013). The Design and Research of Student Attendance System Nodes Based on RFID. In Proceedings of the 2nd International Conference on Computer and Applications: Vol. 17, (pp. 235-238). Harbin: SERSC.

[11] Singhal, Z. & Gujral, R.K. (2012). Anytime anywhere-remote monitoring of attendance system based on RFID using GSM network. International Journal of Computer Applications, 39(3), 37-41.

[12] Saparkhojayev, N., & Guvercin, S. (2012). Attendance Control System based on RFID-technology. International Journal of Computer Science Issues, 9(3), 227-230.

[13] Tiwari, A.S., Tiwari, A.S., Ade, N.M., Sheikh, S., Patel, N.R., & Khan, A.R. (2014). Optimized Design of Student Attendance System Using RFID. In International Conference on Machine Learning, Electrical and Mechanical Engineering (pp. 134-139). Dubai: International Institute of Engineers.

[14] Wahab, M.H.A., Kadir, H.A., Mutalib, A.A. & Mohsin, M.F.M. (2010). Design and development of portable RFID for attendance system. In Z. A. Bakar (Ed.), IEEE International Conference on Information Retrieval & Knowledge Management (pp. 173-178). Danvers: IEEE.

[15] Yadav, R., & Nainan, S. (2014). Design of RFID Based Student Attendance System with Notification to Parents Using GSM. International Journal of Engineering Research & Technology, 3(2), 1406-1410.

[16] Bhalla, V., Singla, T., Gahlot, A. & Gupta, V. (2013). Bluetooth Based Attendance Management System. International Journal of Innovations in Engineering and Technology (IJIET), 3(1), 227-233.

[17] Subramaniam, H., Hassan, M, & Widyarto, S. (2013). Bar Code Scanner Based Student Attendance System (SAS). Technology of Information and Communication Journal, 1(3), 173-177.

[18] Joshi, P.S. & Joshi, S.S. (2014). Zigbee Based Student Attendance System Using Lan Networking. IOSR Journal of Electronics and Communication Engineering, 9(6), 7-12.

[19] Nawaz, T., Pervaiz, S., Korrani, A. & Azhar-Ud-Din (2009). Development of Academic Attendance Monitoring System Using Fingerprint Identification. International Journal of Computer Science and Network Security, 9(5), 164-168.

[20] Potadar, M.P., Marathe, V.V., Khose, A.S. & Kotkar, L.A. (2015). Biometric Attendance Recording and Communication System. International Journal of Innovations in Engineering and Technology, 5(2), 230-234.

[21] Shafi, Q., Khan, J., Munir, N. & Baloch N.K. (2010). Fingerprint verification over the network and its application in attendance management. IEEE International Conference on Electronics and Information Engineering, Vol. 2, (pp. 555-559). Kyoto, Japan.

[22] Shoewu, O. & Idowu, O.A. (2012). Development of attendance management system using biometrics. The Pacific Journal of Science and Technology, 13(1), 300-307.

[23] Talaviya, G., Ramteke, R. & Shete, A.K. (2013). Wireless fingerprint based college attendance system using Zigbee technology. International Journal of Engineering and Advanced Technology, 2(3), 201-203.

[24] Verma, P. & Gupta, N. (2013). Fingerprint Based Student Attendance System Using GSM. International Journal of Science and Research, 2(8), 128-131.

[25] Uddin, N., Rashid, M.M. & Mostafa M.G. (2016). Development of Voice Recognition for Student Attendance. Global Journal of Human-Social Science Research: G Linguistics & Education, 16(1), 1-6.

[26] Reda, A., Panjwani, S. & Cutrell, E. (2011). Hyke: a low-cost remote attendance tracking system for developing regions. In Proceedings of the 5th ACM workshop on Networked systems for developing regions, (pp. 15-20). New York: Association for Computing Machinery.

[27] Kar, N., Debbarma, M.K., Saha, A. & Pal, D.R. (2012). Study of implementing automated attendance system using face recognition technique. International Journal of computer and communication engineering, 1(2), 100-103.

[28] Patil, A. & Shukla, M. (2014). Implementation of Classroom Attendance System Based on Face Recognition in Class. International Journal of Advances in Engineering & Technology, 7(3), 974-979.

[29] Patil, C.S., Karhe R.R. & Jain M.D. (2014). Student Attendance Recording System Using Face Recognition with GSM Based. International Journal of Research in Advent Technology, 2(8), 32-35.

[30] Tamimi A.A., AL-Allaf O.N.A., & Alia M.A. (2015). Real-Time Group Face-Detection for an Intelligent Class-Attendance System. International Journal of Information Technology and Computer Science, 7(6), 66-73.

[31] AccuClass – Classroom Attendance Tracking in the Cloud. Retrieved June, 2020, from http://www.engineerica.com/.

[32] Attendance Tracking for Colleges and Universities. Retrieved June, 2020, from http://aplusattendance.com/.

[33] Register Attendees for Events, Conferences and Trade Shows. Retrieved June, 2020, from http://www.jollytech.com/.

[34] Student Attendance Management System. Retrieved June, 2020 from http://www.creatrixcampus.com/.

[35] Student Attendance Tracking, Barcode Scanning for Student Attendance. Retrieved June, 2020, from http://www.mss-software.com/.

[36] Attendance Tracking with Browsers, QR code and Phones. Retrieved June, 2020, from https://www.regular.li/.

[37] Akhila K., Prathyusha B., PavanKumar M., & Amrutha M. (2013). A Novel Approach of Mobile Based Student Attendance Tracking System Using Android Application. International Journal of Engineering Research & Technology, 2(4), 2139-2143.

[38] Attendance Tracker. Retrieved June, 2020 from https://www.knack.com/

[39] BePunctual – School Class Attendance System. Retrieved June, 2020, from http://www.bepunctual.com/

[40] MyAttendanceTracker. Retrieved June, 2020, from https://www.myattendancetracker.com/

Increasing Users Safety Supported by Internet Emergency Call System

Z. Nyikes
Óbuda University / Bánki Donát Faculty of Mechanical and Safety Engineering, Budapest, Hungary
nyikes.zoltan@uni-obuda.hu

Abstract - Nowadays the internet supported applications and the available possibilities reformed our life. This can be said based on surveys that showed that people use the internet independently on age or locality. Besides this practise the digital competency and the safety awareness are low level. The organized real-time digital defence solution of the civil users and the small and middle enterprises are not available yet. Currently, the available service supports means help after the incident by the manual and static way. In this paper, I introduce own invented system that means real-time and dynamic defence solution possible for the public users and for the undefended users who are not defended by laws obligation. The case in points incidents is cyber-grooming, virus attacks and system disaster.

1. Introduction

When everything is all right in our life, we don't need help. In an ideal world, we would not need an organization that protects people or helps them out in a time of harm, because harm would be non-existent.

Active help in certain situations may save lives, especially when there is a dangerous situation. Since the formation of the middle class, various help providing organizations work in the form of institutions, in almost every country.

Good examples for this is the police, the fire service and the ambulance. We alert these organizations, when an unexpected situation occurs, that we are unable to handle legally, physically, or in absence of knowledge.

In the case of this situation, they try to help on the emergency event location, or they try to provide help by telephone. With the advancement of communication, calling for help, and the reaction for the call become faster and more accurate.

2. Emergency Situations and Maintenance Nowadays

2.1. The Emergency Call System

With the appearance of the phone network, calling for help through a phone become possible. The call center operator had to be asked to alert, or connect the help providing the facility.

The first emergency call system was established in London, at 1937 July 1st, using the call number 999, which was quickly extended to the entire country.

At 1946 the Southern California Telephone Company started to use the 116 call number in Los Angeles.

At 1959 they adopted the 999 call number in Winnipeg, Manito and Canada, pressed on by Stephen Juba, governor of Winnipeg. The city changed its number to 911 in 1972, to adapt to American standards.

In the sixties - before the introduction of the 111 New Zealand emergency call number - the city of Auckland, 40 call center has operated, each with there owns emergency call number, and the local numbers had to be found in the cities five hundred pages long phonebook.

This problem was partially solved in the United States, Canada and the United Kingdom, there 0 had to be called in case of danger. In modern times these numbers can be called quickly on a mobile phone, even if there is no SIM card inside the mobile.

The number is different for each country, usually an easy to remember, three-digit number, which can be called quickly. In certain countries, multiple emergency call number exist in different situations. Inside the European Union, since the nineties, the number 112 is used. Mobile phones and SIM cards contain certain pre-programmed numbers.

When the user calls for an emergency call number, that is known by the device, the network automatically transmits the call onto the locally used emergency call number. Most GSM device can this number can be called even if the keyboard is closed, the SIM card is missing, or even if the caller uses the number as a PIN code.

A. The Emergency Call System in Hungary

In our country, in 2011. on the 7th of June, the government decision of the creation of a free emergency call system. The calls arrive at one of two police centers - Szombathely and Miskolc - and received by experienced, law enforcement, fire protection, and medical operators.

They record the notifications to so-called "intelligent datasheets" which immediately arrives at the local police, firefighter, or ambulance department. Therefore help may arrive under the shortest possible time and based on accurate information.

Calls for the police's 107 and the firefighters 105 arrive at 170 different places in the country. Calls for the ambulances 104 arrives onto 26 different centers. Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.

2.2. Cyber-Grooming

In today's digital age, children are not only exposed to the immediate community that visibly surrounds them, but also to that less visible and less tangible world of cyberspace. As a result, they are exposed to a new level of vulnerability that did not exist before. It can be argued that the emergence of communication technologies in our every day lives may be considered a contributing factor to the increase in ways in which children may be sexually victimised. Children may be victimised online in several ways: they may become the subjects of indecent images; they may be groomed for sexual abuse which takes place offline or they may be groomed online and the abuse may be carried out via the use of webcams, for example.[2]

B. What is Cyber-Grooming?

Online grooming is a modus operandi that is commonly associated with sexual abuse and it can involve both children and adults. However, it is a method that perpetrators can use to prepare a person to commit several different other crimes. [2]

In the United Kingdom, there is a law, the legislation in the Sexual Offenses Act 2003, which contains "sexual grooming". The intention is to intervene at an early stage to prevent any sexual contact. [3]

Definition by law: A process by which a person prepares a child, significant adults and the environment for the abuse of this child. Specific goals include gaining access to the child, gaining the child's compliance and maintaining the child's secrecy to avoid disclosure. This process serves to strengthen the offender's abusive pattern, as it may be used as a means of justifying or denying their actions. [3]

This process can continue over days, weeks, or even years. This behaviour can take place both when people are nearby and distant, that is, via the Internet. By law, the police can apprehend an offender before they become physically involved with a child, resulting in more protection for children. Furthermore, the risk presented by offenders convicted under Section 15 can be managed more easily because they will have to register as sex offenders. It does allow the police to conduct proactive policing via the Internet; by using the anonymity that the Internet affords; police officers can play child sex offenders at their own game. These operations entail police officers posing as vulnerable children in Internet chat rooms and waiting for offenders to prey on them. Offenders, who arrange to meet these undercover police officers, will then be arrested subject to the availability of evidence that proves their sexual intent. Chief police officers highlight the difficulty of obtaining proof of sexual intent, which varies in difficulty depending on the medium used to sexually groom children. Internet grooming is much more likely to make the grooming process explicit, and it is more likely that evidence of the sexual nature and intent of the behaviour can be obtained, that is, via the computer trail that remains (or through records of conversations with undercover police officers). In contrast, there is much less likely to be such tangible evidence in 'face to face' sexual grooming interactions. Here, the identification of behaviours as sexually motivated relies on people being witness to it and recognising it as such. If the public and/or police are unable to recognise and identify the full range of sexual grooming behaviours, the scope of the legislation will be further reduced. Thus, a greater understanding of sexual grooming is required to aid the identification of this behaviour and to increase public awareness. This organisation provides a free helpline to adults who have, or who are thinking about, offending against a child, and for adults concerned about the sexual behaviour of known adults and children towards other children. [3]

Offenders contact the helpline anonymously in the time before committing an offence and the trained helpline staff can help them manage their risk and prevent them from offending. Increased awareness more generally could aid the identification of an individual's sexual grooming behaviour as this person grooms the environment and significant adults in the community, or when they groom the child victim; yet before any contact offence takes place. Thus, the public needs to be made aware of how a potential offender would groom adults, children and the environment. It is important, however, that this is managed in a way whereby appropriate help and intervention is sought. Increasing awareness of sexual grooming without increasing awareness of appropriate interventions, or of the adverse impact of extremely negative behaviour towards the 'groomer', is likely to result in vigilante culture and potential offenders going 'underground' as discussed previously. [3]

C. Perils of Cyber-Grooming in Virtual Worlds

The EU has explicitly admitted to making the Internet a safer place for kids and minors. The discussion in the EU on protecting children is finally focused on the question to just block or completely delete harmful websites [4]. This is not targeted to online environments (like virtual worlds, browser games, online apps), but is stuck to an out-dated content-oriented (not: communication-oriented) view. Kids primarily explore the internet by starting to play games, not by surfing the Web. Those online environments are especially attractive because of their possibilities to interact and communicate with other players (like shared game experience, chats) [5]. There are certain offers with a design and game mechanism that is particularly suitable for children, where they are especially exposed to enter close emotional relationships with others. Besides other legal issues, this is intensively exploited by pedo-criminals in a targeted manner [6]. Such initiations of sexual interactions with minors are called Cyber-Grooming. This important, but not yet sufficiently covered topic was primarily addressed by the symposium "Protection of Children and Minors on the Internet – Perils of Virtual Worlds" on 19 September 2012 in Brussels. Starting from a criminological overview of the phenomenon, aspects of law, society and IT for protecting kids and minors against Cyber-Grooming have been considered by respective experts, and first experiences with virtual police offices in an online game for kids have been presented [7]. In the following, the political consequences have been discussed with representatives of the EU Commission and parliament. All participants agreed that the Internet is an important part of today's media reality and that providing related skills as well as adequate protection of minors is a central goal of our efforts. [8]

2.3. Protection of Informatics Systems Against Codes

The internet and the smartphone can not only save lives today, with there malfunction and disrupted operation, but these items can also create problems, that in the worst case, cost lives.

Unfortunately, many don't pay enough attention to the protection, and operation of informatics systems. It can come from the user's negligence or the lack of digital competence.

But new malicious codes appear daily, ones that beyond the user, can accurately attack complete systems, like hospitals, airplanes or banks data. This would be a problem in any case, but as we see people's life may be in danger because of these attacks. [9]

2.4. Protection Against Online Harassment

Beyond malevolent programs, we may talk of malevolent people, online harassers too. These people harassing their victims, while hiding behind the anonymity of the internet. [10]

In situations like this, the victims are defenceless, vulnerable and alone. In any case, they have no idea what to do, or who to ask for help. Many times, the police report follows a harassment case, but before the victim could reach the police, the "clues" become too old for use.

2.5. Safety of Informatical System's Operation

May occur, that the problem isn't caused by hostile computer programs or harassers, but a malfunctioning application, or the operating system brakes down in a way that the user is unable to handle. [11]

The solution might be in a system, that works similarly to emergency call systems. On its model, there would be an online service, which would be capable to provide immediate help, stopping the escalation of the virus code or provide a way to fix the system's error.

In the above mentioned two cases, beyond technical help, an investigation would be possible to find the perpetrator's and put them under prosecution. While in the third case, only technical help is possible.

3. The Internet Emergency System Realizing and Applicability

With the appearance of the internet and especially smartphones, multiple new applications were released, like the complex emergency call systems, the user can call help by the push of a single button and providing geopositional data.

Many car factories started to develop onboard devices, which in case of accidents, sudden change of speed or leaving of the road, immediately alerts the dispatcher center. But emergency call systems are used on different forms of travel, like in sailing, since the start of the 1900s. With the help of the "SOS" Morse code - sent through radiotelegraph, an abbreviation of "Save Our Souls" - they could signal distress.

This was later used in different territories. As radio technology advanced, the "Mayday" distress call, or the "Pan Pan" emergency call, furthermore the "Securite" safety call, got into use.

Applications and services, already exist, that in case of a road accident, automatically activate the emergency call system, according to the geopositional data, and the user may contact the operator in the call center, who in turn acts according to the situation.

To the formation of the system, a dispatcher center is required, which beyond technical help, can serve as authoritative functions. This can be integrated into the United European emergency call dispatcher center, using the phone number 112, or it can be operated by the National Institute of Cyber protection.

The National Institute of Cyber protection, in accordance with the I. law of 2013 [8] is the Hungarian, governmental event management center, which also provides the so-called GovCERT service with governmental authority. Furthermore, there are other organizations, providing help against malfunctions and harassments.

Otherwise, the supporting system might appear within the currently operational customer service of phone, and broadcasting companies. However, currently exist no available service, for citizen users, within the market's supply of, that fulfils my mentioned requirements.

Through the Hun-CERT, a service is available, where the investigation may be started, by a notification on a fixed user interface (by the description of an event).

However, that can only be started after the incident happened, which takes away valuable time from applicants. The mission of Hun-CERT is to help the Hungarian Internet Society, and the Hungarian Internet Providers, dealing with internet hazards. [9]

However, beyond a static datasheet, and a phone number available at 24/7, they are unable to provide help. There are different solutions against online harassment, online many helpful sites, and in better case, phone numbers are available.

These are also static and not real-time services. My outlines may work as independent organizations, taking pressure off from the above-mentioned systems.

By function, this would cover a continuously growing and advancing territory. I wish to call this system "Internet Emergency Call System" (IECS).

3.1. The Implementation of the Internet Emergency Call System

In practice, at the push of a button, by the use of a virtual private network - VPN - and through a predefined and separate channel, the dispatcher center would be contacted, which in turn would provide remote assistance to the victims.

If needed, an investigation would immediately start to find the perpetrators. In accordance with the emergencies type, different functions could be activated. In case of a hacker attack, the internet could be cut off, so the malicious code infecting the software can't be spread further on the network.

With the help of VPN the dispatcher center takes control over the computer, beginning to delete the viral code while finding evidence for the later investigation.

In this case, the dispatcher center may utilize artificial intelligence, capable of faster and more accurate intervention. An operator may oversee the operation or can even overtake it.

In case of online harassment, the users' internet connection would not be cut off, but the dispatcher center could provide remote assistance and instruction for the victim, by overseeing the messages or web calls of the harasser, by the dedicated VPN system.

Figure 1. IECS - Internet Emergency Call System Applicability (by author)

In a possible function, the dispatcher center may use the victim's profile to communicate with the harasser as an expert. In this case, there is a way to collect evidence against the perpetrators. Here I find the work of a dispatcher necessary, for the required empathy of the task.

In case of technical malfunction, a dedicated VPN would also grant access to the dispatcher center, to provide remote help, and try to restore the system. The use of artificial intelligence supervised by an operator can be also a solution here. These solutions would make the use of the human workforce more effective. The proposed system is shown in Fig.1.

3.2. *The Applicability of the Internet Emergency Call System*

The Internet Emergency Call System may fill in a current space. Seeing the advancement of computer science in the last decades it would be daring to imagine what territories will computer science appear in. [12]

By past experiences, we can conclude that computer systems will receive more and newer attacks, the prevention of which and to deconstruct the current situation, we have to find new solutions. [13]

We expect the use of IoT systems to spread further, therefore users and providers will be more vulnerable, and have to get ready for more attacks by the internet. [14] Therefore my conceived emergency call system is essential. In my research, I found no signs of a similar system in practice.

A system like this could serve a currently grey area unprotected by law or economical interest with any similar solution. [15] Like the computer systems of small and medium companies, which despite not serving a large number of workers, users, or databanks, still significant for there numbers even just by looking at Hungary.

These small and medium companies have a significant role in the local production and service industry. Therefore the governments' current Irinyi-program, wishes to support the local small and medium-sized companies in the framework of the current, fourth industrial revolution.[16]

Furthermore, my recommended system would be significant for individual internet users. Nothing that a singular internet user don't mean a critical mass, but the number of users is significant almost covering the entire population. [17]

We have to mention those schools, that by being private, not under the effect of the above mentioned L. law of 2013. To digitalize schools, and to develop the student's digital competence, the up to date digital infrastructure of schools, and the protection of that infrastructure is a key question.[18]

Figure 2. IECS button on the screen (by author)

Therefore the protection of these institutions would have paramount importance, where the recommended system of mine, could also be applied. The design of the application is shown in Fig. 2.

My concept could also be used in case of organisations or enterprises that are in contact with local authorities or government-owned companies, but also not inclouded under the information security law but are part of the supply chain, providing a weak point for certain protected organizations.

4. CONCLUSIONS

The described emergency call system and the prospects of its use and realisation are conceived in this paper. I presented the emergency call system on the telephone, as the founding of my concept, its creation and state in Hungary. Also, I showcased the use of an online emergency call system, its protection against malicious codes, online harassment and malfunctions.

I showed the concept for the creation and applicability of an internet emergency call system. Within these lines, I detailed the possibilities of the creation of the Internet Emergency Call System and the applicability of the Internet Emergency Call System.

My design, in case of realization, serves the safety, of the users and the operators. The new solution, I showcased, may serve the protection of the users and the various informatical systems not protected by the law, it can revolutionise the solutions and possibilities of defence. Furthermore, it may stop and repels the current tools of cybercrime.

Expectedly informatics systems will touch our homes (smart home), motorways and mass transportation (self-driving cars, intelligent train etc.), therefore defensive solutions must be established, capable of providing real, immediate and authoritative intervention for the above-mentioned users and systems.

The Internet Emergency Call System fits well into Europe's Digital Agenda program made by the European Union, and into the Digital Wellbeing program of the Hungarian government, just like into the Irinyi Program. Further, it would help the realisation and spread of Smart City projects as the formation of these are expected in Hungary, Europe, and in the whole world soon.

REFERENCES

[1] Z. Nyikes, A Közép-Kelet Európai Generációk Digitális Kompetencia és Biztonságtudatosság Vizsgálatának Eredményei, Hadmérnök, XII. 4. 2017, Nemzeti Közszolgálati Egyetem Hadtudományi és Honvédtisztképző Kar, pp. 159-173, ISSN 1788-1919

[2] E. Martellozzo, Online sexual grooming: Children as victims of online abuse, Cybercrime and its victims. Martellozzo, pp. 108-128. ISBN 9781138639447, http://eprints.mdx.ac.uk/28183/

[3] S. Craven, S. Brown and E. Gilchrist, Current Responses to Sexual Grooming: Implication for Prevention, The Howard Journal of Crime and Justice 46(1), pp. 60-71, 2007, ISSN 0265-5527, DOI: 10.1111/j.1468-2311.2007.00454.x

[4] G. Ehrmann, U. Lucke, M. Mulder, T.-G. Rüdiger, T. Schulz-Spirohn, J. Storbeck and D. Woidke, Protecting Children and Minors in the Internet: Perils of Cybergrooming in Virtual Worlds, Oranienburger Schriften, Articles of the University of Applied Science of the Brandenburg Police, 2013. pp. 83–84., Rainer Griege, ISSN: 1865-1062,

[5] I.E. Berson, Grooming Cybervictims: The Psychosocial Effects of Online Exploitation for Youth, Journal of School Violence, v2 n1 p5-18, 2003, ISSN: ISSN-1538-8220

[6] E. Chase, J. Statham, Commercial and Sexual Exploitation of Children and Young People in the UK: a Review, Child Abuse Review., 14., pp.4-25, 2005, DOI: 10.1002/car.881

[7] N. Lorenzo-Dus, A. Kinzel, 'So is Your Mom as Cute as You?': Examining Patterns of Language use by Online Sexual Groomers, Journal of Corpora and Discourse Studies, 2(1):1–30, 2019, DOI: https://doi.org/10.18573/ jcads.31

[8] H. Pranoto, F.E. Gunawan, B. Soewito, Logistic Models for Classifying Online Grooming Conversation, International Conference on Computer Science and Computational Intelligence (ICCSCI 2015), Procedia Computer Science 59, pp.357 – 365, 2015, DOI: 10.1016/j.procs.2015.07.536

[9] A, Holtai, S. Magyar and B. Puskás, Az Informatikai Fejlesztés és Üzemeltetés Határvonalai, Felderítő Szemle, 2016, (1), pp. 191-203., 2016

[10] Z. Nyikes, Creation Proposal for the Digital Competency Framework of the Middle-East European Region, Key Engineering Materials, Vol. 755, pp. 106-111, 2017

[11] Z. Rajnai and B. Fregan, Kritikus Infrastrukturák Védelme, Kolozsvár, Erdélyi Múzeum-Egyesület (EME), 2016., pp. 349-352., Műszaki Tudományos Közlemények 5.

[12] L. LAW OF 2013 ON STATE AND LOCAL GOVERNMENT AGENCIES ELECTRONIC INFORMATION SECURITY (Hungary)

[13] Z. Rajnai, Információtechnológiai Kutatások a Védelmi Szektorban, Az 5. Báthory-Brassai Konferencia tanulmánykötetei., 709 p., Budapest, Óbudai Egyetem Biztonságtudományi Doktori Iskola, 2014., pp. 1-12., 1-2. köt., ISBN:978-615-5460-38-8

[14] Z. Rajnai, Planification of a Transmission Network, Proceedings of the International Scientific Conference: New Trends in Signal Processing. 213 p., Liptovsky Mikulas: Armed Forces Academy of General Milan Rastislav Štefánik, 2012., pp. 134-141., ISBN:978-80-8040-447-5

[15] Z. Rajnai, A Kritikus Információs Infrastruktúrák Összetétele, Biztonsági Kérdései, Nemzetközi Gépész és Biztonságtechnikai Szimpózium: a Magyar Tudomány Ünnepe tiszteletére, Budapest, 2012., Óbudai Egyetem, Bánki Donát Gépész és Biztonságtechnikai Mérnöki Kar, 2012., pp. 15-22., ISBN:978-615-5018-35-0

[16] T. Kovács, Z. Nyikes and D. Tokody, Komplex Monitoring-Rendszer Használata Vasúti Felépítmény Vizsgálatában az Ipar 4.0-hoz, XVII. Műszaki Tudományos Ülésszak előadásai. ISSN 2393–1280, EME, MTK 6., Kolozsvár, 2017, http://eda.eme.ro/handle/10598/30075?show=full

[17] D. Tokody and F. Flammini, Smart Systems for the Protection of Individuals, Key Engineering Materials, Vol. 755, pp. 190-197, 2017, DOI: 10.4028/www.scientific.net/KEM.755.190, https://www.scientific.net/Paper/Preview/52515

[18] D. Tokody and Gy. Schuster, Driving Forces Behind Smart City Implementations-The Next Smart Revolution., Journal of Emerging Research and Solutions in ICT 1.2, 2016, pp. 1-16., http://eprints.fikt.edu.mk/171/

Industry 4.0 in the Context of Coal Mining

Vladimir Simeunović*, Sonja Dimitrijević *, Dragan Stošić*and Snežana D. Pantelić*
* Mihailo Pupin Institute, Belgrade, Serbia
vladimir.simeunovic@pupin.rs, sonja.dimitrijevic@pupin.rs, dragan.stosic@pupin.rs, snezana.pantelic@pupin.rs

Abstract - Industry 4.0 is one of the most popular topics nowadays. It is an umbrella term that connects different concepts and technologies such as the Internet of Things (IoT), Cyber-Physical Systems (CPS), Smart Factories, etc. It is highly relevant for the coal industry, which needs to respond to future productivity challenges, numerous safety issues, and ecological and environmental protection challenges as well. The main objective of the paper is to reveal the current research trends, opportunities, impact and limitations of Industry 4.0 in coal mining as envisioned in recent literature. Therefore, the paper selects, reviews, summarizes and discusses research initiatives and studies that explicitly put Industry 4.0 in the context of coal mining. Many Industry 4.0 opportunities have been recognized in the reviewed papers, especially those coming from the implementation of IoT, CPS, automation, intelligent systems and big data. Although the impact of Industry 4.0 on coal mining productivity and/or safety has been widely recognized, its impact on ecological and environmental protection has been mostly addressed in the background or neglected.

I. Introduction

Coal is one of the fundamental fossil fuels in the world. The global coal consumption peak is expected in the mid-2020s [1]. It is also expected that the absolute total amount of coal consumption will remain relatively stable. At the same time, the share of coal consumption in total primary energy consumption will be gradually decreasing but staying above 50% [2].

Not only does the coal industry need to respond to future productivity challenges, but also to the increasing pressure from ecological and environmental protection [3]. To this end, digital innovations in coal mining are more and more recognized as key factors for future productivity [4], human safety and environmental protection [3]. Consequently, such innovations can have a much wider impact on individuals, organizations and even communities [4].

Some of the ultimate goals of mining, therefore of coal mining as well, are not new and involve [5]:

- continuity of all processes thanks to the process control systems which monitor and optimize performance
- autonomous moving equipment (machinery) which is controlled from afar
- a few people on or above the ground
- zero environmental footprint
- zero net energy consumption
- safe and healthy working environment

The increasing convergence of ICT, primarily the Internet of Everything, and the industrial value chain has formed the foundation for the next industrial revolution - Industry 4.0 [6].

Industry 4.0 is an umbrella term that connects concepts and technologies of utmost importance primarily for the success of manufacturing and production industries. Industry 4.0 is also highly relevant for mining industries, which are heavily dependent on different types of machinery and equipment.

Industry 4.0 is in line with the ultimate goals of (coal) mining. Moreover, it can add more value in coal mining processes thanks to the research and practice results from different research fields and other industries. Industry 4.0 raises the focus on technologies enabling communication and coordination of machines on and under the ground. Therefore, it can improve the visibility, continuity and performance of coal mining processes. It can also lead to a reduction in the number of people in danger zones and a decrease of environmental impact, and, in general, help meeting the mining goals. Moreover, Industry 4.0 "is boosting the relevance of predictive maintenance for manufacturing and production industries" [7].

Thanks to the Internet of Things (IoT), meaningful human-to-human, human-to-machine and machine-to-machine connections are increasingly used in the context of predictive maintenance. To this end, research and innovation efforts have been put into the development of robotized technologies, machinery with the elements of artificial intelligence (AI), and system automatic controls [8].

The main objective of the paper is to find out the current research trends regarding Industry 4.0 in coal mining. Therefore, the paper selects, reviews, analyses, summarizes, and discusses research initiatives and studies that explicitly put Industry 4.0 in the context of coal mining. In particular, the paper focuses on opportunities, impact and limitations of Industry 4.0 in coal mining as acknowledged in recent literature. Moreover, some research gaps in this, still early, research phase of Industry 4.0 have been identified.

To the best of the authors' knowledge there is no other review on this specific topic. By providing a current cross-section of the novel Industry 4.0 paradigm and the highly important coal mining domain, the review is of interest not only for researchers and practitioners, but also for stakeholders in the domain.

II. Industry 4.0 Concepts

A paradigm shift from centrally controlled to decentralized production processes characterizes the fourth industrial revolution, which has been triggered by the rising capabilities of the communication between people, machines, and resources [6]. Despite the fact that Industry 4.0 is one of the most popular topics nowadays, there is no clear definition of the concept. Even the main promoters, the "Industrie 4.0 Working Group" [9] and the "Plattform Industrie 4.0" [10], have not provided a clear definition along with the vision, the main aims, the basic technologies and the selected scenarios of Industry 4.0 [6].

The term has become publicly known in 2011, when an association of representatives from business, politics, and academia has gathered around the initiative "Industrie 4.0". The main objective of the initiative is to support strengthening the competitiveness of the German manufacturing industry [6]. Moreover, the German federal government has supported the initiative by including Industry 4.0 into "High-Tech Strategy 2020 for Germany".

The three key components of Industry 4.0 according to the first recommendations of the "Industrie 4.0 Working Group" are the Internet of Things (IoT), Cyber-Physical Systems (CPS), and Smart Factories [6].

The IoT enables 'things' equipped with RFID, sensors, actuators, and mobile devices to interact with each other and work collaboratively towards reaching common goals [11]. CPS are systems that integrate computation, networking and physical processes, i.e., deeply intertwine physical and software components that interact with each other in a multitude of ways adapting to the context [6][12]. Some of the examples of CPS are smart greed, autonomous automobile systems, industrial control systems, robotics systems and medical monitoring [12]. Finally, Smart Factories are context-aware systems that integrate IoT and CPS to assist people and machines in task execution. Smart Factories perform their tasks acting upon information from both physical (e.g. position or condition of a machine) and virtual world (e.g., electronic documents, drawings and simulation models) [13].

On top of these three key components, Industry 4.0 relies upon other popular ideas and research topics such as automation, artificial intelligence, big data and cloud computing [14][15]. Moreover, robots, cooperating and coordinating machines, self-decision-making systems, autonomous problem solvers, learning machines, 3D printing, etc. are expected to play a much bigger role in manufacturing [16]. The Industry 4.0 wheel on Figure 1 represents some of the key concepts of Industry 4.0.

Hermann, Pentek, and Otto [6] have found four principles that guide researchers and practitioners towards Industry 4.0 based on the quantitative text analysis and qualitative literature review. These principles are: interconnection, information transparency, decentralized decisions, and technical assistance.

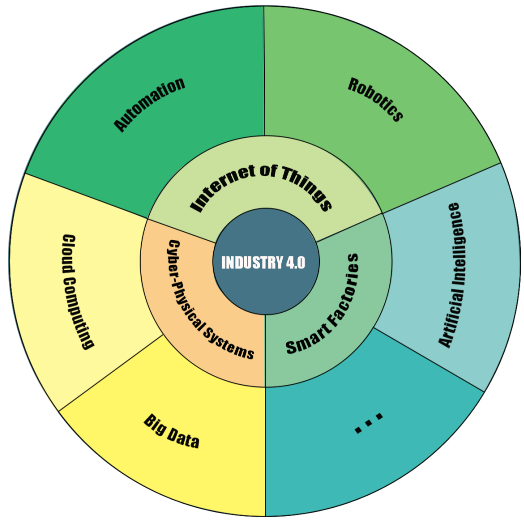


Figure 1. Industry 4.0 Wheel[1]

Based on the more recent literature review, Oztemel and Gursev [16] have defined six design principles of Industry 4.0: interoperability, virtualization, local, real-time talent, service orientation and modularity.

On the one hand, academic research is trying to define the concept and develop respective systems, business models and related methodologies. On the other hand, industry is primarily focused on the features of Industry 4.0 (industrial machine suits and intelligent products) that can enable transformation from machine dominant manufacturing to digital manufacturing, as well as on potential customers that such progress could bring [16].

III. Coal Mining Processes and Machinery

There are two main types of coal mining: underground and surface coal mining. Bord, pillar and long wall mining characterise underground coal mining. Furthermore, open--cast strip mines and truck-and-shovel operations characterize surface coal mining [4].

Among the bord and pillar machinery, there are continuous miners, entry drivers, bolting systems, haulage systems, feeder breakers, and loaders [4]. Longwall shearers are used in the exploitation of coal deposits in long wall mining. They are typically employed together with a powered roof support and a scraper conveyor. A longwall shearer mines the body of coal by moving forward on a longwall conveyor [17].

An open-pit coal mine heavily depends on the excavation and haulage system, which is made up of two primary pieces of machinery: excavators/shovels and trucks [18]. Whereas excavators/shovels dig ore or waste material out of the ground, trucks transport it to the dedicated destination [19][20]. Long-range overland or aerial conveyors are an alternative to out-of-mine transport [5]. In addition, many auxiliary operations are

[1] Special thanks to ADD Design, http://add-design.best/

implemented (auxiliary excavation, relocation of conveyor belts, workmen transport, ...) in order to provide the uninterrupted basic coal mining operations. To this end, machines such as bulldozers, trench excavators, pipe-layers, etc. are also used [21].

Working in coal mines, especially in underground mines, is immensely difficult and dangerous because of the complex and unpredictable environment. Many natural and technical hazards such as wall failures, vehicle collisions (surface coal mining), dangerous gases such as methane, high temperature, noise, the risk of rock ejection, etc. (underground coal mining), are inherent to the coal exploitation process. In addition, coal exploitation has a very negative impact on the environment (e.g., surface deformation, production and storage of waste and gangue, hydrological changes, pollution of water and air), and consequently on the infrastructure located on the surface [22]. Accordingly, some operations in coal mines are focused on mitigating and minimizing prospective hazards.

The Industry 4.0 paradigm is in line with increasing efforts in coal mining to improve energy efficiency by reducing both the cost of mining operations and their environmental impact. To be able to do so, the current mining processes supported by the state-of-the-art mining systems, machinery and equipment should be carefully considered.

IV. Industry 4.0 Research Initiatives in Coal Mining

The reviewed research initiatives have been categorized into:

- research initiatives with a broad focus, and
- research initiatives with a narrow focus

Therefore, the review results have been presented in the following subsections according to the given categories.

A. Research Initiatives with a Broad Focus

Most of the research initiatives and studies that have been reviewed within this paper have a rather broad research focus. This means that they have addressed different Industry 4.0 opportunities in coal mining. The following text summarizes the most significant initiatives and studies that are relevant for the topic of the paper.

U.S. National Institute for Occupational Safety and Health (NIOSH) researchers have reviewed existing sensors and communications network systems used in U.S. underground coal mines to find out if they have the ability to support the Industrial Internet of Things (IIoT) [23]. The authors have equated the term IIoT with the term Industry 4.0. The results have shown that “about 40% of the installed post-accident communication systems as of 2014 require minimal or no modification to support IIoT applications” (p.50). In addition, the researchers have conducted the feasibility study to investigate a specific IIoT application based on existing communication and tracking infrastructure.

The Chinese Academy of Engineering initiated ‘‘Technical revolution in ecological and efficient coal mining and utilization & intelligence and diverse coordination of coal-based energy system”. It has defined three stages (3.0, 4.0 and 5.0) of China’s coal industry development strategy. The following main objectives of China Coal Industry 3.0, 4.0 and 5.0 respectively have been pointed out [3][24]:

- ‘‘reduced staff, ultra-low ecological damage, and emission level near to natural gas” by 2025
- ‘‘near-unmanned mining and near-zero emission” in 2035.
- ‘‘no coal above ground, no staff underground, zero emission and zero damage”

Wang, Xu, and Ren [3] have analyzed the development trends of the China Coal Industry 3.0 and its support for the stages 4.0 and 5.0. They have concluded that the current stage is critical for transition from ‘‘functional” to “quality form” that the following stage should bring.

Palka and Rizaoğlu [22] have proposed the concept of mine 4.0, i.e., the use of Industry 4.0 concepts and technologies in underground hard coal mining. For each concept/technology they have defined the area of application, and the main advantages and disadvantages as well. For example, the authors have recommended the use of [22]:

- autonomous machines in the areas of application such as mining of coal, protection of the roof, crushing of rocks, transportation of spoil, and safety in general;
- big data in the areas such as machine operation parameters and equipment, mining level data, transport and coal processing data;
- autonomous gas meters in the areas such as security and data collection;
- chips identifying employees for monitoring employees;
- innovative modeling software for virtual deposit modeling;
- GIS system for digital mapping of a mining company;
- augmented reality in the areas such as workplace training, vocational training, service and maintenance of machines, maintenance;
- etc.

Among the most common advantages of mine 4.0, Palka and Rizaoğlu [22] have identified improvement of work efficiency, safety and integration of mining company departments. Among the most repetitive disadvantages, there are high costs, limitations of the mining environment and the need for qualified staff.

Sankaranarayanan et al. [25] have analyzed the influence of principles and technologies of Industry 4.0 over coal industries considering eight crucial factors:

consumer, water resources, smart transportation, smart factory, smart grid, smart mining, smart home, and renewable energy. The total interpretive structural modeling (TISM) method has been used in the analysis. The model has distinctly shown the influence of Industry 4.0 principles and technologies over coal industries.

Nepsha et al. [26] have discussed the main problems in increasing efficiency of the coal-dependent electrical power system (EPS). The problems have been categorized into:

- problems that can be solved at the design stage, and
- problems that can be solved at the operation stage.

The authors have proposed the development of an intelligent EPS based on modern technologies such as IoT (Internet of Things) and big data. In perspective, it will involve a unified cyber-physical system ensuring the rational management of the coal-dependent EPS. The authors see such a solution as an introduction into the implementation of the "Industry 4.0" concept, whose importance has been recognized in the Russian coal industry [26].

Samorodova et al. [27] have analyzed issues related to the development of the digital ecosystem in the Kuzbas coal-mining region in Russia within the Industry 4.0 framework. The authors have found that the formation of the digital ecosystem and the use of the "Internet of Things" are the key factors of the new stage in the mining industry restructuring. They believe that Russia has already gained positive experience through the projects "smart mine" and "intellectual open-cast mine" in the Kuzbass mines and coal cuts. This experience should facilitate the transition to a digital ecosystem in the region.

Pałaka et al. [17] have conducted a survey among the group of underground coal mining workers in Poland: mechanical fitters, electricians and miners. The survey is about the development of new technologies for underground coal mining starting from the Industry 4.0 perspective. Based on the results of the survey, the authors have discussed some opportunities arising from the introduction of Industry 4.0 in underground coal mining for improving safety and increasing coal production. Some of these opportunities are the introduction of autonomous machines and intelligent systems, predictive maintenance and automation.

Hao, et al. [14] have reviewed and analyzed studies on intelligent ground control in coal mining with the application of the IoT in the context of Industry 4.0. They have found that this topic is understudied. Moreover, they have proposed an IoT-based architecture of ground control, which includes the perception, communication, service, and interaction layer. Ground control addresses ground stability and concerns risks such as pillar stability [28][29], rock bolt supporting quality [30][31], roof fall [32][33], excavation face deformation [34][35], etc. Hao, et al. [14] have concluded that the key concepts of Industry 4.0 could help approaching ground control with some new insights.

Although not specifically focused on coal mining, Lööw, Abrahamsson, and Johansson [36] have formulated the notion Mining 4.0. They have attempted to envision the future of mining and to illustrate various possible outcomes by formulating two scenarios: one utopian and one dystopic.

B. *Research Initiatives with a Narrow Focus*

Some of the reviewed research studies have rather a narrow research focus. However, their results have been positioned within the broad Industry 4.0 concept.

Starting from the goals of Industry 4.0, Rylnikova, Radchenko, and Klebanov [8] have proposed an approach for intelligent and automated control of machinery considering safety risks and the necessity for human presence. The following zones have been defined based on periodicity of human presence:

- zero entry production areas,
- zones of periodic human presence due to performing main or auxiliary mining operations, and
- zones of permanent human presence.

They have analyzed and discussed the prospective improvements of machinery operating in each zone.

Hoosain, et al. [4] have proposed an intelligent technique for assisting in the control of shuttle vehicles, which have been used in the bord and pillar mining. Shuttle car movements have been modeled based on queuing, path constraints and shortest paths. To this end, the following optimization algorithms have been used: Dijkstra's shortest path algorithm, K-shortest path algorithm and discrete event simulation. Moreover, this research initiative has provided a graphical user interface (GUI) to assist mining operators.

Mardonova and Choi [37] have reviewed the trends of open-source technology for the mining industry. On that basis, they have proposed a low-cost environmental monitoring system for non-metallic underground mines to support mine safety and occupational health issues. The results of the field experiments have shown that despite few limitations related to the precision of the low-cost dust and gas sensors, the mining industry (including coal mining) can benefit from open-source technology deployment when taking into account cost factors.

Evsutin and Meshcheryakov [38] have discussed the possibility to increase the efficiency of mining enterprises (including coal mining) at the cost of the development of common information space. In order to solve security problems at the level of data flow between the participants of the production process, the authors have proposed an approach based on the technology of blockchain and digital watermarking.

V. Discussion and Conclusion

Industry 4.0 puts undivided focus on productivity, safety, ecological and environmental protection. Therefore, coal mining is a particularly challenging domain for Industry 4.0. Yet, Industry 4.0 principles, technologies and experience from different industries and research areas can be very beneficial in coal mining as well.

Research with an explicit focus on Industry 4.0 in the context of coal mining is still scarce. Most of the reviewed research initiatives and studies are from 2019 and 2020, which speaks in favor of the topic novelty and its growing importance. Whereas the leading economies such as Germany and China have developed the long-term strategies of utmost importance for both industry and research, Industry 4.0 initiatives in less developed economies are mostly a result of individual efforts.

Many opportunities of Industry 4.0 have been recognized in the reviewed papers. First of all, IoT and CPS have been typically seen as the basis for the application of many other concepts and technologies such as: automation, intelligent systems, big data and cloud computing. The reviewed research initiatives and studies have primarily addressed the impact of Industry 4.0 on coal mining productivity and/or safety. The double impact of some of the Industry 4.0 concepts and technologies, i.e., their impact on both productivity and safety, has been recognized. For instance, big data could be used not only for monitoring and predicting machine operation parameters in real time, but also for monitoring and predicting safety risks such as the risk of rock ejection or occurrence of hazardous gases. Likewise, the use of autonomous machines for mining of coal and transportation of soil would positively affect both productivity and safety by minimizing human presence on or above the ground.

However, the impact of Industry 4.0 on ecological and environmental protection in the coal industry has been mostly addressed in the background or neglected in the reviewed research initiatives and studies. It is realistic to expect that future studies will address this topic to a much greater extent. To this end, future studies could investigate and propose specific approaches regarding:

- the use of IoT and CPS for collecting ecology-related data such as gas emission, pollution of water and air, local hydro-meteorological data, etc.;
- the use of big data, GIS and easy-accessible software applications with modern data visualizations for monitoring ecology-related data in real time;
- the use of AI for predicting and mitigating preventable ecology risks (e.g., waste leakage, erosion, contamination of soil, groundwater and/or surface water, etc.);
- the use of AI and simulation for finding the most ecology-friendly design solution;
- the use of green technology for limiting or reversing the effects of coal mining activities;
- etc.

The reviewed papers typically have a broad research focus, i.e., address the broad concept of Industry 4.0 and its opportunities in the context of coal mining. In addition, some of the reviewed papers acknowledge the limitations in reaching the opportunities of Industry 4.0, in general, high costs, limitations of the mining environment and the need for qualified staff. However, research studies with a rather narrow research focus (e.g., an approach for intelligent and automated control of machinery, a low-cost environmental monitoring system), if considered from the wide Industry 4.0 perspective, could have substantial contribution to getting closer to achieving some of the ultimate coal mining goals (e.g., continuity of all processes, autonomous moving machinery). At the same time, creative solutions, with the constant advancement of science and technology, can help overcome at least some of the limitations. This is particularly important for economies that cannot afford substantial investments into Industry 4.0.

Taking into consideration the Industry 4.0 aims, concepts/technologies and results from different industries, future studies with a rather narrow research focus could investigate and propose:

- new or improved AI algorithms for autonomous moving machinery and robotics systems, predictive maintenance, identification and prediction of safety hazards, etc.;
- improvements in design of autonomous moving machinery (e.g., sensor processing units, object recognition units, etc.);
- improvements in design of intelligent personal assistants for interfacing with machines;
- improvements in design or application of augmented technology and wearable technologies to be used in miners' training and work environment;
- etc.

The reviewed research initiatives typically address Industry 4.0 in coal mining on a national or subnational level. No study deals with Industry 4.0 in coal mining at the supranational level. For instance, a comparative study that investigates baseline possibilities and preparedness of coal industries in different economies for Industry 4.0 would be valuable. On that basis, a comprehensive analysis and comparison of envisioned or prospective Industry 4.0 strategies would contribute considerably to the body of knowledge and practice in the given domain.

ACKNOWLEDGMENT

The research described in this paper was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] BP (2017). Energy Outlook-2017 Edition. Retrieved from https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2017.pdf

[2] H. Xie, "Symposium report of Chinese Academy of Engineering's consulting project: green exploitation and utilization of coal and

clean energy technological revolution of coal-based diverse coordination". Beijing, 2016.

[3] G. Wang, Y. Xu, and H. Ren, "Intelligent and ecological coal mining as well as clean utilization technology in China: Review and prospects", *Int J. Min Sci Technol*, vol. 29, no. 2, pp. 161-169, 2019.

[4] M. S. Hoosain, B. S. Paul, and W. Doorsamy, "Constraint dependent, shortest path algorithms for real time vehicles movement in the coal mining industry towards Industry 4.0", *In Proceedings of the International Conference on Industrial Engineering and Operations Management,* Pretoria / Johannesburg, South Africa, October 29 – November 1, 2018.

[5] T. Albanese and J. McGagh, "Future Trends in Mining", in P. Darling, Ed., SME mining engineering handbook (3rd ed.). Society for Mining Metallurgy and Exploration, 2011.

[6] M. Hermann, T. Pentek, T., and B. Otto, "Design principles for industrie 4.0 scenarios", In *2016 49th Hawaii international conference on system sciences (HICSS),* (pp. 3928-3937), 2016.

[7] M. Compare, P. Baraldi, and E. Zio, "Challenges to IoT-enabled Predictive Maintenance for Industry 4.0", *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4585-4597, 2019.

[8] M. Rylnikova, D. Radchenko, and D. Klebanov, "Intelligent mining engineering systems in the structure of Industry 4.0", In *E3S Web of Conferences* (Vol. 21, p. 01032), EDP Sciences, 2017.

[9] H. Kagermann, W. Wahlster, and J. Helbig, Eds., Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group, Frankfurt, 2013.

[10] Plattform Industrie 4.0, Neue Chancen für unsere Produktion – 17 Thesen des wissenschaftlichen Beirats der Plattform Industrie 4.0, Berlin, 2014.

[11] D. Giusto, A. Lera, G. Morabito, and L. Atzori, Eds., The Internet of Things, Springer: New York, 2010.

[12] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: A survey", *IEEE Syst J.*, vol. *9, no.* 2, pp. 350-365, 2014.

[13] D. Lucke, C. Constantinescu, and E. Westkämper, "Smart factory-a step towards the next generation of manufacturing", In Mitsuishi, M., K. Ueda, and F. Kimura, Eds., *Manufacturing systems and technologies for the new frontier,* Springer:London, 2008, pp. 115-118.

[14] Y. Hao, Y. Wu, P. G. Ranjith, K. Zhang, H. Zhang, Y. Chen, ... and P. Li, "New insights on ground control in intelligent mining with Internet of Things", *Comput Commun*, vol. *150*, pp. 788-798, 2020.

[15] R.Y. Zhong, X. Xu, E. Klotz, and S.T. Newman, "Intelligent manufacturing in the context of industry 4.0: A review", Eng., vol. 3, no. 5, pp. 616–630, 2017.

[16] E. Oztemel and S. Gursev, "Literature review of Industry 4.0 and related technologies", J. Intell Manuf, vol. *31*, no. 1, pp. 127-182, 2020.

[17] D. Pałaka, B. Paczesny, M. Gurdziel, and W. Wieloch, "Industry 4.0 in development of new technologies for underground mining", In *E3S Web of Conferences* (Vol. 174, p. 01002). The 5th International Innovative Mining Symposium. EDP Sciences, 2020.

[18] S. R. Patterson, E. Kozan, and P. Hyland, "Energy efficient scheduling of open-pit coal mine trucks", *Eur J. Oper Res*, vol. 262, no. 2, pp. 759-770, 2017.

[19] P. Darling, P. SME mining engineering handbook (3rd ed.). Society for Mining Metallurgy and Exploration., 2011.

[20] P. Westcott, G. Pitkin, and T. Aspinall, "Open-cut mining", *Australasian Coal Mining Practice*, pp. 410-458, 2009.

[21] M. Ivanovic and D. Skataric, "ICT Technologies in Optimization of Machines Movement at Open-Pit Coal Mine", *Teh. Vjesn*, vol. *26*, no. 4, pp. 1152-1158, 2019.

[22] D. Palka and T. Rizaoğlu, "The concept of hard coal mine in the perspective of Industry 4.0", *Multidisciplinary Aspects of Production Engineering*, vol. *2,* no. 1, pp. 327-335, 2019.

[23] C. Zhou, N. Damiano, B. Whisner, and M. Reyes, "Industrial Internet of Things:(IIoT) applications in underground coal mines", *Min eng*, vol. *69,* no. 12, p. 50, 2017.

[24] H. Xie, F. Gao, and Y., Ju, "Theoretical and technological conception of the fluidization mining for deep coal resources", J China Coal Soc, vol. 42, pp. 547–56, 2017.

[25] B. Sankaranarayanan, K. Rahul, J. Pradeep, S. G. Ponnambalam, and S. Saravanasankar, "Evaluation of Influence of Principles Involved in Industry 4.0 Over Coal Industries Using TISM", In Industry 4.0 and Hyper-Customized Smart Manufacturing Supply Chains, IGI Global, 2019, pp. 244-262.

[26] F. Nepsha, R. Belyaevsky, V. Efremenko, and K. Varnavskiy, "Modern Problems of Increasing Coal Mines Power Supply Efficiency", In *E3S Web of Conferences* (Vol. 105, p. 03026). EDP Sciences, 2019.

[27] L. Samorodova, L. Shut'ko, Y. Yakunina, O. Lyubimov, and P. Kovacs, "Digital ecosystem of the region and the competitiveness of Kuzbass coal enterprises". In *E3S Web of Conferences* (Vol. 134, p. 03010), EDP Sciences, 2019.

[28] E. Ghasemi, M. Ataei, and K. Shahriar, "An intelligent approach to predict pillar sizing in designing room and pillar coal mines", Int. J. Rock Mech. Min. Sci., vol. 65, pp. 86–95, 2014.

[29] Z. Jian, X. B. Li, X. Z. SHI, W. Wei, and B. B. Wu, "Predicting pillar stability for underground mine using Fisher discriminant analysis and SVM methods", Trans. Nonferr. Met. Soc. China, vol. 21, no. 2, pp. 2734–2743., 2011.

[30] Y. Hao, Y. Wu, P. Li, J. Tao, Y. Teng, and G. Hao, "Non-destructive inspection on anchorage defect of hollow grouted rock bolt using wavelet transform analysis", Eurasip J. Image Video Process, vol. 12, art. no. 146, 2018.

[31] S. Kahraman, J. Rostami, and A. Naeimipour, "Review of ground characterization by using instrumented drills for underground mining and construction", Rock Mech. Rock Eng., vol. 49, no. 2, pp. 585–602, 2016.

[32] Y.L. Tan, F.H. Yu, L. Chen, "A new approach for predicting bedding separation of roof strata in underground coalmines", Int. J. Rock Mech. Min. Sci, vol. 61, pp. 183–188, 2013.

[33] H. Kang, J. Lou, F. Gao, J. Yang, and J. Li, "A physical and numerical investigation of sudden massive roof collapse during longwall coal retreat mining", Int. J. Coal Geol, vol. 188, pp. 25–36, 2018.

[34] A. Ebrahimabadi, M. Azimipour, and A. Bahreini, Prediction of roadheaders' performance using artificial neural network approaches (MLP and KOSFM), *J. Rock Mech. Geotech. Eng*., vol. 7, no. 5, pp. 573–583, 2015.

[35] X. Wu, Y. Jiang, Z. Guan, and B. Gong, "Influence of confining pressure-dependent Young's modulus on the convergence of underground excavation", Tunn. Undergr. Space Technol, vol. 83, pp. 135–144, 2019.

[36] J. Lööw, L. Abrahamsson, J. Johansson, "Mining 4.0—The impact of new technology from a work place perspective", *Mining, Metall & Explor*, vol. 36, no.4, pp. 701-707, 2019.

[37] M. Mardonova and Y. Choi, "Toward Open-Source Hardware and Software for the Mining Industry: A Case Study of Low-Cost Environmental Monitoring System for Non-Metallic Underground Mines", *Mining, Metall & Explor*, vol. *36, no.* 4, pp. 657-674, 2019.

[38] O. Evsutin and Y. Meshcheryakov, "The Use of the Blockchain Technology and Digital Watermarking to Provide Data Authentication a Mining Enterprise", *Sensors*, vol. *20*, no. 12, p. 3443, 2020.

10th International Conference on Applied Information and Internet Technologies - AIIT 2020
October 16th, 2020, Zrenjanin, Serbia

Approach to Rapid Development of Data-Driven Applications for Smart Cities using AppSheet and Apps Script

Nenad Petrović*, Vasja Roblek**, Maša Radenković* and Valentina Nejković*
*University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia
**Faculty of Organisation Studies in Novo mesto, Ulica Talcev 3, 8000 Novo mesto, Slovenia
nenad.petrovic@elfak.ni.ac.rs, vasja.roblek@gmx.com, masa.radenkovic@elfak.rs, valentina.nejkovic@elfak.ni.ac.rs

Abstract - In this paper, we propose an approach to rapid development and delivery of smart city mobile applications based on synergy of AppSheet and Google Apps Script. Two case studies are presented: smart home energy management application and COVID-19 patient risk assessment. According to the achieved results, this approach enables many possibilities and significantly reduces required time.

I. INTRODUCTION

Lately, everything is intelligent: house, appliances, cars and even cities. Well, cities are supposed to be that way. If not before, then soon. However, what is a smart city? Officially, no definition of the term can be found when a city is involved. The smart city is a buzzword that emerged in the early 1990s and the early years of the 21st century and was considered a phenomenon of "urban labelling" [1]. According to McNeill [2] and Yigitcanlar [3], theory and practice about intelligent cities are still in development. The theory of intelligent cities is divided into two directions. The first focuses on the development of environmental technologies (e.g. energy efficiency and lower carbon emissions). The second direction focuses on the emergence of a digital urban infrastructure that enables the transformation of the classical city into the ecosystem of the smart city, which is composed of complex systems with various interactions and dependencies [4].

Moustaka, Maitis, Vakali and Anthopoulos [5] offer a sophisticated mechanism that allows us to become familiar with the D.N.A. of the smart city. A smart city can be imagined as a molecular conglomerate in which human, environmental and socio-economic factors, often unintentionally, interact with each other to produce technological solutions. In this context, smart cities development strategies focus on introducing innovative technological solutions, such as the Internet of Things (IoT), Internet of Services (IoS), artificial intelligence technologies, blockchain technologies, new sustainable materials, introducing new economic models (sharing economy, cyclical economies), as well as the development of smart processes (e-mobility, e-health, e-government, e-education, e-social inclusion) that lead to the continuous development and semantic characteristics of the smart city [6,7].

The digital components enable the development of smart data platforms, which represent an intelligent urban brain and also affect the emergence of a city as a cyber-physical social system [8]. The smart data platform (various European cities such as Amsterdam, Copenhagen, Munich and Vienna implement a smart data platform) is defined as an open, secure and city-wide platform. The platform is organized as a system for collecting, processing, analyzing, interpreting, storing and distributing the collected urban data on mobility, energy, urban life and masses. The task of the platform is to convert the raw data into value-added information. In this way, it ensures better urban planning and quality of life in urban areas. It is expected that the development will soon have an impact on the establishment of an autonomous administration, and data analysis will be based on artificial intelligence. The key goal, of course, is to inform citizens [9,10].

Data-driven mobile applications play an essential role in the information and exchange of different types of data and the quality of the urban population. The development of mobile devices applications (app) in 2010, when the first provider's Apple (Apple App Store) and Google (Google Play) came onto the market, led to the emergence of the so-called app or platform economy, which represents the sum of all economic activities, products and services such as car-sharing, automated driving, public on-call transportation, co-travel, energy, eHealth, mobility, waste, public safety, water management and others through applications to users of mobile phones and other smart devices [11-13].

Different organizations have needs for development of corresponding mobile applications in order to cover aspects of their interests within the targeted domain. However, the development of native mobile applications, due to variability of mobile hardware and operating systems can be quite expensive and time-consuming, as it requires high programming skill and expertise. Despite the emerging popularity and adoption of cross-platform mobile application development tools and environments (such as Xamarin and Flutter), they still require significant amount of time.

In this paper, we present an approach to rapid data-driven development of smart city mobile applications relying on AppSheet and Apps Script technologies, based on data stored within spreadsheets. The goal is to speed-up the development and delivery of smart city mobile applications, even by other domain experts, apart from software developers. Unlike state-of-the-art cross platform frameworks, it requires almost no coding, while it provides more advanced user interfaces, data visualization and AI-enabled capabilities without knowledge of additional frameworks or software packages at the same time.

II. Background and Related work

A. AppSheet

AppSheet [14] is an online platform acquired by Google in 2020, which enables easy creation and distribution of mobile, tablet and web applications starting from cloud data sources, such as spreadsheets and databases without any coding. It mainly targets business use cases, such as CRM, project management and personalized reports. AppSheet analyzes the structure of provided data sources and automatically generates the views that can be shown within the application. The users are able to customize the generated views by showing or hiding particular columns or write formulas for data processing and aggregation. AppSheet is free for prototyping and personal use, while monthly fee has to be paid for commercial applications. Moreover, AppSheet offers advanced machine learning and AI features such as value prediction, optical character recognition (OCR), sentiment analysis and anomaly detection. However, active internet connection and client app are required in order to access AppSheet applications and their features, as they are deployed in cloud.

AppSheet-based mobile applications have been approved as effective solution across many fields. For example, in [15], it was used for health facility survey mobile application, while [16] presents waste monitoring and management use case.

B. Google Apps Script

Google Apps Script [17] is a cloud platform for rapid application development that enables easy integration with G Suite (Gmail, Google Docs, Google Forms, Google Sheets, Google Drive and others). Apps Script relies on JavaScript language with built-in libraries for G Suite. Google provides editor run in web browser, while the scripts themselves are run on Google's servers. Apps Script can be used for adding custom menus, dialogs, and sidebars to Google Docs, Sheets and Forms. Apart from that, the extensions and add-ons for these services can be also developed. Moreover, it is possible to perform interaction and coordination of various services (such as AdSense and Maps). Apps Script makes it easy to develop and deliver add-ons for G Suite to large numbers of users, whether publicly, worldwide or group of private users.

There are many interesting existing Apps Script use cases presented in the existing scientific publications. In [18], Apps Script was used for digital literacy student assessment tasks implemented inside Google Docs. Furthermore, in [19], it was used with Android app and Google Sheets for generation of light vehicle diagnostic reports sent to the owner's email. A similar approach was used in [20] for real-time environmental data logging using Apps Script with Google Sheets.

In this paper, we leverage Apps Script capabilities related to Google Sheets that are further used by AppSheet for smart city mobile application development. More accurately, the triggers for Google Sheets are written, which are activated when certain types of events occur – such as data insertion or update, as shown in Fig. 1.

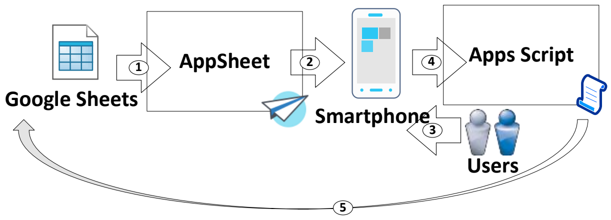


Figure 1. AppSheet and Apps Script synergy for data-driven mobile applications: 1-Spreadsheet data 2-Generated mobile app based on spreadsheet data 3-User input 4-Trigger activation 5-Spreadsheet updates on triggers

III. Case Studies

A. Smart Home Energy Management

The aim of this mobile application is to enable easy and convenient household consumer device monitoring and energy management, targeting the so-called smart grid prosumers (which both produce and consume energy). This case study builds upon the works presented in [21-23], denoted as energy management framework (E-M). In Fig. 2, overview of the smart home energy management application is given.

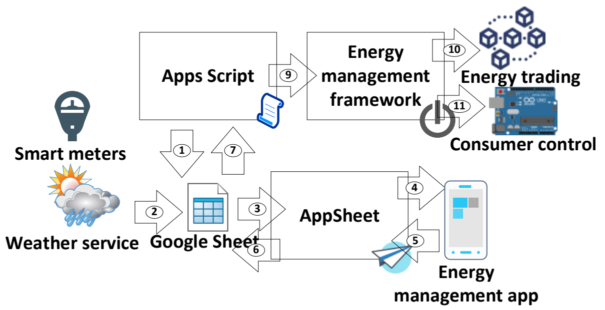


Figure 2. Smart home energy management mobile app: 1-Apps Script trigger for data acquisition 2-Weather and consumption data 3-Collected data 4-Visualization 5-User input 6-Deal propositions and turn on/off commands 7-Apps Script trigger activation 8-Calling energy management framework on the backend 9-Smart contract 10-Household device control

First, the energy consumption values are collected by IoT devices which act as smart meters [21] and inserted into spreadsheet. The collected measurements are further forwarded to the AppSheet's prediction module in order to make load forecast within the household, based on energy consumption history.

Both the collected measurements and prediction results are used to render views shown to the users of the mobile application, providing them overview of current consumption monitoring and future predictions. Apart from collected electric measurements, weather data (temperature and rainfall) is retrieved from online service in order to provide more accurate predictions [21]. Moreover, there is also an option within monitoring screen to turn off the consumer device remotely. For that task, Arduino Uno equipped with actuator devices (such as relay) are used, as described in [23]. In Fig. 3, the screenshots of AppSheet mobile application taken on iOS devices are given.

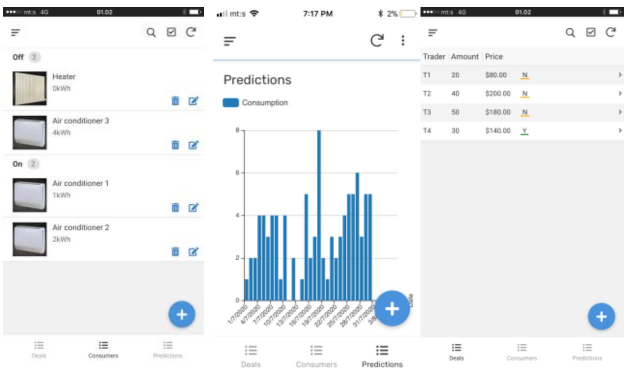


Figure 3. Consumer device monitoring and management: (a) device list (b) consumption prediction (c) trading offers

Periodically, the script is sending the consumption data and predictions to the energy management framework backend. In the background, it performs linear optimization process with respect to program from [21] and generates recommendations – either to perform energy buying and selling. Following the recommendations, the app users can make deal propositions in order to buy or sell energy at a given price range. For energy trading, the blockchain-based transactions are executed relying on [21] and verified using [22]. In this step, the App Script code is responsible for passing the parameters necessary for smart contract generation, to the energy management framework, using the deal propositions parameters stored within spreadsheets. In Listing 1, the described energy trading process is given as pseudocode.

Input: energy needed, user id
Output: energy transaction
Steps:
1. AppSheet: consumption:=PredictConsumption(user id);
2. Apps Script Trigger: energy needed=consumption-available;
3. Apps Script Trigger: Send(energy needed, user id, to: E-M);
4. E-M: trader_list:=Optimization(energy needed, user id);
5. Apps Script Trigger: Receive(trader_list, from: E-M);
6. Apps Script Trigger: UpdateGoogleSheet(trader_list, user id);
7. AppSheet: UpdateView();
8. User: trader:=Take trader from list;
9. Apps Script Trigger: SendOffer(trader);
10. Trader: Accept or reject deal offer;
11. Apps Script Trigger: if(accept) then
12.   E-M: ExecuteTransaction(trader, energy needed);
13. Endif;
14. End.

Listing 1. Energy trading leveraging mobile app and Apps Script trigger for coordination with energy trading framework

B. *COVID-19 Patient Risk Assessment*

This application enables quick COVID-19 (potential) patient risk assessment based on user-provided health state information (such as body temperature, symptoms, chronic diseases). The application relies on semantic fuzzy reasoning framework (S-F) from [24] for patient risk assessment which collects textual data about COVID-19 from publicly available online scientific publications and transforms them into facts stored within the semantic knowledge base. In Fig. 4, an overview of the second case study application is given.

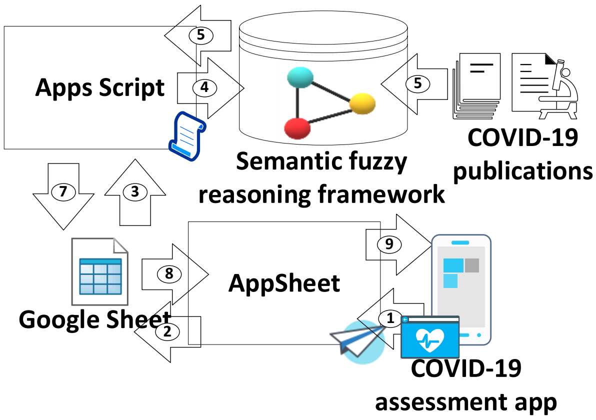


Figure 4. COVID-19 patient assessment app: 1-Health state, symptoms and medication information 2-Spreadsheet update 3-Apps Script trigger 4-Calling assessment framework 5-Retrieveing coronavirus-related literature 6-Assessment results 7-Spreasheet update as trigger 8-AppSheet application update 9-Result visualization

In Fig. 5, the two screenshots of the application taken on iOS devices are given.

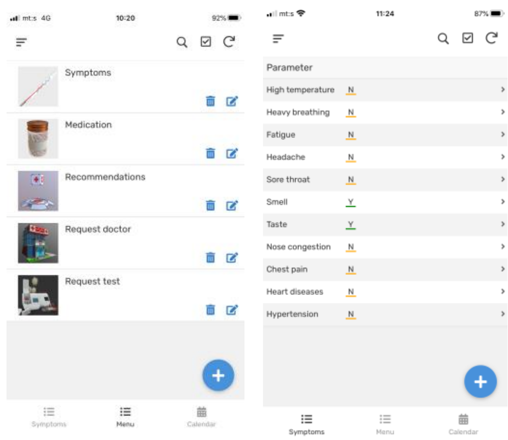


Figure 5. Consumer device monitoring and management: (a) device list (b) consumption prediction (c) trading offers

In what follows, the patient assessment procedure is described. First, the patient enters the information about symptoms (such as fever, cough) and general health information (chronic diseases, age). Once the information is provided, the Apps Script Trigger is activated upon data insertion into Google Sheet table. It forwards the symptom list and health state data to the semantic-fuzzy (S-F) module, which then constructs semantic knowledge base containing the facts about the patient’s symptoms and health state leveraging public online literature. After that, the reasoning is performed by matching the health state and symptoms with assessment, such as “high risk”, “the patient should visit doctor” or “request a COVID-19 test”. Once the results become available, they are sent back to the Apps Script Trigger which updates the Google Sheet shown to the user after rendering by AppSheet. In Table 3, the pseudocode of the underlying patient assessment mechanism is given.

In Listing 2, the pseudocode of the underlying patient assessment mechanism is given.

Input: symptom list
Output: state assessment
Steps:
1. User: Provide health state input;
2. Apps Script Trigger: Send(symptom list, to: S-F)
3. S-F: publications:=GetPublicationsForKeyword(symptom_list);
4. S-F: knowledge_base:=ExtractKnowledge(publications);
5. S-F: assessment:=Reasoning(symptom list, knowledge_base);
6. Apps Script Trigger: Receive(assessment, from: S-F);
7. Apps Script Trigger: UpdateGoogleSheet(assessment);
8. App Sheet: UpdateView();
9. End.

Listing 2. Automated patient assessment algorithm leveraging Apps Script and semantic knowledge base

IV. EVALUATION AND RESULTS

For purpose of evaluation the following devices were used: a laptop equipped with Intel i7 7700-HQ quad-core CPU running at 2.80GHz with 16GB of DDR4 RAM and 1TB HDD acting as Edge server deploying the external services (energy management and semantic-fuzzy), iPhone 6s Plus (dual-core, 2GB of RAM) as representative of iOS smartphones and Honor 9 Lite (octa-core, 3GB of RAM) which belongs to Android smartphones. As active internet connection was required to run the apps, 4G network with 50Mbps download/upload speed was used.

Several aspects of evaluation were taken into account: application development time speed-up (compared to equivalent cross-platform application development time), application delivery time (how much time is needed to deliver the modified application without re-installing), storage required for AppSheet client (installation package and installed app), RAM memory utilization and average prediction accuracy of AppSheet’s predictor in case of energy demand prediction. When it comes to capabilities relying on Apps Script and external services, the overall response time for energy trading (*Case study A*) and semantic knowledge base construction (*Case study B*). In Table I, the results achieved during the evaluation are given.

According to the achieved results, the proposed approach of mobile application generation using AppSheet taking Google Sheet as an input significantly speeds up the mobile application development compared to cross-platform solutions. For comparison, equivalent Xamarin apps were developed. Considering the delivery time, it was shorter in case of Android device, as it is equipped with more powerful CPU and larger RAM, but AppSheet client is larger in case of iOS, but still significantly faster than native and cross-platform apps. Moreover, smart home management occupies more RAM than the COVID-19 assessment application, as it includes more advanced data visualization capabilities and leverages value prediction. Furthermore, the consumption prediction model in the first case study achieved accuracy around 89% for day-ahead forecast, which is just slightly worse than [21]. However, setting up the prediction module in this paper relied on easy-to-use AppSheet’s user interface taking just several minutes, while in [21], a custom deep learning module based on TensorFlow in Python was developed, which required much more effort and time.

TABLE I. EVALUATION RESULTS

Aspect	Value	
	Case A	***Case B***
Development speed-up [times]	12 times	8 times
Device type	iPhone 6s Plus	Honor 9 Lite
Delivery time [s]	4.36	3.59
Client app size [MB]	iOS	Android
	.ipa:59 total:96	.apk:13 total:77
RAM utilization [MB]	258	163
Prediction accuracy [%]	89%	Not applied
Service response [s]	2.51	4.91 (cold start: 414)

V. CONCLUSION AND FUTURE WORK

The proposed approach of data-driven smart city mobile applications seems promising, as it enables domain experts to create mobile applications for their needs, without advanced programming knowledge. In our case studies, it accelerates the development and delivery of mobile applications significantly - around 12 times for the first case study, while it was around 8 for the second (compared to Xamarin cross-platform framework, which still requires solid programing skill). Furthermore, the consumption prediction model in the first case study achieved satisfiable accuracy. The applications were run successfully on both Android (Redmi Note 7, Honor 9 Lite) and iOS devices (iPhone 6s Plus and SE).

In future, it is planned to adopt the same approach on mobile applications for energy-aware fog, cloud and hybrid cloud computing infrastructure management scenarios, building upon the work from [25].

ACKNOWLEDGMENT

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

[1] N. Komninos, "Intelligent cities: Variable geometries of spatial intelligence," *Int. Build. Inter*., vol.3, no. 3, pp. 172-188, 2011.

[2] D. McNeill, "Global firms and smart technologies: I.B.M. and the reduction of cities," Trans. of the Inst. of Brit. Geograph., vol 40, no. 4, pp. 562-574, 2015.

[3] T. Yigitcanlar, "Smart cities: an effective urban development and management model?", Aust. Plan., vol. 52, no. 1, pp. 27-34, 2015.

[4] V. Roblek, "The smart city of Vienna", in Smart City Emergence: Cases from Around the World, Amsterdam, NL: Elsevier, 2019, 5, pp. 105-127.

[5] V. Moustaka, A. Maitis, A. Vakaland, & G. L. Anthopoulos, L. G. "CityDNA Dynamics: A Model for Smart City Maturity and Performance Benchmarking," Companion Proceedings of the Web Conference 2020, Taipei, Taiwan, pp. 829-833, April, 2020.

[6] H. Han, & S. Hawken, "Introduction: Innovation and identity in next-generation smart cities," *City, Cult. and Soc*., vol. 12, pp.1-4, 2018.

[7] M.S., Sepasgozar, S. Hawken, S. Sargolzaei, & M. Foroozanfa, "Implementing citizen centric technology in developing smart cities: A model for predicting the acceptance of urban technologies", Tech. For. and Soc. Ch., vol. 142, pp.105-116, 2019.

[8] G.C. Cassandras, "Smart cities as cyber-physical social systems," Eng., vol. 2, no. 2, pp. 156-158, 2016.

[9] I. Cordón, J. Luengo, S. García, F. Herrera, & F. Charte, “Smartdata: Data preprocessing to achieve smart data in r,” *Neurocomp*., Vol. 360, pp. 1-13, 2019.

[10] V. Roblek, & M. Meško, "Smart city knowledge management: Holistic review and the analysis of the urban knowledge management," The 21st Annual International Conference on Digital Government Research, Seoul, South Korea, pp. 52-60, June, 2020.

[11] M. Kenney, & J. Zysman, "The platform economy: restructuring the space of capitalist accumulation," *Camb. J. of Reg., Econ. and Soc*., vol. 13, no. 1, March, 2020.

[12] C. Troise, E. Ferrara, M. Tani, & O. Papaluca, "Perspectives of the App Economy: Tenets of the Innovative Phenomenon," *Int. Bus. Res*., vol. 13, no. 3, pp. 1-14, 2020.

[13] V. Roblek, V. Dimovski, & M. Meško, "The complexity of the Age Management in the Framework of Smart Urban Environment," 3rd American University in the Emirates International Research Confer-ence, Dubai, U.A.E., pp.1-9, August, 2020.

[14] Google Apps Script [online]. Available on: https://developers.google.com/apps-script, last accessed: 23/09/2020.

[15] P. Sylim, C. C. Santos-Acuin, "Development of a Low-Cost Electronic Data Collection Tool for a Health Facility Survey: Lessons Learned in the Field", Journal of the International Society for Telemedicine and EHealth, 4, e27, pp. 1-8, 2016.

[16] S. K. Arumugam, R. Muhamad, K. Yahya, "Mapping of construction waste for eco-costs per value ratio (EVR) index using Google My Maps in Shah Alam, Malaysia", ICONBUILD & RCCE 2019, pp. 1-9, 2019. https://doi.org/10.1088/1757-899X/849/1/012046

[17] AppSheet [online]. Available on: https://www.appsheet.com/, last accessed: 23/09/2020.

[18] L. Siiman, M. Mäeots, "Assessing Students’ Digital Literacy with Interactive Computer-Based Tasks Created in Google Apps Script", EDULEARN ’19, pp. 10265-10271, 2019. https://doi.org/10.21125/edulearn.2019.2578

[19] P. Rajput, R. Parkeh, "On-Board Diagnostics based remote emission test for Light Motor Vehicles", CONECCT 2020, pp. 1-6, 2020.

[20] T. DeBell et al., "OPEnS Hub: Real-Time Data Logging, Connecting Field Sensors to Google Sheets", Frontiers in Earth Science, May 2019 volume 7, pp. 1-6, 2019.

[21] N. Petrović, Đ. Kocić, "Data-driven Framework for Energy-Efficient Smart Cities", Serbian Journal of Electrical Engineering, Vol. 17, No. 1, Feb. 2020, pp. 41-63. https://doi.org/10.2298/SJEE2001041P

[22] N. Petrović, M. Tošić, "Semantic Approach to Smart Contract Verification", Facta Universitatis Series Automatic Control and Robotics 19(1), pp. 21-37, 2020. https://doi.org/10.22190/FUACR2001021P

[23] N. Petrović, Đ. Kocić, "Sistem za relejnu zaštitu zasnovan na Arduino Uno u okviru podacima-vođene arhitekture pametne mreže", IEEESTEC – 12th Student Projects Conference, pp. 331-334, 2019.

[24] N. Petrović, "Tackling the COVID-19 Conspiracies: The Data-Driven Approach", ICEST 2020, pp. 1-4, 2020.

[25] N. Petrovic, M. Tosic, "SMADA-Fog: Semantic model driven approach to deployment and adaptivity in Fog Computing", Simulation Modelling Practice and Theory, 102033, pp. 1-25, 2019. https://doi.org/10.1016/j.simpat.2019.102033

Convolution neural network models to detect COVID-19

Shaymaa Adnan Abdulrahman[1], Abdel-Badeeh M. Salem[2]
[1]Department of computer Engineering, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq
PhD Student at Ain Shams University, Egypt
[2]Department of Computer Science ,Ain Shams University, Cairo , Egypt
Shaymaaa416@gmail.com , abSalem@cis.asu.edu.eg

Abstract - COVID-19 is considered dangerous at the present time Therefore, it is necessary to identify people with this disease to limit its spread through preventive measures in health centers and hospital . There are three different convolutional neural network(CNN) based models namely; ResNet(50), Inception(V3) and Inception-ResNet(V2) was suggested in this study to detect people infected with emerging COVID-19 disease . Confusion matrices by these three models are given and analyzed when used(3-fold) cross validation. The results that we obtained through the three proposed models indicate ResNet50 model provides the highest classification performance with 95% accuracy while 97% accuracy for InceptionV3 and 87% accuracy for Inception-ResNetV2).

I . INTRODUCTION

COVID-19 was declared by world health organization as a dangerous and raid spread virus .Formerly COVID-19 was called SARS-CoV-2 [1] [2]. The emergence of this virus was in the Chinese city of Wuhan unite it spread to all countries of the world. COVID-19 source is not known yet, but it is a new virus and an evolving virus and is considered one of the viruses that transmits between humans and animals [3] .Fever, fatigue, cough, shortness of breath, headache, diarrhea are all symptoms of people with COVID-19 disease. The method of spreading the disease among people is by means of respiratory drops in , in addition to human contact . For this reason, home quarantine and reduced human contact and early detection of disease limit the speed of the viruses spread . Models showed advanced intelligence such as artificial intelligence (AI) and machine learning have the potential to predict the spread of the epidemic or diagnose infectious diseases through the use of (data / information) and analysis such as the use chest X-rays of COVID-19 patient. These techniques contribute to identifying confirmed cases of the injury. Moreover, clinical examination, medical diagnosis and Reverse transcription polymerase chain reaction (RT-PCR). Reverse transcription polymerase chain reaction indicates it blood sample of COVID -19 [4] .Currently, the (RT-PCR)test today is the actual and proven test for people with emerging. In addition,(point-of-care) test Immunoglobulin M (IgM), and detection of antibodies can be used to detecting infected case [5] .Since the (RT-PCR) test is important and necessary , it requires a CDC guide's sample collection qualified , microbiology expertise time from (4)hrs up to (6)hrs . Therefore, medical imaging is very necessary candidate in screening for the COVID-19 cases. Radiologist's diagnosis involves computed tomography (CT) scans, (chest X-ray) radiographs . COVID-19 symptoms can be effectively detected using(CT) or by X-ray images. Based on the chest (CT)scans During recovery radiologists can detect the (COVID-19) pneumonia and the stages of patient recovery or deterioration. Automated artificial intelligence models, can accurately provide early detection for the diagnosis of the case of COVID-19 through detecting the early, lung damage signs in the image[6] [7].

The main objective of this study , applied deep convolutional neural network (CNN)for the classification process of (COVID-19) Chest X-ray images to normal and (COVID-19) classes.

The structure of the paper is organized as follows. Section 2: Proposed work , Data-set , Pre-processing, Segmentation , Deep Machine Learning Algorithms, Convolutional Neural Networks (CNNs), Deep Transfer Learning , Section 3: Conclusion.

II . PROPOSED MODEL

A. Data-set

There are two data set applied in this study such as(chest X-ray image dataset taken from kaggle repository ("www.kaggle.com [Accessed 16th March. 2020]").[8] All medical images are for healthy people without pneumonia, and other dataset is A COVID-19 X-ray image) . This is the dataset was developed by Cohen JP [9], by using images from various open access sources . Usually dataset contain (125) X- ray images diagnosed with COVID-19. please see: (https://github.com/muhammedtalo/COVID-19). Table 1 refer to Clinical symptoms of (COVID-19) patients and.

TABLE 1 : CLINICAL SYMPTOMS OF COVID-19

Feature	Total Patients
Male	43
Female	82
average age of these subjects	approximately 55 years.

There are several stages in the proposed model for the purpose of detecting cases infected with COVID-19 . The following figure 1 represents the suggested stages .

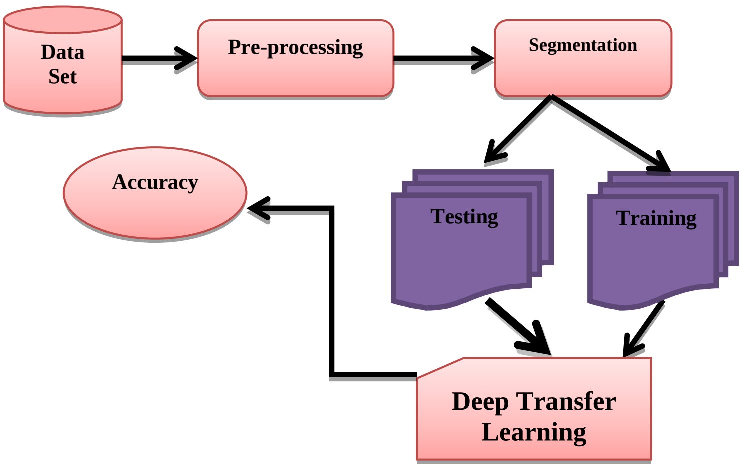


Figure 1 : proposed model

B. Pre-processing

Usually data collected tend to be not wholly completes such as (data irregularity , noisy , detection missing of data , prevent the errors , conflicting,) and decrease the data to be analyzed would lead to massive payouts, for decision making . The machine learning and analytics, the selection of features is refer to the process of selecting a subset of features like(variables, predictors) to use in model building. Component selection approaches used for many reasons (H, Witten &EibeFrank . [10] .

- Simplifying the models for easier the researchers to understand
- The training process is shorter
- Reducing the status of curse dimensions
- Improvement generalization by using minimizing over fitting.

To removal of noise from A COVID-19 X-ray image is an important step, in the preprocessing. There are different types of filters that are used are this stage such as (Gaussian filter , .MPEG7 Histogram Filter , Gabor Image Filter , Pyramid of Rotation-Invariant Local Binary Pattern Histograms Image Filter and Fuzzy 64-bin Histogram Image Filter , Anisotropic Diffusion Filter, ….ect) .In this step used (Gabor Image Filter) : is refer to texture analysis. which implies that it mainly analyzes if there is some different frequency data in the picture, in specific directions across the point or area of, analysis in a regional region. Gabor Image Filter can define by a sinusoidal wave (a plane wave 2-D Gabor filters) multiplied by a (Gaussian function) . A set of Gabor filters with different frequencies and orientations may assist in extracting useful features from images [11] .In the discrete domain the(2 -dimensional Gabor filters) are given by(1) (2).

$$G_c[i,j] = \; Be\;2\sigma^2 \quad \cos(2\pi f\,(i(\cos\theta + j\sin\theta)) \ldots\ldots\ldots(1)$$

$$G_s[i,j] = \; Ce\;\frac{(i^2+j^2)}{2\sigma^2} \quad \sin(2\pi f\,(i(\cos\theta + j\sin\theta)) \ldots\ldots(2)$$

C. Segmentation

Segmentation process is a hot topic due to its wide range of applications . There are a large number of researchers have used the segmentation process in their research in order to facilitate dealing with the available data , especially medical images . There are a number of techniques used in the segmentation process such as (Histogram Bias Correction, Global Thresholding , Multi-Level Thresholding , Edge Histogram Descriptor ,Edge Detection…….etc) . The segmentation Edge Detection was applied in our search .

Edge Detection There is another type of Segmentation called (Segmentation Edge Detection) . This type of segmentation has been applied in our research .The edge of the object is in the form of discontinuous local features of the images that is the most significant part of the images changes in local brightness. For example (texture changes , gray value of the mutation and color mutation . and so on[12] .The use of discontinuities to detect the edge; so as to achieve the purpose of images segmentation. There is always a gray edge between (two) adjacent regions with different gray values, in the images . , and there is a case where the gray value is not continuous. This discontinuity can often be detected using derivative operations, and derivatives can be calculated using differential operators[16] . While Parallel edge Detection (PeD) is often done through means of a spatial domain differential operator to perform images segmentation by convoluting its template and images. Usually Parallel edge detection.(PeD) is applied as a method of image preprocessing. The widely (first-order) differential operators are Prewitt operator, Roberts operator and .Sobel operator [13] .While (second-order) differential operator has nonlinear operators such as Laplacian, Kirsch operator and Wallis operator.

. Sobel operator (SO) is mainly applied for edge detection also it is technically a discrete differential operator applied to calculate the approximation of the gradient of the images luminance functions. The (SO) is edge detection operator based on the first derivative. As a result of the operator in the introduction of a similar local average operation; so the noise has a smooth effect also can effectively eliminate the impact of noises . The influence of the Sobel operator on the position of the pixel is weighted which is better than the Prewitt operator and the Roberts operator. Usually Sobel operator contain two sets of (3x3) matrices, which are transverse and longitudinal templates . and are plotted with the images plane, respectively, to obtain the difference between, the horizontal and the longitudinal difference. In actual use the following ,two templates are used to detect the edges of the image.

$$GX=\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}$$

Detect horizontal edge (transverse template)

$$Gy=\begin{bmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{bmatrix}$$

Detect vertical edge (longitudinal template)

The horizontal (H)and vertical (V) gradient approximations of each pixel of the images can be combined to calculate the size of the gradient by using the following formula:

$$G=\sqrt{G_I^{2}+G_Y^{2}} \text{ ....(3)}$$

The gradient can then be calculated by using the next equation

$$\Theta = \arctan\left(\frac{G_Y}{G_I}\right) \text{.........(4)}$$

. **Laplace operator** (LO) indicate an isotropic operator, second -order differential operator . and refer to more appropriate when it is only concerned with the position of the edge regardless of the pixel gray scale difference around it[24]. One of the most important feature of Laplacian Operator(LO), is response to isolated pixels is stronger than the edge or line, and therefore applies only to noise-free images .[14] The presence of noise in the images lead to this that Laplacian Operator need to perform low pass filtering before detecting the edge . Therefore; the usual segmentations algorithms combines the Laplacian operator with the smoothing operator to generate a new templates ..Laplacian operator(LO) is the simplest isotropic differential operator ,with rotational invariance . Laplacian Operator (LO), is more suitable when using with digital image processing

III. DEEP MACHINE LEARNING ALGORITHI

Convolutional Neural Networks (CNNs) This type of CNN network is widely used due to the analysis and interpretation of images , especially medical images. Where this network is extracting the spatial and temporal features in an images Usually the CNN contain number of layers have a weight-sharing technique which, helps in reducing computation efforts [15] [16]. Usually the CNN contain number of layers, have a weight sharing technique which helps in reducing computation efforts. CNN architecture is consists of several blocks of input Layer, Hidden Layer (convolutional layer, pooling layer), fully connected layer, and Output Layer. However, for selecting an anchor, the anchors are divided into two categories (positive / negative). Built deep convolutional neural network (CNN) for the classification process of (COVID-19) Chest X-ray images to normal and (COVID-19) classes based on dataset taken from [8][9] . Used CNN based on (ResNet (50)and Inception(V3) , and Inception-ResNet(V2) models. Figure 2 shows the structure of ResNet-50.

.

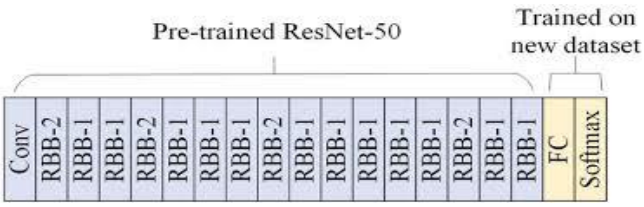


Figure 2: Structure of (ResNet-50)

Usually Residual building block. (RBB) refer to the most vital element in ResNet-(50). Usually RBB is based on the idea of .skipping block of convolutional layers ,by using shortcut connections. These shortcuts are important for optimizing trainable parameter in error back - propagation to avoid the vanishing -exploding gradients problems which, can help to construct deeper (CNN) structure to improve, final performance for fault diagnosis. Figure 3 refer to (CNN) including pre-trained ResNet (50), Inception(V3) and Inception ResNet(V2) models for the prediction of (COVID-19) patients and normal .

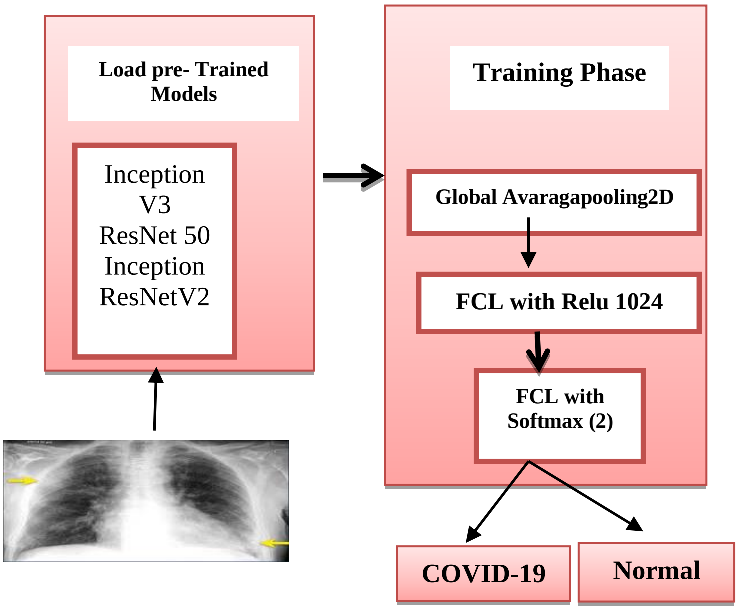


Figure 3 : Representation of pre-trained models for the prediction of COVID-19 patients and normal

Residual neural (ResNet) method is an improved version of convolutional neural networks . (ResNet) adds shortcuts between layers to solve a problem. Thanks to this, it prevents the distortion that occurs as the network gets deeper and, more complex. Also , bottleneck blocks are applied to make training faster in the (ResNet) model [17]. ResNet (50) is a (50) layers network trained on the X-ray dataset . Inception(V3) is a kind of convolutional neural network (CNN) models . Usually consists of numerous convolution and maximum pooling step . While the last stage, it contains a fully connected neural networks [18]. As with the ResNet(50) models, the network is trained with X-Ray medical data-set.

IV. Result & Analysis

CNN models (ResNet50, Inception(V3) and Inception ResNet(V2) were pre-trained with random initialization weights using the Adam optimizer. The dataset used was randomly split into two independent datasets with 70% and 30% for train and test respectively.
There are four criteria were used for the performances of deep transfer learning methods such as [34].

Accuracy = (TN + TP) / (TN + TP + FN + FP)……………(1)

Recall = TP / (TP + FN) ………………………………..…(2)

Specificity = TN / (TN + FP) ……………………….…….(3)

Precision = TP / (TP + FP)……………………….……..(4)

TP, FP, TN and FN given in Equation (1) – (4) represent the number of True Positive., False Positive,. True Negative and False Negative., respectively[35]
Pre-trained methods such as ResNet(50), Inception(V3) and Inception ResNet(V2) have been trained and tested on chest (X-ray) images. Training accuracy and loss values for fold-3 of the pre-trained models are given in Figure (4) and Figure (5), respectively. The training stage has been carried out up to 5th epoch to avoid overfitting for apre-trained methods . It can be seen from Fig 4 that the highest training accuracy is obtained with the ResNet(50) model . Inception(V3) and Inception-ResNet(V2) models have similar performance. However, it is seen that ResNet(50) shows a fast training process than other models
The pre-trained models give very high initial values; the initial values are below 75% due to the low number of data. The training loss values of ResNet(50) Inception(V3) and Inception ResNet(V2) are shown in Fig 5. When the loss figure are analyzed; it is seen that the loss values decrease in (3) pre-trained models during the training stage. It can be said that, the ResNet(50) model both decreases loss values faster and approaches zero.

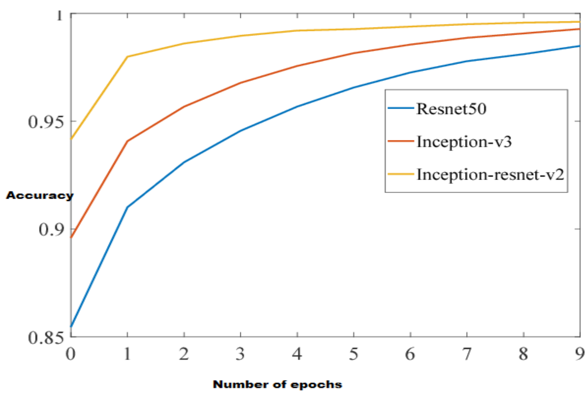


Figure 4. The performance of three pre-trained models (Training accuracy for fold-3)

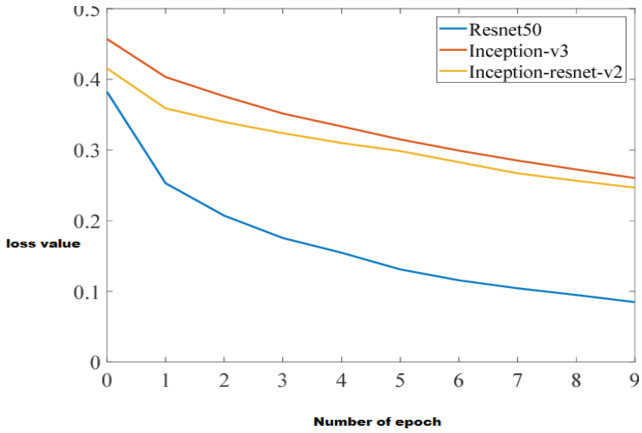


Figure 5. The performance of three pre-trained models (loss value)

There are more test that were applied in this study , accuracy, Specificity , Precision, Recall was tested . To calculate the mentioned scores . Confusion matrices for all the architectures were obtained Figure 6 : refer to confusion matrices of (COVID-19)and normal test result of the models are given.
First: Inception(V3) pre-trained model classified (10) of the COVID-19 as True Positive for fold- (3)and classified 10 of the normal as True Negative. **Second** : ResNet(50) model also classified 10 of the COVID-19 as True Positive for fold-3 and classified 10 of the normal as True Negative. **Third**: Inception ResNetV2 classified 10 of the COVID-19 as True Positive for fold-3) and classified 9) of the normal as True Negative.

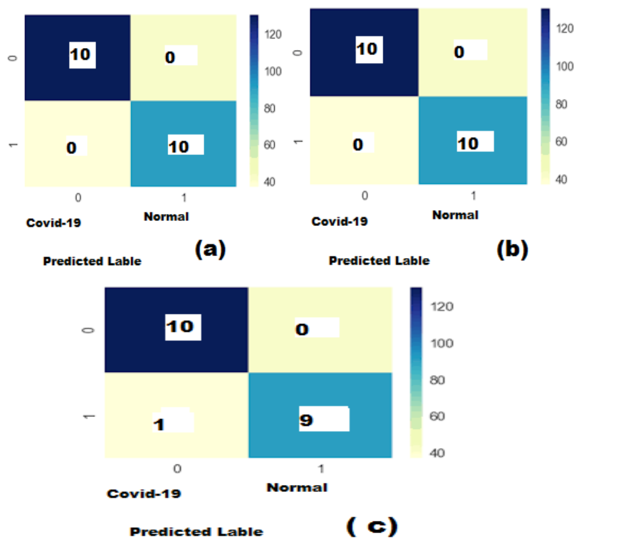


Figure 6 : confusion matrices of (COVID-19)and normal with fold k=3 results: (a) Inception (V3) (b) ResNet50 (c) Inception-ResNetV2

In addition besides the confusion matrix (CM), receiver operating characteristic curve (ROC) plots and areas for one model is given Shown fig 7. Inception(V3) and ResNet(50) pre-trained models appear to be very high

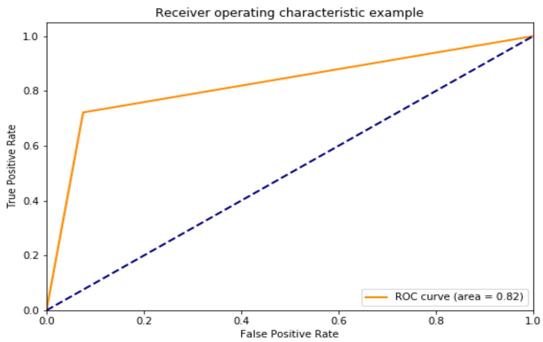


Fig 7: ORC Curve in case ResNet(50), Inception(V3) and Inception ResNet(V2)

The comparisons of three models by using the test data are shown in Table 2 . We have obtained the best performance as an accuracy of (95%) , recall of (94%,) and specificity value of (100%) for ResNet(50) pre-trained model. The lowest performance values have been yielded an accuracy of(81%), recall of (79%), and specificity value of(88%) for Inception-ResNet(V2).As a result, the ResNet(50) model provides superiority over the other two models both training and testing stage.

Table 2 : Comparison of the proposed method

		TP	TN	FP	FN	ACC	Rec	Spe	Pre
InceptionV3	Fold-1	8	10	0	5	80	73	100	100
	Fold-2	10	10	0	0	100	100	100	100
	Fold-3	10	10	0	0	100	100	100	100
	Mean					93	91	100	100

		TP	TN	FP	FN	ACC	Rec	Spe	Pre
ResNet(50)	Fold-1	8	10	0	4	80	73	100	100
	Fold-2	10	10	0	0	100	100	100	100
	Fold-3	10	10	0	0	100	100	100	100
	Mean					95	94	100	100

		TP	TN	FP	FN	ACC	Rec	Spe	Pre
InceptionResNetV2	Fold-1	7	6	2	2	75	88	70	70
	Fold-2	10	8	2	0	92	100	88	89
	Fold-3	8	10	1	0	92	100	88	89
	Mean					81	79	88	87

The main advantages of our study can be summarized as follows:

Step 1: Two dataset : first , Chest (X-ray) image have been used in the study when obtained from kaggle repository for healthy people , second , dataset is A COVID-19 X-ray image was developed by [18] . (https://github.com/muhammedtalo/COVID-19)

Step2 :pre-processing when used Gabor Image Filter to remove noise .

Step 3 : In this step used Segmentation Edge Detection used as Segmentation process

Step 4: Used deep convolutional neural network (CNN)for the classification process of (COVID-19) Chest X-ray images to normal and (COVID-19) classes .

These result from three models obtained in our study will help the health institution and doctor to make a more accurate decision . Finally it is possible to use second methods in the segmentation process also use additional filter for the purpose of giving better result

V. Conclusion

Early detection of people with corona disease is very important and necessary to prevent its spread among people . In this study, we proposed a deep transfer learning based approach, using chest (X-ray) images obtained from (COVID-19) patients and normal to predict ,(COVID-19) patients automatically

the results show that the ResNet(50) pre-trained model yielded the highest accuracy of 95% among the three models. In addition, four criteria were used for the performances of deep transfer learning methods and evaluate proposed work such as (Accuracy, Recall, Specificity , Precision).. Considering the performance results obtained, it is seen that the pre-trained ResNet50 model provides the highest classification performance with 95% accuracy among other two proposed models 93% accuracy for InceptionV3 and 81% accuracy for Inception-ResNetV2).

References

[1]http://www.nhc.gov.cn/jkj/s7916/202001/44a3b8245e8049d2837a4f27529cd386.shtml. Accessed March 18 , (2020).

[2] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet.395(2020):497-506.

[3] Shaymaa Adnan Abdulrahman, WaelKhalifa, MohamedRoushdy, Abdel-Badeeh M. Salem" Comparative Study for 8 Computational Intelligence Algorithms for Human Identification" Journal , Computer Science Review Vol 36 2020 https://doi.org/10.1016/j.cosrev.2020.100237

[4] H. H. Elmousalami and A. E. Hassanien, "Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations," arXiv preprint arXiv:2003.07778, 2020.

[5] Fang, Yicheng, et al. "Sensitivity of chest CT for COVID-19: comparison to RT-PCR." Radiology (2020): 200432.

[6] C. Molnar, Interpretable machine learning: Lulu. com, 2019.

[7] Lamia Nabil Mahdy, Kadry Ali Ezzat1, Haytham H. Elmousalami , Hassan Aboul Ella3, Aboul Ella Hassanien , " Automatic X-ray COVID-19 Lung Image Classification System based on Multi-Level Thresholding and Support Vector Machine"medRxiv ,2020 https://doi.org/10.1101/2020.03.30.20047787

[8] www.kaggle.com [Accessed:16th March 2020]

[9] Cohen JP (2020) COVID-19 image data collection. https://github.com/ieee8023/COVID-chestxray-dataset.

[10]Shaymaa Adnan Abdulrahman, Abdel-Badeeh M. ," A efficient deep belief network for Detection of Corona Virus Disease COVID-19" Vol. 02, No. 01, PP. 05-13, 2020 , DOI: 10.5281/zenodo.3931877

[11] Li Chaorong , Huang Yuanyuan , ZhuLihong " Color texture image retrieval based on Gaussian copula models of Gabor wavelets, Pattern Recognition , vol 64, pp 118-129, 2017

[12] Davis L S. A survey of edge detection techniques[J]. Computer graphics and image processing, 1975, 4(3): 248-270.

[13] Kundu M K, Pal S K. Thresholding for edge detection using human psychovisual phenomena[J]. Pattern Recognition Letters, 1986, 4(6): 433-441.

[14]. Rajaraman, S.; Candemir, S.; Kim, I.; Thoma, G.;Antani,S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. Appl. Sci. 2018, 8, 1715.

[15] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6.

[16] M. Goyal, R. Goyal, and B. Lall, "Learning Activation Functions: A new paradigm of understanding Neural Networks," arXiv preprint arXiv:1906.09529, 2019.

[17]. Douarre, C.; Schielein, R.; Frindel, C.; Gerth, S.; Rousseau, D. Transfer learning from synthetic data applied to soil–root segmentation in x-ray tomography images. J. Imaging 2018, 4, 65.

[18] Ahn, J. M., Kim, S., Ahn, K. S., Cho, S. H., Lee, K. B., and Kim, U. S. A deep learning model for the detection of both advanced and early glaucoma using fundus photography.PLoS ONE 13(11): e0207982, 2018

IT retraining at the Faculty of Technical Sciences in Čačak: experiences and recommendations

Marija Blagojević*, Danijela Milošević*
* University of Kragujevac, Faculty of Technical Sciences Čačak, Serbia
marija.blagojevic@ftn.kg.ac.rs, danijela.milosevic@ftn.kg.ac.rs

Abstract - The paper presents the experience in the process of IT retraining at the Faculty of Technical Sciences in Čačak. The cycle of retraining was realized through two cycles: 2018 and 2019. The training was conducted in both cycles through two courses: Java programming and Web application development. A significant percentage of participants managed to complete training and internships and find employment in the field of information technology. The paper also presents the subjective opinions of participants about the training as well as an overview of the implementation of the training.

Keywords: IT, retraining, internship, evaluation

I. Introduction

According to the data of the Government of RS given on the [1], the IT sector is one of the largest net exporters in Serbia and has grown at a rate of over 20% every year, with strong potential for future growth. The sector currently employs about 20,000 workers. The average salary per worker in the IT sector is almost three times higher than in other industries. Each worker in the IT sector contributes about 40 thousand euros a year to Serbia's exports.

According to the latest research in Serbia, about 1,500 new qualified IT workers appear on the market every year. It is necessary to increase that number at least 5 times, to over 7000 IT workers a year, in order for the export goals to be achieved. For that reason, one of the projects of the Ministerial Council for IT is retraining in the field of information technologies.

There are two reasons for additional education, the first refers to people with formal IT education who do not have knowledge and experience for current jobs in the market and the second for retraining talented, interested candidates without IT education who are looking for employment in this sector. The fact that about 500 people are currently employed in IT companies in the Moravica district has been identified, and that at the moment there is a need for at least another 100, with real demand that can be twice as high, especially in the field of JAVA programming and web application development. In the Moravica district there are about 10 IT companies with approximately 30 employees and 50 IT companies with up to 10 employees. In the Moravica district, the IT market has a constant need for IT experts that companies profile with training in various narrower IT disciplines. It is not possible to identify a large number of IT seniors so IT juniors are a real need with additional training through work and training.

The analysis was conducted through interviews with representatives of relevant IT companies in the Moravica district. After the analysis, the need for acquiring new knowledge and retraining of employees in the IT sector was recognized.

The Faculty of Technical Sciences in Čačak trains IT experts, but there has been a constant growth of high school students interested in the field of Information Technology. In addition, the Faculty of Technical Sciences received a request for recommendations for employment in the field of information technology on a weekly basis.

After the adoption of the Proposal for the Improvement of the IT Sector in Serbia, the Government of the Republic of Serbia has determined the manner of implementing the retraining program for individuals talented for IT into junior programmers. The project was organized and implemented by the United Nations Development Program in Serbia (UNDP). One of the performers of the program is the Faculty of Technical Sciences in Čačak in partnership with the Science and Technology Park Čačak.

The process of retraining in information technologies is becoming very common both in Serbia and abroad. Numerous authors have described this process and outlined the main advantages and disadvantages. Since they are mostly adult students, special attention is paid to the motivation for retraining [2]. Hilton [3] points to the need for an increased number of workers in the IT sector but also points to a basic problem in retraining: “graduates of these programs who lack substantive work experience in the field for which they have trained will sometimes have difficulty finding work in that field. Authors like Joia [4] proposed a new model for workers' retraining in Brazil. Model supposes development and use of a Web-based instruction system. Authors in [5] described results of two empirical researches with the graduates of pedagogical requalification studies. Requalification is adopted in different fields, and presented research gives overview in IT field. Requalification is also important for professional training, in both dimensions (initial and continuous training [6]). According to [7] adult learners have some specific characteristics which could effect on learning process, like “they take as much responsibility for their learning as they want to take”.

The project met with great interest from potential participants throughout Serbia. This paper will describe the experiences and subjective assessment of students. Also, the recommendations that came from the two cycles of IT retraining will be presented.

II. METHODOLOGY

The retraining process was implemented through four activities:

A. Selection process

The selection process at the Faculty of Technical Sciences in Čačak was performed through two activities:

• Combined testing to check digital literacy and logical thinking. The mentioned testing was performed using electronic tests.

• Oral interview with candidates where interest in training, level of existing knowledge, motivation and readiness to learn is checked.

The final ranking list was formed by adding the points of the combined test and the oral interview.

B. Training

The training was conducted through two courses: Java programming and Web application development

• Java programming-The purpose of training for Java programming is to acquire knowledge and skills in the field of Java programming with the aim of increasing the number of qualified IT professionals in the Moravica district. The goal of this training is the basic introduction, introduction and acquisition of skills in the field of Java programming in a way that follows current world trends. Java, as one of the most used languages today, is an excellent basis for a professional career in several directions: development of console, desktop, web and Android applications. Java programming training was realized using traditional teaching with the support of the Moodle learning management system within which the Java programming course was developed [8].

• Web application development-the goal of realization of this training is basic introduction, familiarization and acquisition of skills in the field of Web application development in the clearest way. Students will acquire the necessary theoretical and practical knowledge and skills to build modern web applications and sites using different programming languages, libraries, frameworks and tools through training. Through practical examples and projects, participants will gain experience in teamwork and to solve specific programming tasks. As part of the course, more languages and technologies are required to work in the area of Web application development, Web design and web programming, but the primary programming language was JavaScript. Training is designed to introduce participants to the most important techniques, technologies and tools from the domain of Web application development, Web design and programming. . The training of Web application development was realized using traditional teaching with the support of Office 365 learning system under which the Web application development course was developed [9].

C. Internship

The retraining program in the field of information technology, in addition to theoretical classes, includes the development of practical skills and competencies of students through professional practice. Professional internship programs are designed in cooperation with representatives of IT companies and enable participants to acquire a functional knowledge base and develop skills in accordance with the needs of the IT business sector. This concept provides students with quality education through practical training for career development in the field of information technology. Also, in this way, participants establish direct contacts with potential employers and have the opportunity to receive appropriate recommendations, which positively affects their possibility when applying for a job, after the completion of IT retraining.

Partner companies that will provide internships to participants were selected according to the results achieved in the work, which means many years of stable business, steady growth and employment of young university graduates. Coordination of professional practice was realized by NTP Čačak as a partner organization on the project.

D. Mentoring and career guidance and counseling

During the complete duration of the training, mentoring and career guidance of students was organized. A mentor is available to the participants who meets regularly with the participants, has contact by phone and e-mail and who accompanies the participants and advises them during the training. The mentor provides the participants with all relevant information for their training and employment. The goal is to eliminate all potential doubts during the training and solve the risk of dropping out of the course, as well as to provide assistance in finding potential employment and facilitate communication between the student and the lecturer. Mentoring and career guidance will also be conducted through training hours during which participants will be able to write a resume and actively look for a job.

III. RESULTS AND DISCUSSION

The chapter on results is conceived through two sections: the first refers to the review of candidates' performance through the percentages of completion of training, practice and employability; the second relates to participants' subjective assessments of training. Through all implemented activities, recommendations for potential future cycles of retraining in information technologies have been defined.

Information on the IT retraining process is given in the faculty website, within special sections dedicated to this process [10, 11].

Table 1 shows the percentages of successfully completed training and internships for both retraining cycles.

TABLE I. OVERVIEW OF SUCCESS IN TRAINING AND PRACTICE IN RELATION TO THE INITIAL NUMBER OF CANDIDATES

	Percentage of success in completed training		**Success rate of completed internship**	
	Java programming	Web application development	Java programming	Web application development
1. cycle	87,5	94,11	87,5	94,11
2. cycle	90	100	90	100

In both cycles (except for the Web Application Development course in the second cycle) there are participants who dropped out of the training. Participants who regularly attended classes, took tests and defended the final project successfully completed the training and internship in IT companies.

In terms of employability, the Web Application Development group showed better success in both cycles. The total employability on both courses in the first cycle is 43.3%. By courses, 5 out of 14 participants who successfully completed the Java programming course and 9 out of 16 participants who successfully completed the Web Application Development course. In the second cycle, a significantly lower percentage of retraining in IT was observed. Attendees generally remained in their current jobs, which can be analyzed through the current situation of the corona virus pandemic.

Within both cycles, participants were able to do a training evaluation, through periodic surveys in order to improve the teaching process and organization. In addition, meetings with participants were organized. After the completion of the training and practice, the monitoring of the employability of the participants was organized, with mentoring and monthly meetings.

Figure 1 shows how participants experienced the information presented in the course. The largest number of participants believe that the information is given at the appropriate level, while 12.5 and 18.8, respectively, believe that the information is too advanced, or at the basic level. These percentages depend mostly on the prior knowledge of the participants. Although retraining is being done from other branches in information technology, a certain number of students have already studied some areas of programming independently, while a certain percentage of students have started learning from the beginning within the retraining process.

In Figure 2, participants were able to assess the quality of space and equipment at the Faculty of Technical Sciences. The largest number of participants is extremely satisfied with the equipment (68.4%), while there are a certain number of participants who are not satisfied. The faculty is continuously working on raising the level of equipment, especially computer classrooms.

Figure 3 evaluates soft skills training. Even within the soft skills, the participants are extremely satisfied with the conducted training.

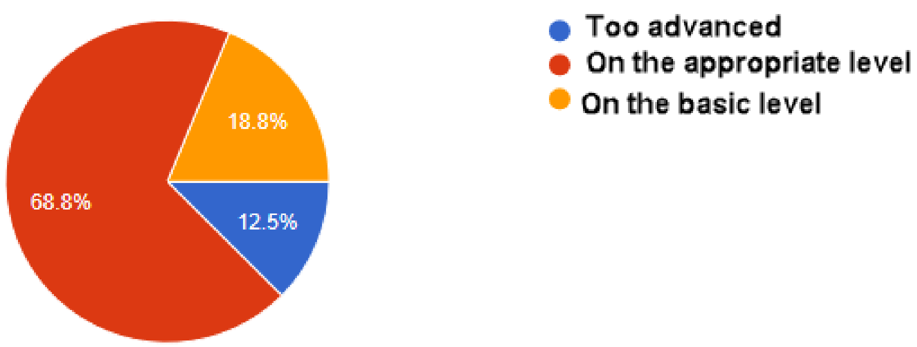


Figure 1. Information during course

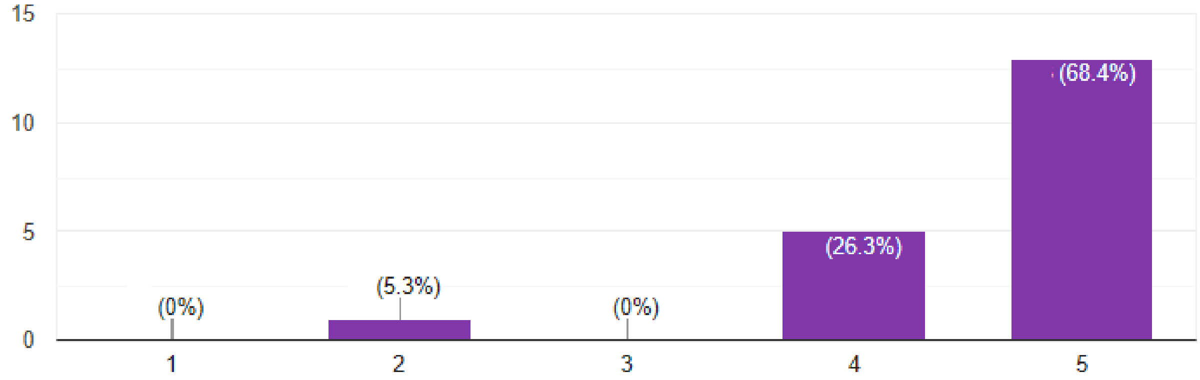


Figure 2. Quality of clasrooms at faculty

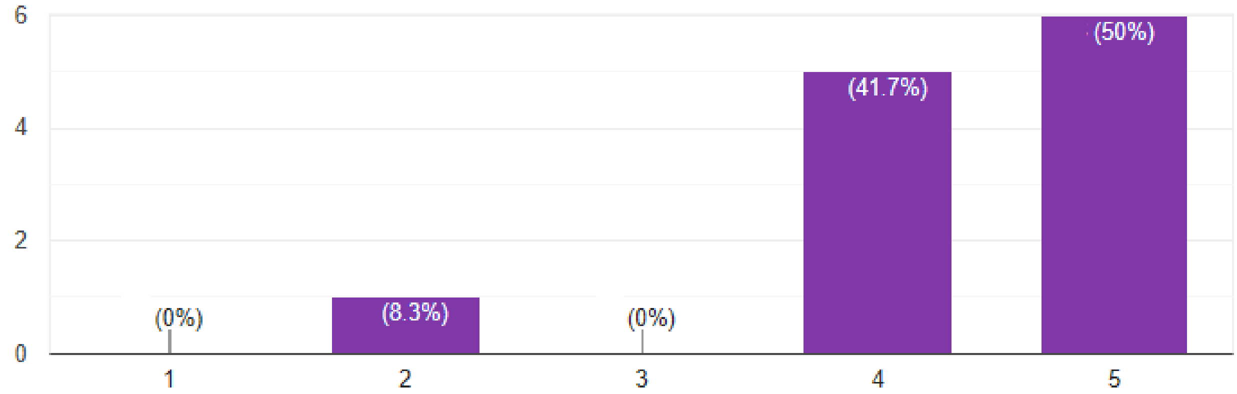


Figure 3. Quality of soft skills course

IV. Conclusion

The most important activities during the implementation of the project are related to teaching within the training and organization of practice. These are key activities that contribute to the acquisition of the necessary knowledge and skills for retraining students and their future successful work in IT companies. In addition to the mentioned activities, study visits to IT companies and attending organized masterclass lectures are also important as an opportunity to meet experienced IT experts with many years of experience in IT. Also, monitoring employability is one of the important project activities.

The sustainability of project activities and results is reflected in the strengthening of links between the faculty and the business sector, which during the implementation of the project has gained much importance. Employers have realized that they can rely on quality training by teaching staff at FTN Čačak, as well as professional support in terms of finding a suitable candidate for specific jobs in the IT profession. The cooperation of FTN Čačak with the Science and Technology Park (NTP Čačak) has been further improved by working on this project. NTP provides on a regular basis spatial and logistical capacities for the formation of new and growth of existing young IT companies. Successful cooperation of project partners FTN and NTP Čačak results in further ideas for new project applications in the field of cooperation between science and economy.

Acknowledgment

This study was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, and these results are parts of the Grant No. 451-03-68/2020-14/200132 with University of Kragujevac - Faculty of Technical Sciences Čačak. The study was also supported by project "Serbia on Your Fingertips – Digital Transformation for Development", IT requalification program, "JAVA programming and Web application development", United Nations Development Program (UNDP), coordinator: dr Danijela Milošević, 15.1.2018-31.10.2019. (1st cycle) and 28.12.2018-30.4.2020. (2nd cycle).

References

[1] IT obuke, dostupno na itobuke.rs, last access July 28th 2020.

[2] S. Ryan, “A Model of the Motivation for IT Retraining”, Information resources management journal, vol. 12, no 4. DOI: 10.4018/irmj.1999100103, 1999.

[3] M.Hilton,” Information Technology workers in the new economy”, Monthly Labor Review, vol. 24, no 6, 2001, p. 41.

[4] Joia, L. “A new model for workers’ retraining in Brazil”, Journal of Workplace Learning, retrieved from: https://www.emerald.com/insight/content/doi/10.1108/13665629910276070/full/html

[5] R. Bakutytė, L. Ušeckienė, D. Illiško, “The development of implementation of learner-centered assessment in pedagogical requalification studies”, Social Velfare, vol. 2, no. 6, 2016, DOI: http://dx.doi.org/10.21277/sw.v2i6.264

[6] L.Vranciu, G. Arionesei, D. Neamţu, “The Evolution of Adult Training Process in Suceava County during 2004-2013, in the Tourism Field”, Procedia - Social and Behavioral Sciences, vol. 142, 2014, pp. 616-622.

[7] K. Illeris, “Adult education as experienced by the learners”, Journal International Journal of Lifelong Education, vol. 22, no.1, 2010, pp. 13-23.

[8] Course Java programming, Moodle LMS: https://eucenje.ftn.kg.ac.rs/course/view.php?id=189

[9] Course Web application development, Office 365: https://teams.microsoft.com/_#/school/conversations/Op%C5%A1ti?threadId=19:231393f737a641e395f51b77cf082942@thread.skype&ctx=channel

[10] 1st cycle of IT precvalification at Faculty of Technical Sciences: http://ftn.kg.ac.rs/aktuelnost-163

[11] 2nd cycle of IT precvalification at Faculty of Technical Sciences: http://ftn.kg.ac.rs/aktuelnost-164

Beacon and beacon-less indoor assisted navigation

A. Dimitrievski*, A. Misev *, S. Savoska ** and V. Trajkovik*
* Faculty of Computer Science and Engineering, Skopje, Macedonia
** Faculty of Information and Communication Technologies, Bitola, Macedonia
ace.dimitrievski@gmail.com, anastas.mishev@finki.ukim.mk, snezana.savoska@uklo.edu.mk, trvlado@finki.ukim.mk

Abstract - In this comparative paper, we review the technologies for indoor navigation for persons with disabilities and persons that require monitoring in Ambient Assisted Living (AAL) environments. We focus on the distinction between beacon-based and beacon-less technologies. We present the categorization of indoor navigation technologies based on the approach taken to determine the location. We identify features of localization technologies that provide classification based on constraints for deployment. We also propose using floor plans to generate a navigation graph, and we give the algorithm for graph creation.

I. Introduction

Indoor assisted navigation could improve accessibility and mobility and provide independence for persons with visual impairment. This technology could also be used in smart living environments, especially in ambient assisted living (AAL) care facilities and homes to provide localization of the person that requires monitoring. In this paper, we focus on less invasive technologies that preserve user privacy; thus, we don't consider solutions such as video monitoring and facial recognition.

Indoor navigation presents a challenge that researchers have addressed using emerging technologies. It is a new field of research that is continually improving. The indoor navigation has two fundamental problems it needs to solve: localization in the environment and mapping of the next steps. There exist several technologies for the localization of visually impaired persons [1]. For this paper, we separate them in beacon-based and beacon-less.

To complete indoor navigation, once the location is determined, we need to guide the person to the desired destination using path planning. When the person taking advantage of the indoor navigation system has mobility or visual impairments, the system should provide not merely the shortest path but a personalized preferred path.

In the next section, we list approaches to the indoor navigation problem. Section III gives an overview of the beacon-based technologies. In section IV, we discuss beacon-less techniques and describe and classify beacon-less technologies. Section V provides a methodology and algorithm for generating navigation graphs. Section VI concludes the paper.

II. Indoor Navigation Approaches

Various approaches can determine indoor localization. The indoor navigation technologies, according to Adler et al. [2], fall under the following categories:

- **Inertial** Systems use pedestrian dead reckoning (PDR) or other IMU tracking techniques. The basic premise of this approach is to estimate relative movements and determine the current position as the sum of individual vectors.
- **Map Matching** Systems that use any previously generated or recorded maps, including patterns of environmental characteristics for the position estimation, such as received signal strength indicator (RSSI).
- **RSS** Systems that use RSS for range estimation.
- **Time of Flight (ToF)** category contains all approaches that use some form of TOF estimation to calculate the distance to another network member.
- **Sound** systems, including ultrasound beacons or other sound sources with known position to estimate their distance to known anchor beacons.
- **Other** Systems that use different spatial depended on environmental properties than described above, including light, magnetic fields, visual object recognition.
- **Multiple** Systems using multi-modal sensing.

For each approach, there can be multiple technical implementations and algorithms to calculate the location. Some implementations can fall under more than one approach. For example, a system can use both sound and ToF to triangulate the location of a device using sound beacons.

The Bluetooth low energy BLE beacons fall under the RSS category. In this category also fall the Wi-Fi signal based and other radio signal based. All such technologies use triangulation algorithms to determine the location of the person.

III. BEACON TECHNOLOGIES

The beacon technologies rely on permanently places beacons with well-known spatial coordinates. These beacons emit a signal that is used in triangulation. Depending on the type of signal, such as with audio signal, a Doppler effect may also be used to measure velocity. When discussing beacon technologies, we usually assume that the beacons are based on BLE. The BLE radio protocol provides new opportunities for the indoor location. It supports portable battery-powered beacons that can be easily distributed at low cost, giving it distinct advantages over Wi-Fi [3]. BLE beacons have omnidirectional radiation, and due to the laws of electromagnetic spreading, the strength of the signal always decreases with the inverse square of the distance from the source, also known as the inverse-square law. The Received Signal Strength Indicator (RSSI) provides information on the distance from the beacon. A minimum of three beacons are required for spatial localization of the receiver (usually a mobile device like a smartphone [4]). There are two main commercial standards for BLE beacon technologies; one is iBeacon, developed by Apple Inc., and the other is Eddystone, developed by Google Inc. The most crucial advantage of iBeacon is that it is very energy efficient, which translates to possible quick deployment of small size beacons that only need to be powered by a battery and eliminates the necessity to rely on any existing infrastructure as Wi-Fi networks [5]. Beacons have a logarithmic proportional correlation between the number of beacons in an area and precision. Adding more beacons gives better results, but after a certain number, a level of saturation is achieved [6]. The optimal layout of beacons in an indoor positioning system was presented in [7].

Beacon technologies can include various methods to determine the location, including:

- Direct RSSI measurement and triangulation.
- Fingerprinting/Scene Analysis which includes: probabilistic methods that rely on the likelihood of the user being in position 'x' provided the RSSI values; Artificial Neural networks (ANN) used in many classification and forecasting scenarios; k-Nearest Neighbor (kNN) algorithms relies on the online RSSI to obtain the k-nearest matches; and Support Vector Machine SVM, primarily used for machine learning (ML) and statistical analysis that has high accuracy [8,9]

IV. BEACON-LESS TECHNOLOGIES

A detailed survey of indoor localization systems and technologies is provided by Zafari et al. in [9]. In summary, the following beacon-less techniques are identified:

- Channel State Information (CSI) – This technology has higher granularity than the RSS as it can capture both the amplitude and phase responses of the channel in different frequencies and between separate transmitter-receiver antennae pairs. Potentially it can have more stable measurements and higher localization accuracy.
- Angle of Arrival estimates the angle at which the transmitted signal impinges on the receiver by exploiting and calculating the time difference of arrival at individual elements of the antennae array. As this requires having multiple antennas, it has limited use with smartphone localization as the measurement is done by the external system and the data then needs to be sent to the smartphone.
- Time of Flight (ToF), Time Difference of Arrival (TDoA) - the principle used by GPS, and Return Time of Flight (RToF) have limited use in indoor navigation because the time precision required would be expensive to deploy. However, in theory, it can be used with ultrasound as the sound speed is slow. One such android based application is presented in [10].

From the same survey, we select the following technologies, whose characteristics are shown in Table I:

- Wi-Fi-based systems usually use or extend existing Wi-Fi infrastructure for indoor navigation. Various techniques, including RSSI measuring, are used to determine indoor location. Although it is possible to build a practical Wi-Fi-based indoor localization system, developing such a system for a large indoor area is not an easy task [11]. The positioning experiment for a pedestrian shows that the reported position is shifted from the actual position before arrival to the vicinity of structures with steel such as elevators. Those factors cause instability of electric waves [12]. Such errors can be corrected using dead reckoning or probabilistic methods. This technology should be considered for indoor assisted navigation as it reuses existing investment in Wi-Fi access points.
- Ultra-Wideband (UWB) has been a particularly attractive technology for indoor localization because it is immune to interference from other signals (due to its drastically different signal type and radio spectrum). UWB can be used to locate people with high precision, but it cannot identify the target individual, so it has limited use for assisted indoor navigation in crowded areas.
- Visible light communication (VLC) uses AoA for localization. As this requires receivers not usually found on smartphones, any implementation would require an additional module that could connect to the smartphone. Due to this limitation, it has limited use for assisted indoor navigation.
- Passive Infrared sensors require sensors to be installed in the environment and are used to detect human motion against the background. These sensors could be used when a small number of people are present. In our previous research, we have shown that these sensors could be used not only to identify the presence of a person but also to detect activities in daily living [13,14].

TABLE I. LOCALIZATION TECHNOLOGIES

Technology	*Range*	*Line of sight*	*Requires infrastructure*	*Required network connection*	*Precision*
Wi-Fi-based	Long	No	Yes	No	Low to medium
Ultra-wideband	Medium	Usually	Yes	Yes	High
Visible light communication	Medium	Yes	Yes	Yes	Medium
Passive infrared (PIR) sensors	Medium	Yes	Yes	Yes	Low to medium
Acoustic signal-based	Medium	Usually	Yes	No	Medium to high
RFID	Short	Yes	Yes	No	High
Computer vision	Medium	Yes	No	No	Medium
Visual/depth	Short/medium	Tes	No	No	Low to high
Magnetic localization	Long	No	No	No	Low
Inertial navigation systems	N/A	N/A	No	No	Low

- Acoustic Signal-based localization technology leverages the ubiquitous microphone sensors in smartphones to capture acoustic signals emitted by sound sources and estimate the user location with respect to the sound source. Although acoustic-based systems have been shown to achieve high localization accuracy, due to the smartphone microphone limitations of receiving only audible band acoustic signals, the transmission power should be low enough not to cause sound pollution. Techniques for signal modulation can help improve localization in a noisy environment. This technology has plausible use for assisted indoor navigation. This approach can also be used in reverse where the phone speaker could emit a sound that, in turn, will be received by connected microphones to triangulate the person receiving assisted navigation. Such systems include the Active Bat system, where a network of wires link the receivers fixed on the ceiling with the network of receivers connected to the server [15].
- RFID is a generic term used to describe a system that transmits the identity of an object or person wirelessly using radio waves. RFID technology is most used to automatically identify objects in large systems [16]. This technology has a very short range, and although cannot be used to locate a person in an indoor area, it can be used to identify objects such as entrances to rooms and can be used in a feedback look for opening automatic doors or requesting a floor in elevators from the smartphone.
- Computer vision systems could use known images and 3D models of the building to determine the location based on the structural objects and landmarks registered by the smartphone camera and calculate the distance based on perspective and field of view. Some types of camera systems are based on AOA. Other techniques use pattern recognition with image processing [17]. This technology requires intensive computation, which can be offloaded to the edge nodes or gradually moved to the phone once the algorithms and processing power become sufficient. This approach has great potential for assisted in indoor navigation. An alternative approach would be to use facial recognition of the person using assisted navigation. This approach has privacy implications and cannot be used in all areas.
- Visual/Depth sensors are used in many research fields due to many available resources, making their implementation easy and fast. This technology can be subdivided into structured light technology, pulse light technology, and stereo camera [17]. Some phones use structured light technology for face unlocking of the phones. As most modern smartphones have multiple cameras, the stereo camera approach can be used. Still, the implementation would vary for each device, and as the cameras are not separated enough, the precision would be hard to achieve.
- Magnetic localization uses magnetic sensors to estimate localization or orientation. Compasses and inclinometers can be fused to estimate the 3 DOF orientation of an object [17]. Most smartphones are equipped with magnetic sensors, and this technique could be used in the sensor fusion approach.
- Inertial Navigation Systems (INSs), aka Dead Reckoning, is a device that approximately determines the current position by knowing the past position and the velocity in which it moves. The dead reckoning is a navigation technology that requires to begin with a known position, and then it will add and track changes. These changes can be in the form of Cartesian coordinates or velocity [16]. Usually, IMUs consist of three main sensors: the accelerometer used for acceleration calculation and linear motion sensing, the gyroscope for angular motion sensing, and the

magnetometer. This technique is used primarily for counting steps that a person takes. However, the accumulative increasing error in addition to the need for initial position specification makes IMU not reliable, especially in indoor environments where no GPS can be added [17]. Inevitably, measurement errors are present within the sensor data, and the triple integration of them results in a potentially cubic growth in time (drift). INSs for aviation, marine, and the military use highly accurate sensors that keep the error sources very small and permit tracking for many hours. These are too bulky and expensive for pedestrian navigation [18].

We did not consider techniques such as floor tiles and other more "exotic" technologies that require expensive installations and have only been tested in laboratories in extremely controlled environments or need dedicated hardware other than a smartphone.

V. Path Planning

In addition to localizing users, a navigation system can provide directions from the user's current location to a user-specified destination, which involves planning a path and turning it into easy-to-follow directions. As the user follows directions, the system will dynamically update its estimation of its location and generate a new direction once the previous has been completed. Path following algorithm tries to limit the inaccuracy and latency, which is inherited from the positioning system and to consider human factors [19]. Path-planning algorithms use graphs to represent the environment [20]. The graph is usually generated from the 2D grids and could also be based on 3D shapes [21]. When planning a path using graph-based approaches, the environment is divided into sets of nodes and edges connecting these nodes. The nodes are generated from the grid, where each empty cell yields one node. Depending on the path planning algorithm and constraints, these nodes might be any object type, such as hallway intersections, doors, or obstacles. Most of the current navigation systems use either Dijkstra or A* algorithm [19,22].

While optimizing path and avoiding obstacles has been extensively studied, much of the research was done to navigate robots [23]. While some of the models could be applied to assist the navigation of a person, we identify two constraints in such a scenario: bandwidth and precision. The bandwidth constraint is the limit of feedback data a person could receive and process. For example, in audio feedback, this is the amount of verbal information that could be obtained and be beneficial without overwhelming the person; in tactile feedback, the limiting factor is the variety and duration for acknowledging the feedback. The precision constraint refers to the ability to estimate distances and angles for people and the variation in standard measures such as length of a person's step. Considering these constraints, the navigation system should work with a large margin of error and have a limited guidance set of instructions. Algorithms used for navigating robots will have to be adjusted. For example, the algorithm for navigating a robot could safely use a path through a narrow distance between two obstacles, but such a route should be avoided for a person.

An additional consideration of path planning is given based on impairments or persons. Obstacles differ based on the type of impairment and severity. For example, escalators introduce difficulty to persons with visual impairment, and persons with a specific type of mobility impairment, escalators are an insurmountable obstacle. They should be removed from the path graph. Elevators, on the other hand, provide an accessible alternative for persons with mobility impairments. However, options for avoiding elevators should be provided to accommodate persons with claustrophobia. In certain premises, some areas might be off-limits to visitors or require visitors to be escorted, adding additional complexity to the path planning. These considerations on the application level translate to distances between the graph nodes, resulting in different graphs for different people.

Floor plan maps can be used to acquire a semantic plan [24]. When building the graph from the floor plan, we propose a two-step process. The first step identifies the points of interest that should be precisely defined; these include doors, starting points of stairs and escalators, and furniture. The second step is the partition of the transit area, such as hallways, and define some points based on the available area. These nodes should not be too far apart that a small change of direction could prevent the person from reaching the next point. A sample floor plan is shown in Fig. 1. Here we see several points of interest with graph nodes such as the nodes D, H, I, J, K, Q. The remaining space contains nodes in a grid-like pattern. To minimize the number of nodes after the initial graph from the grid is generated, we do a pre-processing to find a cluster of closely adjacent nodes that are fully connected and replace them with one node located in the geometrical mean of the 2D space. The points of particular interest, such as doors and elevators, can either be extracted from the floor plan by locating such objects. If we don't have the original plan, we can do vertical and horizontal pass to find narrow openings in the walls.

The graph is then connected so that any two nodes have a path clear of obstacles, and the distance between them is less than a given threshold. For this floor plan, we generate the bidirectional graph shown in Fig. 2. Here nodes G and E are not connected to avoid the sofa chair, and nodes F and I are not connected as the distance is too great, and it would be difficult for a visually impaired person to move precisely at the given angle. The weights are then calculated with consideration for distance, possible obstacle, preference for the forward direction, and difficulty to follow instructions. For example, "turn 90 degrees left" should have a lower weight than turn "45 degrees left". In [19], six possible instructions are identified: go forward; turn left a bit; turn left; turn right a bit; turn right; turn around. Multi-path is allowed as the person might pass the node, and another path should be available instead of instructing them to go back. Figure 3 presents the flow diagram for this algorithm.

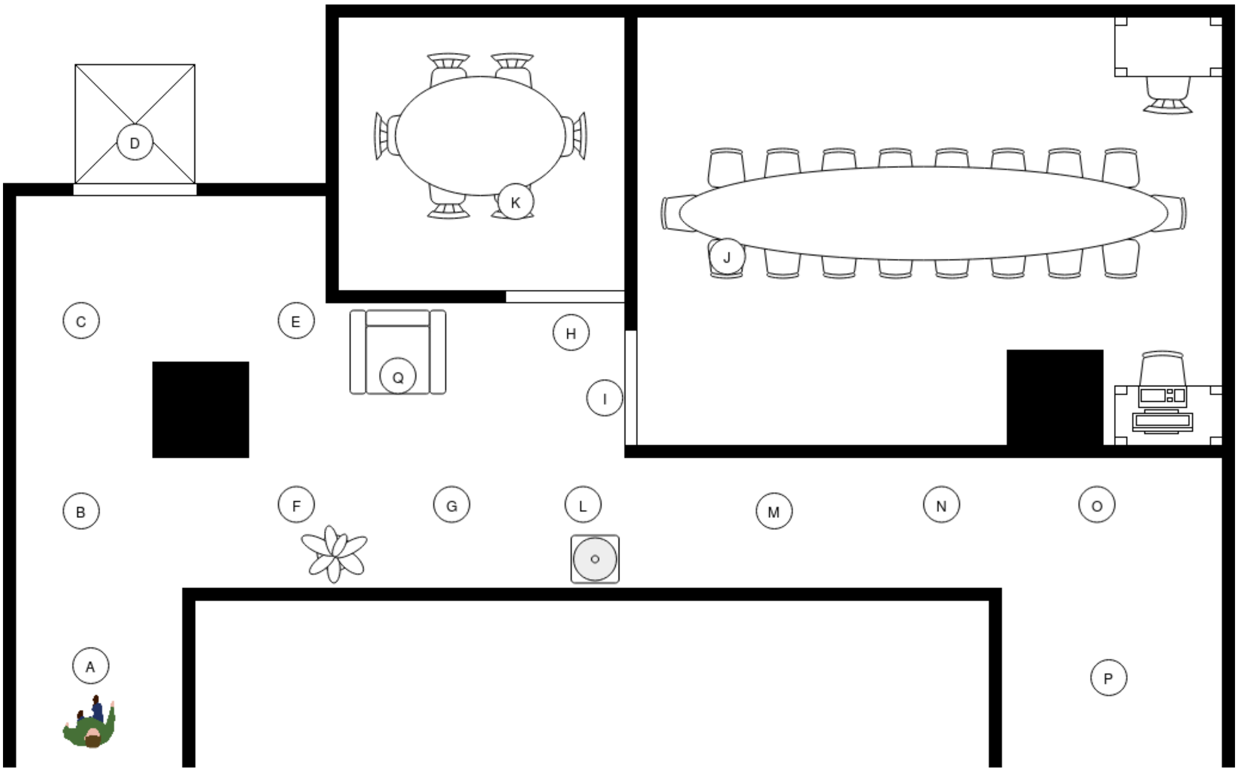


Figure 1. Example of floor plan for indoor navigation

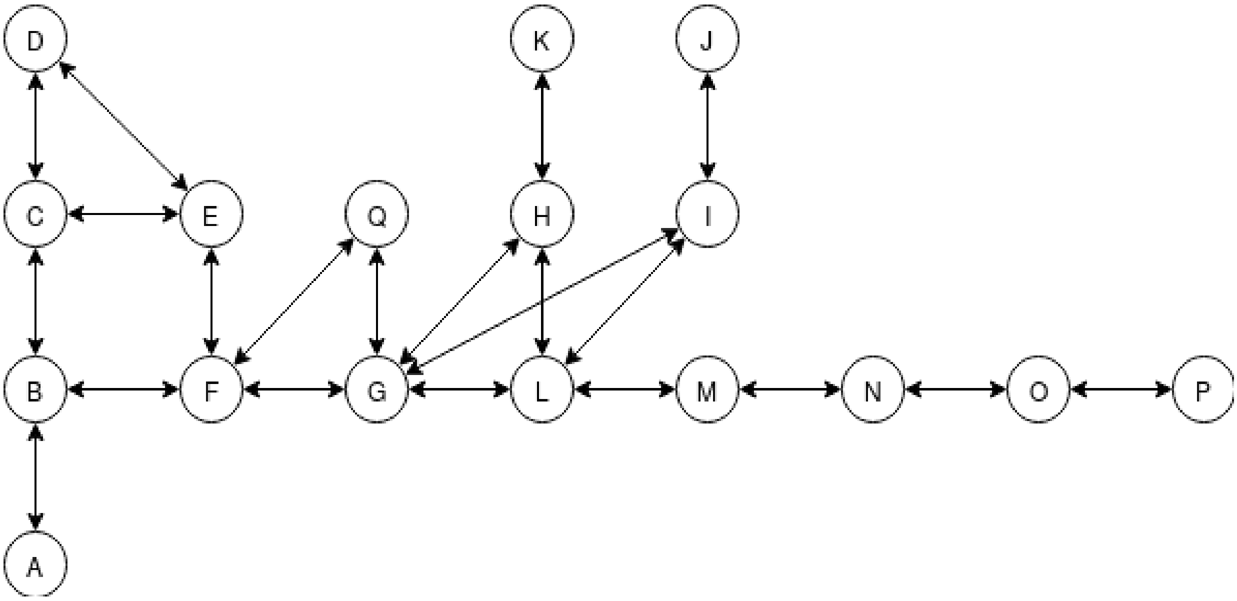


Figure 2. Example of a graph generated from the floor plan in Figure 1

VI. CONCLUSION

In this paper, we made a comparative study of indoor navigation technologies focusing on the difference between beacon-based and beacon-less techniques. We noted the multiple approaches for determining the location of a person. We described each technology and provided classification based on multiple environmental constraint parameters.

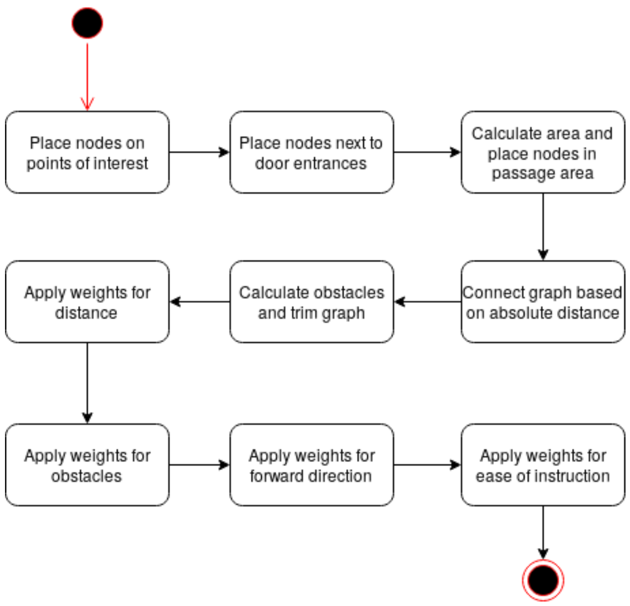


Figure 3. Flow chart diagram for graph generation algorithm

As we have shown, many approaches, technologies, and techniques exist for indoor localization and navigation, each presenting a set of advantages and challenges to overcome. Implementing a fusion of these technologies, new systems are designed, and existing systems improved.

In the previous section, we developed a methodology to use floor maps to generate a navigation graph, and we proposed an algorithm for graph creation with consideration of constraints. Having accurate data on the indoor environment, including obstacles, is essential for indoor navigation as the localization techniques. Our approach aims to reduce the complexity by pre-creating a navigation graph as an alternative to a real-time calculation that may face scaling challenges for larger areas.

REFERENCES

[1] N. Raut, S. Jangam, L. Palve, P. Ukarande, and V. Pawar, "A survey on indoor navigation for blind people," 2019.

[2] S. Adler, S. Schmitt, K. Wolter, and M. Kyas, "A survey of experimental evaluation in indoor localization research," in 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN). IEEE, 2015, pp. 1–10.

[3] R. Faragher and R. Harle, "Location fingerprinting with bluetooth low energy beacons," IEEE journal on Selected Areas in Communications, vol. 33, no. 11, pp. 2418–2428, 2015.

[4] Y. Zhuang, J. Yang, Y. Li, L. Qi, and N. El-Sheimy, "Smartphone-based indoor localization with bluetooth low energy beacons," Sensors, vol. 16, no. 5, p. 596, 2016.

[5] H. K. Fard, Y. Chen, and K. K. Son, "Indoor positioning of mobile devices with agile ibeacon deployment," in 2015 IEEE 28th Canadian conference on electrical and computer engineering (CCECE). IEEE, 2015, pp. 275–279.

[6] M. Ji, J. Kim, J. Jeon, and Y. Cho, "Analysis of positioning accuracy corresponding to the number of BLE beacons in indoor positioning system," in 2015 17th International Conference on Advanced Communication Technology (ICACT). IEEE, 2015, pp. 92–95.

[7] M. Laguna, J. O. Roa, A. R. Jim' enez, and F. Seco, "Diversified local search for the optimal layout of beacons in an indoor positioning system," Iie Transactions, vol. 41, no. 3, pp. 247–259, 2009.

[8] S. Subedi and J.-Y. Pyun, "Practical fingerprinting localization for indoor positioning system by using beacons," Journal of Sensors, vol. 2017, 2017.

[9] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2568–2599, 2019.

[10] M. C. P'erez, D. Gualda, J. M. Villadangos, J. Ure˜na, P. Pajuelo, E. D´ıaz, and E. Garc´ıa, "Android application for indoor positioning of mobile devices using ultrasonic signals," in 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN). IEEE, 2016, pp. 1–7.

[11] D. Han, S. Jung, M. Lee, and G. Yoon, "Building a practical wi-fi-based indoor navigation system," IEEE Pervasive Computing, vol. 13, no. 2, pp. 72–79, 2014.

[12] Y. Inoue, A. Sashima, and K. Kurumatani, "Indoor positioning system using beacon devices for practical pedestrian navigation on mobile phone," in International conference on ubiquitous intelligence and computing. Springer, 2009, pp. 251–265.

[13] A. Dimitrievski, E. Zdravevski, P. Lameski, and V. Trajkovik, "Towards application of non-invasive environmental sensors for risks and activity detection," in 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2016, pp. 27–33.

[14] P. Lameski, A. Dimitrievski, E. Zdravevski, V. Trajkovik, and S. Koceski, "Challenges in data collection in real-world environments for activity recognition," in IEEE EUROCON 2019-18th International Conference on Smart Technologies. IEEE, 2019, pp. 1–5.

[15] W. Sakpere, M. Adeyeye-Oshin, and N. B. Mlitwa, "A state-of-the-art survey of indoor positioning and navigation systems and technologies," South African Computer Journal, vol. 29, no. 3, pp. 145–197, 2017.

[16] M. A. Al-Ammar, S. Alhadhrami, A. Al-Salman, A. Alarifi, H. S. Al-Khalifa, A. Alnafessah, and M. Alsaleh, "Comparative survey of indoor positioning technologies, techniques, and algorithms," in 2014 International Conference on Cyberworlds. IEEE, 2014, pp. 245–252.

[17] F. Alkhawaja, M. Jaradat, and L. Romdhane, "Techniques of indoor positioning systems (IPS): A survey," in 2019 Advances in Science and Engineering Technology International Conferences (ASET). IEEE, 2019, pp. 1–8.

[18] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1281–1293, 2013.

[19] H. Wu, A. Marshall, and W. Yu, "Path planning and following algorithms in an indoor navigation model for visually impaired," in Second International Conference on Internet Monitoring and Protection (ICIMP 2007). IEEE, 2007, pp. 38–38.

[20] P. Bhattacharya and M. L. Gavrilova, "Roadmap-based path planning-using the Voronoi diagram for a clearance based shortest path," IEEE Robotics & Automation Magazine, vol. 15, no. 2, pp. 58–66, 2008.

[21] M. Xu, S. Wei, S. Zlatanova, and R. Zhang, "Bim-based indoor path planning considering obstacles," ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci, pp. 417–423, 2017.

[22] N. Fallah, I. Apostolopoulos, K. Bekris, and E. Folmer, "Indoor human navigation systems: A survey," Interacting with Computers, vol. 25, no. 1, pp. 21–33, 2013.

[23] M. Filipenko and I. Afanasyev, "Comparison of various slam systems for mobile robot in an indoor environment," in 2018 International Conference on Intelligent Systems (IS). IEEE, 2018, pp. 400–407.

[24] S. L. Joseph, X. Zhang, I. Dryanovski, J. Xiao, C. Yi, and Y. Tian, "Semantic indoor navigation with a blind-user oriented augmented reality," in 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2013, pp. 3585–3591

A Conjecture on the Exchange Rate of Bitcoin

Amandeep Kaur*, Deepak Chahal**, Latika Kharb**
* Delhi Technological University, Delhi, India
** Jagan Institute of Management Studies, Sector-05, Rohini, New Delhi, India
*akm0394@gmail.com, **deepak.chahal@jimsindia.org, **latika.kharb@jimsindia.org

Abstract - Bitcoin is one of the widely traded crypto currencies and are been around since October 2008 and it is a decentralized digital currency which is transferred person to person without any intermediary that means there are no financial institutions or any Government control measures. Bitcoins are highly liquid and have low transaction costs. This paper focuses on forecasting bitcoin exchange rate to Indian Rupee using Autoregressive Integrated Moving Averages (ARIMA) model. Data has been collected from the database of coingecko.com spanning from January 2016 to March 2019. Augmented Dickey Fuller (ADF) test has been utilized to check the stationarity of the time series. The ARIMA model parameters are based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis. The parameter with the lowest RSS score has been used for forecasting. The forecasted values depicts rise in the exchange rate. The findings of this study will help investors in taking a rational decision about future investments in Bitcoins.

Keywords: Fiat Currency, Crypto Currency, ARIMA model, Logarithmic Return

I. INTRODUCTION

As an emerging decentralized architecture and distributed computing paradigm underlying Bitcoin and other cryptocurrencies, blockchain has attracted intensive attention in both research and applications in recent years [1]. A Crypto currency is a virtual currency (having no physical embodiment) intended to fill in as a vehicle of trade, it is a means of carrying out a transaction digitally independent of any regulating financial institution, government control or any central authority. It requires no intermediary to execute a transaction being completely decentralized in its working, security of transactions is guaranteed by cryptographic algorithms. Crypto Currencies are been around since 1980s. Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Bitcoin Cash (BCH), Stellar Lumens (XLM), EOS (EOS), Litecoin (LTC), Cardano (ADA), Monero (XMR), Tether (UDST), TRON (TRX), Dash (DASH), IOTA (MIOTA), Binance Coin (BNB), NEO (NEO), Ethereum Classic (ETC), NEM (XEM), Tezos (XTZ), Zcash (ZEC), VeChain (VET), Bitcoin Gold (BTG), Maker (MKR), OmiseGO (OMG), 0x (ZRX), Dogecoin (DOGE), Decred (DCR), QTUM (QTUM), Ontology (ONT), Lisk (LSK), Zilliqa (ZIL), Aeternity (AE), Bitcoin Diamond (BCD), BAT (BAT), BitShares (BTS), Nano (formerly RaiBlocks) (NANO), Bytecoin (BCN), ICON (ICX), Pundi X (NPXS), Siacoin (SC), DigiByte (DBG) Steem (STEEM), Verge (XVG), Populous (PPT), Bytom (BTM), Aurora (AOA), Chainlink (LINK), Waves (WAVES), Metaverse ETP (ETP), Augur (REP), and Golem (GNT) are the top 50 Crypto Currencies [2]. Bitcoin is a popular cryptocurrency that records all transactions in a distributed append-only public ledger called blockchain[3].

Bitcoin is not the first introduced crypto currency, but it has been the most successful amongst all in today's arena having largest share in terms of total market capitalization value [4]. It is an open source peer to peer system of electronic cash [5] i.e. no intermediaries are involved where transaction takes place through a common ledger called Blockchain. Before 2017 Bitcoin's share of total crypto currency was in excess of 70 percent however, as the new crypto currencies were circulated amongst investors in Initial Coin Offerings (ICOs) its dominance rate was dropped down to 32.48 percent on January 13, 2018 [6]. At present the world's largest crypto currency acquires 50.9 percent of total capitalization of entire market according to the database of CoinMarketCap.

First Bitcoin was introduced in October 2008 by pseudonymous person Satoshi Nakamoto in its research paper. Bitcoins are subject to fluctuations as the supply is limited to 21 million only which cannot be changed and are expected to be achieved latest by 2140. Near to 4.3 Billion Bitcoins are left which are yet to be circulated which means around 16.7 million are presently available, out of the present availability 30% is expected to be lost forever as a result of hard drive crashes and misplaced private keys [7] [8].

Bitcoin is the most popular cryptocurrency nowadays. Inspired by the success, both industry and academia seek to apply Bitcoin's core technique, Blockchain, to other fields like finance, healthcare and Internet-of-Things [9].

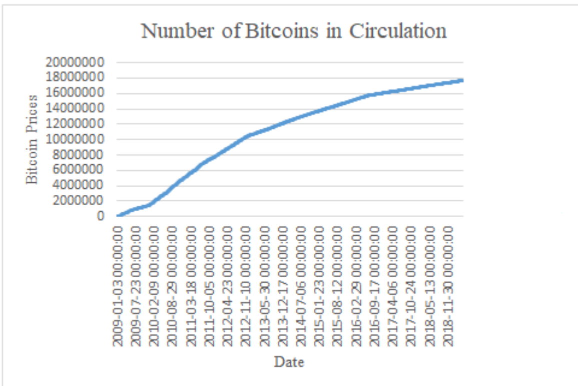


Figure 1. Number of Bitcoins in Circulation

From the above graph we can say the number of bitcoins in circulation has only increased from 50 in January 2009 to 17619000 in March 2019 an all time high in number [10]. They are divisible up to eight decimal places 0.00000001. Since it does not have any central location nor any governing body they are very difficult to hack and are controlled by a consensus of market participants. In order to make a transaction one needs to have a digital wallet on Bitcoin network. Bitcoin Wallet is an application which allows the user to send or receive the bitcoins, it keeps a complete record of all incoming and outgoing transactions, each of the transactions are linked to the wallet and are store for future reference. The wallet includes two keys viz., a public key and a private key. A public key is a key used to send bitcoins to user's address, verifying signatures in a transaction to ensure everything is in order and to authenticate and finalize the transaction, it consists of 27-34 alpha numeric characters while a private key is a secret number that allows users to unlock and spend the bitcoins for purchases.

Bitcoins are not backed by any tangible assets except for the trust which people have in it like in case of Fiat Currencies. Fiat currencies are issued by the Regulatory Authority whose value is controlled by the relationship between supply and demand and the issuing authority's stability. Bitcoins can be acquired through buying (from a person or exchange), Barter (selling product or service in exchange for Bitcoins) and Mining (run software to find bitcoins).

II. RESEARCH METHODOLOGY

All the data has been sourced from secondary sources. The study employs the daily settlement prices of Bitcoins collected from the database maintained by coingecko.com spanning from January 2016 to March 2019. All the compiled values are represented in INR. The Autoregressive Integrated Moving Average (ARIMA) model has been utilized to estimate the future exchange rate movement of Bitcoin. The underlying assumption for ARIMA model is the stationarity of the variable which is about to be forecasted. Stationarity signifies that the time series exhibit mean reversions i.e. the values of the variable fluctuates over time around a constant mean. To ensure the stationarity of the time series Augmented Dickey-Fuller test (ADF) has been employed. The daily settlement prices of Bitcoin are represented graphically as follows:

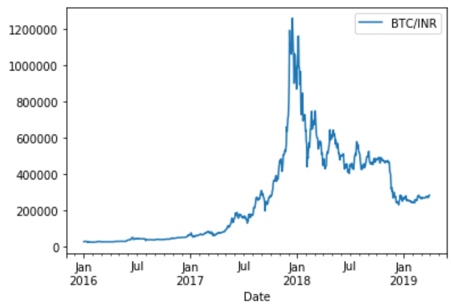


Figure 2. Bitcoin Exchange Rate in INR from January 2016 to March 2019

From the graph it can be ascertained that Bitcoin prices were not stationary they showed a constant trend from January 2016 to December 2016, it happened to increase from January 2017 to June 2017, a small shortfall in the prices is observed from July 2017 to September 2017, it began to rise again from October 2017 to January 2018 however shortfall in prices was observed from August 2017 to September 2017. The rising prices showed a drastic fall in February 2018 which continued till March 2018, from April 2018 prices have showed somehow a positive movement but it was not as high as in January 2018. November 2018 and December 2018 was again a period of downfall, from January 2019 the downward trend was transformed to a constant one.

A. Unit root test

A Unit Root is a feature of Stochastic process that evolve with the passage of time in time series data which might pose problems in statistical inference. Augmented Dickey-Fuller (ADF) test is an indirect method for checking the stationarity in the time series through the presence of a unit root. The non- stationary time series viz., Bitcoin/INR are transformed into stationary by taking the logarithmic return through the application of the following mathematical expression:

$$R_t = Log\,(\frac{P_t}{P_{t-1}}) \qquad (1)$$

Where,

R_t = Logarithmic Return

P_t = Daily Closing Prices at Time t

P_{t-1} = Daily Closing Prices at Time t-1

Logarithmic returns are the symmetric representations of returns/growth rate by the same multiple. Constant values are represents zero returns [11]. The distribution of daily logarithmic returns for Bitcoin exchange rate in INR is represented in figure 3.

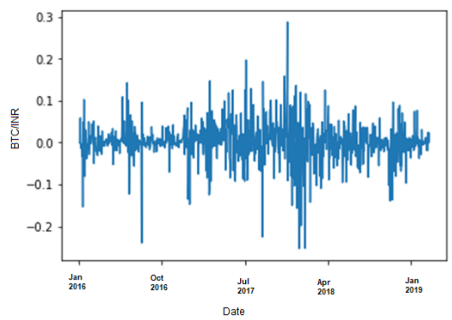


Figure 3. Logarithmic Return of BTC/INR series spanning from January 2016 to March 2019

The obtained time series has been checked for unit root through the application of Augmented Dickey fuller (ADF) test by using the following regression equation:-

$$\Delta Y_t = \propto_0 + \beta Y_{t-1} + \sum_{j=1}^{p} \gamma_j \, \Delta Y_{t-j} + e_t \qquad (2)$$

Where,

Δ = Difference Operator
Y_t = Prices in time period t
$\propto_0$ = Constant
e_t = Error term
β= Coefficient on time trend

p = Lag of the Autoregressive process

The hypothesis to be tested under ADF test is:
H_0: The given time series are stationary
H_a: The given time series exhibit non-stationarity

The null hypothesis will be accepted if the computed test statistic is less than the critical values. The Augmented Dickey fuller test for the above stated hypothesis is tested on the Bitcoin exchange rate in INR and on logarithmic return series of the data.

```
from statsmodels.tsa.stattools import adfuller
print("Results of Dickey-fuller test;")
dftest = adfuller(btc["BTC/INR"], autolag = "AIC")

dfOutput = pd.Series(dftest[0:4], index = ['Test Statistic', 'P-Value', 'Lags Used',
                     'Number of Observations Used'])
for key,value in dftest[4].items():
    dfOutput['Critical Value{%s}' %key] = value
dfOutput
```

Figure 4. ADF test on Bitcoin exchange rate in INR

```
print("Results of Dickey-fuller test:")
dftest = adfuller(diff_Test, autolag = "AIC")
dfOutput = pd.Series(dftest[0:4], index = ['Test Statistic', 'P-Value', 'Lags Used',
                     'Number of Observations Used'])
for key,value in dftest[4].items():
    dfOutput['Critical Value{%s}' %key] = value
dfOutput
```

Figure 5. ADF test on Natural Logarithmic Return Series

TABLE I. AUGMENTED DICKEY-FULLER TEST RESULTS

Variables	***Test Statistics***	***Critical Values***
BTC/INR (Original Prices)	-1.927233	1%: -3.438244 5%: -2.865024 10%: -2.568625
BTC/INR (Logarithmic Return)	-7.332575e+00	1%: -3.438225e+00 5%: -2.865016e+00 10%: -2.568621e+00

```
Out[11]:
Test Statistic                  -1.927233
P-Value                          0.319360
Lags Used                       14.000000
Number of Observations Used    831.000000
Critical Value{1%}              -3.438244
Critical Value{5%}              -2.865024
Critical Value{10%}             -2.568625
dtype: float64
```

Figure 6. ADF test results on Bitcoin Exchange rate in

```
Out[21]:
Test Statistic                -7.332575e+00
P-Value                        1.117839e-10
Lags Used                      1.100000e+01
Number of Observations Used    8.330000e+02
Critical Value{1%}            -3.438225e+00
Critical Value{5%}            -2.865016e+00
Critical Value{10%}           -2.568621e+00
dtype: float64
```

Figure 7. ADF test results on Log return series

The ADF test results on Bitcoin exchange rate in INR depicts the data is non stationary as the test statistics is greater than the critical value whereas the ADF test on Logarithmic return series depicts the stationarity in the data because test statistics is lower than the critical values at 1%, 5% and 10%. Hence, the Ha cannot be accepted, the Ho stands true in case of natural logarithmic return series.

It can be inferred that the time series is stationary and integrated of order 1.

B. ARIMA MODEL

ARIMA model is made out of Autoregressive (AR) and Moving Average (MA) model, the model is then integrated with data of difference process. The differencing of data is done to ensure the stationary characteristics. The combination of this is known as Autoregressive Integrated Moving Averages (ARIMA) Model. The Autoregressive (AR) model specifies the present values of a variable relies on the values that the variable took in the past periods plus an error term, it can be represented by the AR (p) equation, and the notation AR (p) portrays an Autoregressive model of order p. The AR (p) model is characterized as:-

$$Y_t = C + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t$$
$$Y_t = C + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \varepsilon_t \qquad (3)$$

Where,
$\emptyset_i, \ldots, \emptyset_p$ = Parameters of the model
C = Constant
ε_t = Random Error
Y_t = Prices in the time period t

The moving-average model signifies that the output variable is a linear combination of the residuals/ error terms occurring simultaneously in the past. The notation MA (q) refers to the moving average of the order q. The MA (q) is defined as:-

$$Y_t = \mu + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \cdots + \varphi_q \varepsilon_{t-q}$$
$$\mathrm{Y_t} = \mu + \varepsilon_t + \sum_{i=1}^{q} \varphi_i \varepsilon_{t-i} \qquad (4)$$

Where,
$\varphi_i \ldots . \varphi_q$ = Parameters of the model
$\varepsilon_t \ldots \varepsilon_{t-q}$= White noise error terms
μ = Mean of the series

Combining the above two models viz., Autoregression (AR) and Moving averages (MA), we can frame the Autoregressive Integrated Moving Averages (ARIMA) model. In an ARIMA model the future estimation of a variable is a direct mix of AR and MA models i.e. the future values are based on the past values plus past error terms, I stands for order of differencing I(1) , it can be defined by the following mathematical expression:-

$$Y_t = C + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \cdots - \varphi_q \varepsilon_{t-q} \qquad (5)$$

Where,

Y_t = Actual Prices in time period t
ε_t = Random Error
$\emptyset_p$ and φ_q = Model Parameters [12]

The ADF test allows us to go further in time series forecasting i.e. to find out suitable values of p in AR model and q in MA model. The values of p and q can be examined from the Autocorrelation Function and Partial Autocorrelation Function plot of Log return series ranging from lag 1 to lag 10. Autocorrelation Function (ACF) measures the correlation between the current value and the value at the previous time spots. Partial Autocorrelation Function (PACF) measures the correlation at two time spots given that we consider both observations are correlated to observations at other time spots observations.

```
plt.subplot(121)
plt.plot(lag_acf)
plt.axhline(y=0,linestyle="--",color="gray")
plt.axhline(y=-1.96/np.sqrt(len(diff_Test)),linestyle="--",color="gray")
plt.axhline(y=1.96/np.sqrt(len(diff_Test)),linestyle="--",color="gray")
plt.title("Autocorrelation Function")
plt.show()
```

Figure 8. Code for Autocorrelation Function

```
plt.subplot(122)
plt.plot(lag_pacf)
plt.axhline(y=0,linestyle="--",color="gray")
plt.axhline(y=-1.96/np.sqrt(len(diff_Test)),linestyle="--",color="gray")
plt.axhline(y=1.96/np.sqrt(len(diff_Test)),linestyle="--",color="gray")
plt.title("Partial Autocorrelation Function")
plt.tight_layout()
```

Figure 9. Code for Partial Autocorrelation Function

Figure 10 shows the plot of correlogram ACF of logarithmic return series of BTC/INR. It can be inferred that Autocorrelation at lag 8 is exceeding the upper significance limit and the rest of the coefficients are within the confidence limits. Figure 11 represents the PACF correlogram of the lag 1 to 10 for the natural logarithmic differenced time series, the PACF also infers that the coefficients are within the confidence limits except the one outlier at lag 8 which can be assumed as an error and happened by chance.

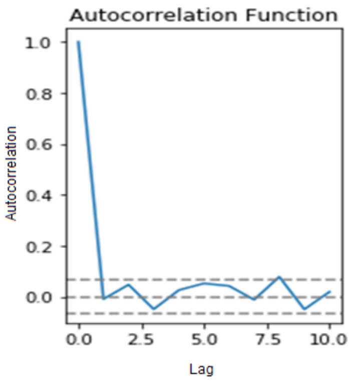


Figure 10. ACF plot of Log return BTC/INR series from January 2016 to March 2019

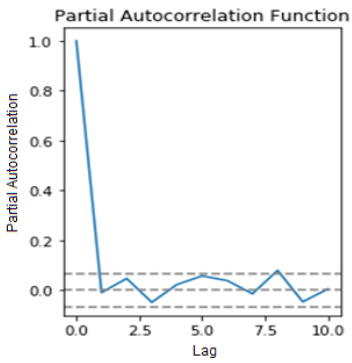


Figure 11. PACF plot of Log return BTC/INR series from January 2016 to March 2019

TABLE II. ACF AND PACF COEFFICIENTS FOR LAG 1 TO 10

Lag	*ACF*	*Lag*	*PACF*
1	-0.00906954	1	-0.00908029
2	0.04729978	2	0.04733345
3	-0.04889219	3	-0.04833872
4	0.02595758	4	0.02321362
5	0.05237301	5	0.05786591
6	0.04244054	6	0.03907171
7	-0.0115967	7	-0.01403525
8	0.0782741	8	0.08059895
9	-0.04942635	9	-0.04693021
10	0.0192636	10	0.0053552

In order to go further in forecasting by using ARIMA, it is necessary to find suitable parameters i.e. p, d and q which corresponds to Autoregressive, Differencing and Moving average components respectively.

The values of p and q can be estimated from the ACF and PACF correlograms and d is the order of differencing used in the time series i.e. 1.

The expected ARIMA (p, d, q) models can be: ARIMA (1,1,0), ARIMA (0,1,1), ARIMA (2,1,1), ARIMA (0,1,0), ARIMA (2,1,0). To select the best suitable model for forecasting out of the above 5 proposed models Residual Sum of Square (RSS) score has been chosen as the decision making criteria. Residual sum of squares is the unexplained variance between the regression function and the data set, it is also known as Squared Errors of Prediction (SSE). The smaller the RSS value, the better it is.

The results of the ARIMA (p, d, q) models are

TABLE III. ARIMA MODELS WITH RSS VALUE

Model	*RSS Value*
ARIMA (1,1,0)	1.9214
ARIMA (0,1,1)	1.9215
ARIMA (2,1,1)	1.9067
ARIMA (0,1,0)	1.9216
ARIMA (2,1,0)	1.9172

The lowest RSS value is 1.9067 for ARIMA (2, 1, 1) model where p=2, d=1, q=1. Hence ARIMA (2, 1, 1) model has been utilized to forecast the future exchange rate movement of Bitcoin in INR. The predicted values are generated in array form for coming 365 days.

```
abc=results_ARIMA.forecast(steps=365)
```

Figure 12. Code used for Forecasting Bitcoin Exchange Rate

The Blue line represents the future predicted values, the grey shaded area represents the confidence area which means the values cannot rise beyond the upper confidence limit and cannot fall beyond the lower confidence limit. Besides recording transfers of currency, the Bitcoin blockchain is being used to save metadata — i.e. arbitrary pieces of data which do not affect transfers of bitcoins [13].

The future Bitcoin prices are expected to increase in future with slight fluctuations and hence an investor can consider it as a good investment opportunity.

0	12.5603
1	12.5624
2	12.5669
3	12.5682
4	12.5735
5	12.5771
6	12.5794
7	12.5837
8	12.5853
9	12.5888
10	12.5912
11	12.5942
12	12.597
13	12.5998
14	12.6024
15	12.6054
16	12.6079
17	12.6108
18	12.6134

Figure 13. Sample of Predicted values of Bitcoin Exchange Rate in Array form

```
results_ARIMA.plot_predict(1,1211)
```

Figure 14. Code used for creating plot for Predicted Bitcoin exchange Rate

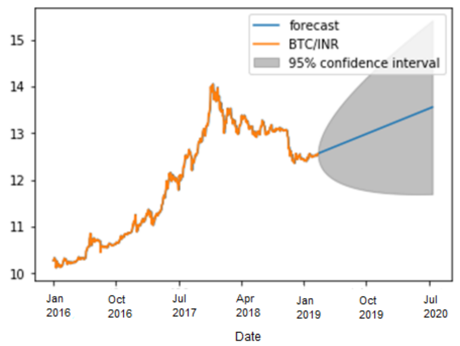


Figure 15. Graphical Presentation of Bitcoin Forecasted Exchange Rate

III. CONCLUSION

The objective of this study is to forecast the future exchange rate movements of one of the widely traded cryptocurrencies viz., Bitcoin. The daily settlement prices of Bitcoin spanning from January 2016 to March 2019 are taken into consideration for forecasting. The stationarity of the time series has been checked by Augmented Dickey fuller test, the test results shows the log return series are stationary as the test statistics = -7.332575e+00 is less than the critical values at 1%, 5%, 10 , i.e. -3.438225e+00, -2.865016e+00 and -2.568621e+00 respectively. Further different parameters for ARIMA model has been proposed and the one with lowest RSS value has been selected i.e. ARIMA (2,1,1) having RSS = 1.9067. ARIMA (2,1,1) has been used to forecast the future bitcoin exchange rate movements. The output is generated in an array form for coming 365 days. From the generated values it can be interpreted that Bitcoin exchange rate in coming future is expected to increase hence, it can be considered as a good investment opportunity.

REFERENCES

[1] Y. Yuan and F. Wang, "Blockchain and Cryptocurrencies: Model, Techniques, and Applications," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 48, no. 9, pp. 1421-1428, Sept. 2018. doi: 10.1109/TSMC.2018.2854904.

[2] https://www.investinblockchain.com/top-cryptocurrencies/, access date: 29-04-2019.

[3] M. Conti, E. Sandeep Kumar, C. Lal and S. Ruj, "A Survey on Security and Privacy Issues of Bitcoin," in IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3416-3452, Fourthquarter 2018. doi: 10.1109/COMST.2018.2842460

[4] Tian Guo, NinoAntulov-Fantulin, Predicting short-term Bitcoin price fluctuations from buy and sell orders, 2018. Doi: https://arxiv.org/pdf/1802.04065v1.pdf

[5] Satoshi Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System.

[6] www.coindesk.com/bitcoins-share-of-total-crypto-market-slips-back-toward-50 access date: 02-05-2019.

[7] https://coincentral.com/how-many-bitcoins-are-left/ access date: 07-05-2019

[8] Cathy YH Chen, Wolfgang Karl Hardle, Ai Jun Hou, Weining wang, Pricing Cryptocurrency Options: The case of CRIX and Bitcoin, IRTG 1792 Discussion paper, 2018, Vol: 4

[9] L. Tseng, "Bitcoin's Consistency Property," 2017 IEEE 22nd Pacific Rim International Symposium on Dependable Computing (PRDC), Christchurch, 2017, pp. 219-220. doi: 10.1109/PRDC.2017.39

[10] www.blockchain.com, access date: 16-05-2019.

[11] Lukas Pichl, Taisei Kaizoji, Volatility Analysis of Bitcoin Time Series, Quantitative Finance and Economics, 2017.

[12] G. Peter Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, Elsevier, Pg. 159 – 175, Vol.50, 2002.

[13] Bartoletti, M., Bellomy, B. & Pompianu, L. J Grid Computing (2019) 17: 3. https://doi.org/10.1007/s10723-019-09473-3

Connected Health Systems Supported by Blockchain: An overview

V. Trajkovik*, G. Gavrilov** and S. Savoska***
* Faculty of Computer Science and Engineering, University Ss Cyril and Methodius, Skopje, R.N. Macedonia
** School of Computer Science and Information Technology, University American College Skopje, R.N. Macedonia
*** Faculty of information and communication technology, University St. Kliment Ohridski Bitola, R.N. Macedonia
vladimir.trajkovik@finki.ukim.mk, gavrilovgoce@yahoo.com, snezana.savoska@uklo.edu.mk

Abstract - This paper analyzes the application of blockchain based approach for exchange of healthcare data within ambient assisted living systems, transforming them into connected health systems. When it comes to the consequences of abuse to the complex health-related data, their portability and discretion ask novel approaches in management of healthcare data. This paper analyzes the possibility and the potential of the blockchain technology in the systems dealing with healthcare data exchange. The purpose of this research is to justify the reasoning why to use distributed consensus introduced by blockchain technology in the domain of the connected health systems.

I. INTRODUCTION

Rapid development of the data and communication technologies, together with recent COVID 19 pandemics, pose new opportunities and challenges in the healthcare sector. The advances in mobile technologies and wireless sensor networks can revolutionize the way health related data is gathered, disseminated, and used by healthcare policy makers, providers, patients and citizens in general. The ubiquities of these technologies can transform the healthcare services from the services that are tightly coupled to the health ordinations and hospitals into pervasive connected health services. Activities and services that require physical presence of the patients can then be performed at any-time and any-where. This can provide additional flexibility in patients' life. It can also help the healthcare policy makers and providers to focus on preventing diseases and promoting health.

The main goal of this paper is to present connected health data exchange cycle based on blockchain technology. The rest of this paper is organized as follows. Section II explains the background and related work in Connected health area while Section III introduces connected health data exchange cycle. Section IV elaborates on blockchain data exchange principles deployed in connected health. Section V summarizes the technical challenges related to authentication, trust and access control, while section VI concludes the paper.

II. BACKGROUND AND RELATED WORK

The development of health information technologies in the last decade creates a broad range of new opportunities to improve the access to health services for citizens and especially healthcare delivery [1], [2]. The other trend to consider is increase of healthcare costs [3]. Healthcare stakeholders should control costs while maintaining the quality of health care services on the same or higher level.

Implementation of electronic health records (EHR), electronic medical records (EMR) and personal health records (PHR), is viewed as a step towards achieving improvements in the health care system in many European countries [4]. EHR is defined as: 'a repository of information regarding the health status of a subject of care in computer process able form, stored and transmitted securely, and accessible by multiple authorized users' [5], [6]. EHR can be understood as a repository of patient data in digital form, which stored and exchanged securely. PHR is a health record where the patient himself maintains health data and information related to the care of a patient. PHR provides a summary of an individual's medical history which is accessible online, under strictly defined access permissions.

Ambient and Assisted Living (AAL) covers range of activities starting with creating care models and algorithms, continuing with different ICT solutions connected with AAL (robotic assistance for the elderly, measurement and sensing technologies, and devices connected to internet of things within the smart housing) [7]. The typical AAL services include but are not limited to monitoring of elderly people vital signs, home rehabilitation, robotic assistance for the elderly, as well as, people with chronic conditions, peoples with different impairments, or wider population [8]. If we take into consideration the environmental factors that influence on wide population as health disturbance factors, measured by many instruments and IoT sensors connected with data collection, we have a huge pool of big data that have to be considered as the health factors [9].

AAL systems need to ensure high-quality-of-service. Essential requirements of AAL systems are usability, reliability, data accuracy, cost, security, and privacy [10].

According to [11], in order to achieve these requirements, it is essential to involve citizens, caregivers, industry, researchers, and governmental organizations in the development cycle of AAL systems, so that end-users can benefit more from the collaborative efforts. The use of ambient intelligence can provide environment adoptable to the user and influence the ability for AAL systems to express the power of a human being. In this context, the usage of advanced information and communication technologies (e.g. social networks) could help connect people and organize community activities within AAL.

Connected Health (CH) describes the new paradigm of a technology-enabled model of health and lifestyle management. It is extension of AAL set up to provide, besides conventional healthcare, preventive or remote treatments, resorting to an electronic information structure based on internet, sensing, communications and intelligent techniques in support of health-related applications, systems and engineering. It refers to a conceptual model for health management, wherein devices, services or interventions are designed around patients' needs, being health related data shared in such a way that patients, caregivers, policy makers, education and research, companies, third sector organizations and funders in the health and welfare sectors, are seamlessly connected in order to provide healthcare in the most proactive and efficient manner possible [12].

CH services increase the quantity and complexity of healthcare information. Therefore, developing information systems that are capable not only to store and retrieve health data, but offer continuous monitoring of health data that is very important for both healthcare stakeholders. For these reasons, extracting data according to healthcare established standards is important [13].

CH systems also should support the process of decision making and diagnosing, searching through large amounts of health data and facts, classifying them and identifying issues that directly relate to a given medical condition. In this way they could offer citizens to be directly involved in their health care, providing information that will assist in making decisions about their own health. Patients will have a greater role in the decision-making processes related to their health as they could be empowered with the ability to gain access and manage information that fits with their personalized needs. Ultimately, they will be able to shape their health as a reflection of the health model of the whole community [14].

III. Connected Health Data Exchange Cycle

One of the advantages of CH systems should be integrating data from AAL systems and smart homes with data from EHR, EMR or PHR. Presenting these health-related big data can lead to more efficient and informed decisions by physicians, nurses, patients, and informal caregivers [15]. Aggregating data from different sources, including medical devices and integrating them with data from health records enables a more comprehensive view of health data, but creates various data integration challenges [16]. A convenient way to integrate data from EHRs, PHRs, AAL, home care, and self-care systems is to integrate them in a highly connected, robust, and reliable edge-based platform, offering uninterrupted availability and taking care of data throughput, network efficiency and data privacy [17].

A general architecture of CH systems uses mobile devices for collecting data from environmental and body network sensors, but also for easy access to specific personal data. Raw data is preprocessed, filtered by noise, and then integrate them with different EMR. These data can be then processed using healthcare algorithms that transform them into relevant information which is distributed and used by different services. This model includes the processing of data aggregated from social networks needed to give different recommendations to the users and medical centers along with data collected by different sensors [18].

The exchange of the original data and the CH services results between the CH stakeholders is one of main task in CH. This process focuses on connecting patients, healthcare providers, policy makers and other relevant stakeholders, with an aim to guarantee the smooth data flow between them. The purpose of CH is timely sharing, patients' condition monitoring and accurate information presenting through smart combining uses of health data, devices and communication software/hardware platforms. In this way, CH aims to bring patients, clinicians and health science researchers all together to help the society to answer one question: how to connect patients, therapists and caregivers to deliver the optimum health results in an era of stretched resources and increasing demands [19].

Figure 1 shows typical date exchange in CH. The end users normally include patients/outpatients and the elderly. The sensing devices are used for collecting health related data and providing real time monitoring results. The healthcare providers can use the sensed data from the end users for predictive analysis [20]. During this process, the data converges from the end users to the healthcare institutions. Based on the provided data, CH propose suggestions and summative conclusions to the healthcare professionals, and in return, they can give more accurate and personalized treatments for the end users based on the feedbacks and analysis results from the CH systems. Through this way, the predictive analysis results obtained using machine learning (ML) or Artificial Intelligence (AI) in general, are distributed to the end users, who benefit from contributing their own personal data.

CH system stakeholders can be of different size and form, including but not limited to clinical centers, hospitals, laboratories, insurance companies, pharmacies, emergency centers, nursing homes, and public health centers. The data that is subject to transfer is typically structured in different formats and described with different terminologies, making the interfaces and integration process more complex and with high cost. Additionally, the data that is subject to a transfer is scattered across the storages of the organizations that are involved in the process, which often leads to inconsistent data handling processes and erroneous and incomplete data records.

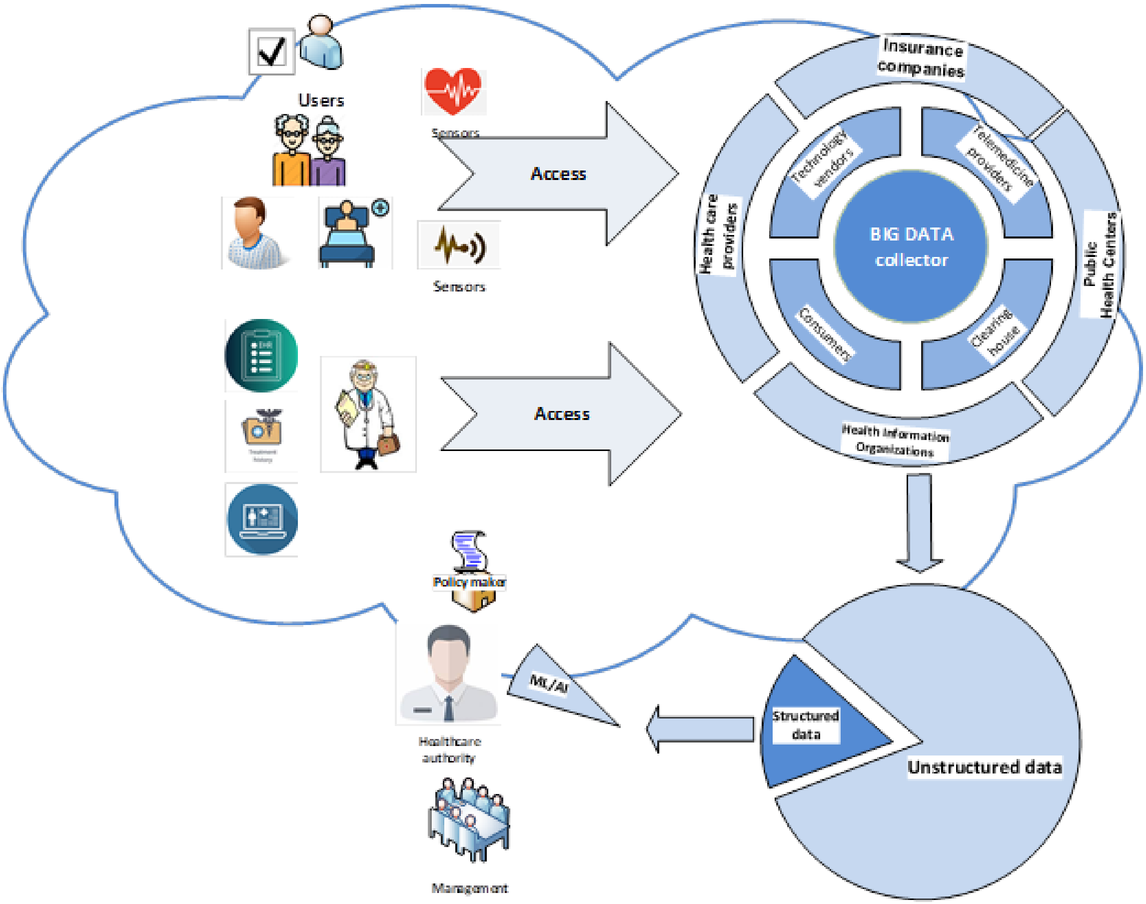


Figure 1. Connected health information exchange

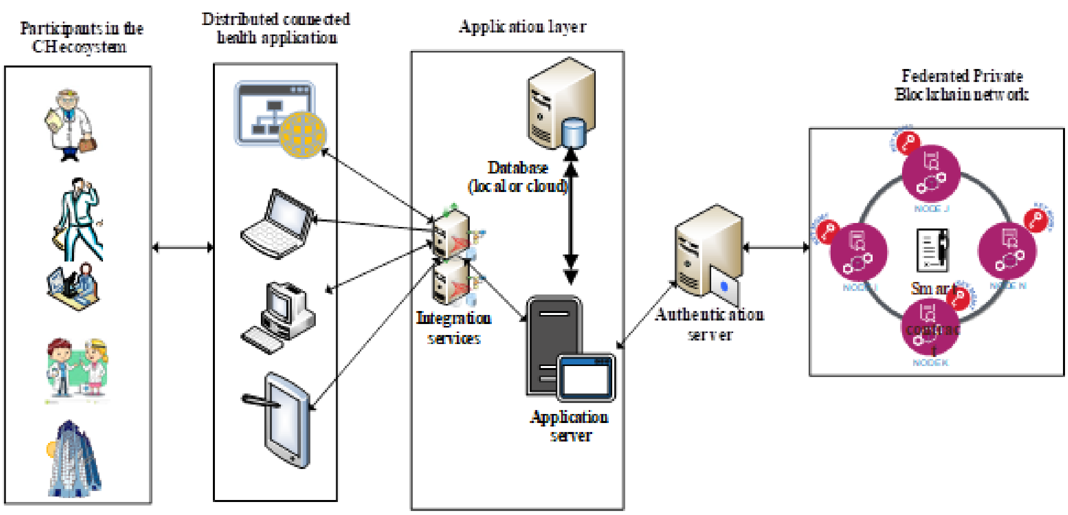


Figure 2. Logical architecture of the blockchain model for CH

The data gathered from EHRs is holding a huge amount of useful information that can be transformed into lifesaving actions. In addition, with different analytical approaches of the data, the costs of the diagnostics and treatments can be greatly reduced with more focused and predicable management [21].

As the research in CH matures, data generated from medical devices will benefit from the analysis by well-established machine learning techniques. There is also potential that new research in machine learning and artificial intelligence can be used on data generated from the sensors used in CH [22].

CH cannot only provide services in evidence-based medicine but can also educate patients to gain health knowledge to achieve better self-management and self-care. As recently data analysis has been becoming more and more popular, this technology has been applied widely to assist health problem discovering and clinical decision making [23].

IV. Blockchain Based Data Exchange Principles

The blockchain infrastructure can be used to create a powerful catalog of health-related records that references different data sources and connect the patients, healthcare providers, laboratories, researchers, health insurance organizations and many other participants in the CH system. Blockchain can help to simplify the way how the parties involved in the CH system exchange the data and collaborate between themselves.

Blockchain in its essence is a distributed system that stores records for specific transactions. It is a distributed ledger of peer-to-peer transactions, which are grouped in blocks that are connected between them. Each block contains a hash, or numeric digest of its content, that verifies the integrity of the containing transactions. The hash of the next block in the blockchain network is dependent on the hash of the current block; the hash of the current block is dependent on the hash of the previous block. This effectively makes the entire blockchain history immutable, as changing the hash of any block would also change the hash of future blocks. The functioning blockchain does not depend on a central, trusted authority, rather than, the responsibility of functioning is distributed to all nodes which participate in the network. Because is missing central authority that will verify the validity of the blockchain, a mechanism for reaching network consensus must be employed. The smart contract plays a vital role in performing the agreement among various stakeholders involved in the CH system. A smart contract is an integral and inseparable part of the blockchain-based applications. A smart contract represents a computer protocol that follows specific rules, codes and constraints agreed by all participants. It is an agreement made among various involved stakeholders in the defined CH system. The CH data can be encrypted and shared with the whole ledger available within the respective network.

These core principles enable decentralized interactions (processes of storing, exchanging and accessing data) between each participant in the network, bypassing the need for intermediaries and regulatory bodies to ensure trust [24].

Blockchain technology provides a transparent, decentralized, authenticated platform that applies a consensus-driven approach to facilitate the interactions of multiple entities in the network through the use of a shared ledger. Because is missing central authority that will verify the validity of the blockchain, a mechanism for reaching network consensus must be employed. By using the permission mechanism for tracking which organization is able to see which medical records in the ledger, we can utilize the blockchain infrastructure to create an immutable chain of content, supported by decentralized network. The blockchain “smart contract” can execute exchange logic on the distributed layer and connects patients with healthcare providers and guarantee the privacy and visibility of data only upon approval (or notification) [25].

I this way blockchain technology can offer a solution that not only helps to securely store and sharing of medical and healthcare data but also to assure the confidentiality of each patient's data by giving the patients, as well as their medical and health data ownership [26]. Many current development projects in healthcare have blockchain technology in the center of their development [27].

V. Blockchain Model

To implement the blockchain technology in CH, decentralized identity management built on top of a multi-block consortium private network can be used. Identity management system is using the decentralized smart-contract standard that defines the method for ownership and transferability of the referrals.

The Figure 2 represents a logical architecture of the blockchain model for CH [28,29]. Application layer, a Database, and an authentication and authorization server are the main components of the model. Authentication and authorization of the users, system make by validating transactions in the blockchain network. The users (the patient, doctor, physician, pharmacists, researchers, employees in clinical centers, hospitals, laboratories, insurance companies, emergency centers, nursing homes, public health centers) communicate with the system through the CH applications. This CH application can be web, mobile, or a standalone application that interacts with the application server via integration services to perform the desired functions. For delivering a requested medical and health data, the application server needs to communicate with the database and the authentication server. To accomplish these activities, users must be authenticated and authorized.

Authentication server receives work from the application server, to check the authenticity of the user and to authorize access to the database. After receiving a permission validation from the authentication server, user’s access to the database from the application server is allowed. Authentication server authenticates and authorizes the user and able to interact with the blockchain network. It has an intermediary role between the blockchain network and the application server and vice

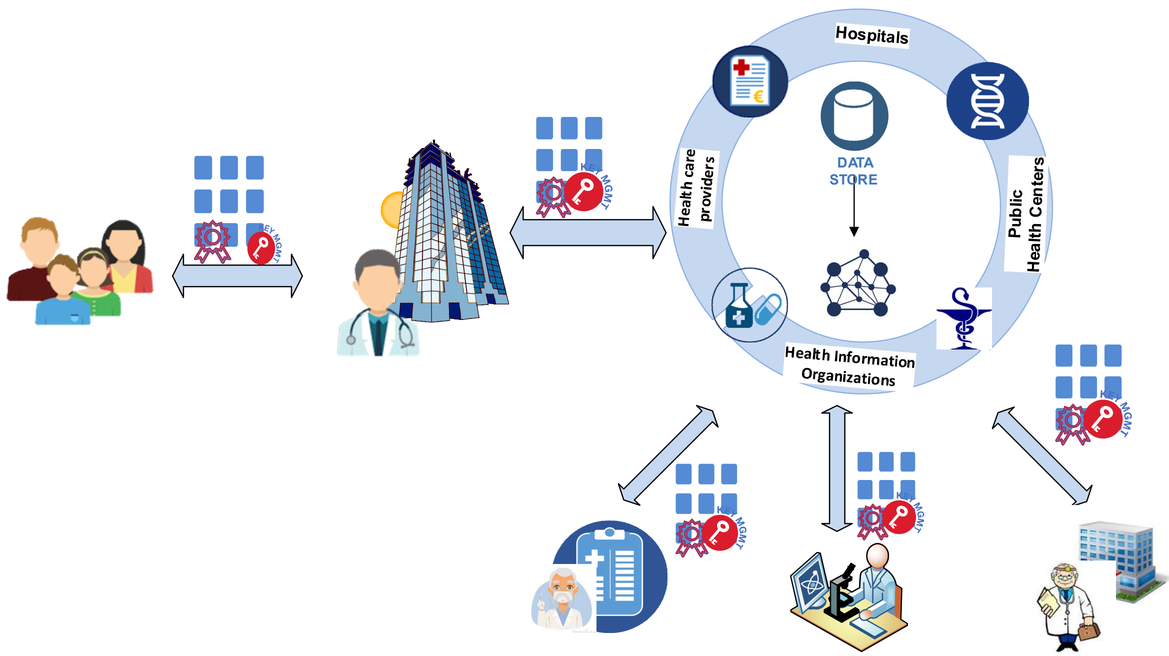


Figure 3. Creating shared data in blockchain based CH marketplace

versa. The data operations and various data access requests for immutability and integrity protection are recorded by the blockchain network. The nodes in the blockchain network follow the rules in the smart contract, validate and broadcast transactions. It keeps the network in running state, maintain the ledger and run the consensus protocol.

The user first must registers in the trusted services offices by providing personal details (such as ID, Biometrics, and PIN), together with the user's public key. Since the registration process is a one-time process, users provide their details using their mobile devices or web application.

Patients produce data by executing a smart contract in private blockchain network that tokenizes a trusted data and decide who to share it with. During this process are used metadata such as user's ID, quantity, healthcare institution, ID of the users of that data (see Figure 3). Patients digitally sign consent forms deciding who has access to their medical data. Healthcare providers and patients produce data from various sources: imagery, biological exams, connected medical devices, genetic tests, surveys and questionnaires.

CH system stakeholders (healthcare organizations, clinical centers, hospitals, laboratories, insurance companies, pharmacies, emergency centers, nursing homes, public health centers) store data “off-chain” and list it on the blockchain along with permissions given by the patient. Healthcare organizations receive a token representing a valid data issued by the patient and after that store in private blockchain network.

A Big data marketplace is built between CH system stakeholders and healthcare providers. Each stakeholder (healthcare authorities, policy makers, scientist doctor, and physician) queries the database to gain access to the information they need. Upon receiving the request query by the blockchain network from the user side, it will check the validity of the user's rights and whether the patient has granted the access permission to that particular user. If the check is successful, it performs the query operation.

VI. CONCLUSION

The existence of diverse health devices and apps with a combinations of the Internet of Things have contributed to transfer of a large amount of medical data daily. Access and sharing health care data before, create basis for novel healthcare services that utilize those large amounts of data. This extensive connected but distributes database of citizens' health information creates significant privacy, security, and availability issues.

Blockchain technology in healthcare information systems has brought immense opportunities in terms of not only providing secure and efficient data storage but also sharing and control access to the data. System model for identity and access management using blockchain technology is proposed in this paper. With the progress in electronic health and interoperability, healthcare data store in the cloud and patient data privacy protection regulations, new opportunities are appearing for health data management, as well as patients' opportunity to access and share their health data. This is possibility that enable each of the healthcare players to protect their data in the CH systems.

REFERENCES

[1] Koceska, N., et al. A telemedicine robot system for assisted and independent living. Sensors, 19(4), pp. 834-851. (2019).

[2] Starič, K. D., et al. A. Can we predict obstetric anal sphincter injury?. European Journal of Obstetrics & Gynecology and Reproductive Biology, 210, pp. 196-200. (2017).

[3] Bergmo, T.S. How to Measure Costs and Benefits of eHealth Interventions: An Overview of Methods and Frameworks. Journal of Medical Internet Research, Vol. 17, No. 11, e254. (2015).

[4] Roehrs A., et al. Personal Health Records: A Systematic Literature Review, Journal of Medical Internet Research, Vol. 19, No 1, e13. (2017).

[5] ISO/DTR 20514:2005 Health Informatics-Electronic Health Record- Definition, Scope, and Context. [Online] Available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39525, (2005).

[6] Snezana Savoska, Ilija Jolevski, Blagoj Ristevski, Natasha Blazeska-Tabakovska, Andrijana Bocevska, Boro Jakimovski, Ivan Chorbev, Vassilis Kilintzis, Design of Cross Border Healthcare Integrated System and its Privacy and Security Issues, In Proceedings of Computer and Communications Engineering, Workshop on Information Security 2019, 9th Balkan Conference in Informatics, Volume 13, 2/2019, Pp.58-64

[7] Goleva, R. I., et al., V. AAL and ELE platform architecture. In Ambient assisted living and enhanced living environments, Butterworth-Heinemann. pp. 171-209. (2017).

[8] Memon, M., et al. Ambient Assisted Living Healthcare Frameworks, Platforms, Standards, and Quality Attributes. Sensors. 14, pp. 4312-4341. (2014).

[9] Savoska, Snezana and Ristevski, Blagoj and Blazheska-Tabakovska, Natasha and Jolevski, Ilija (2019) *Towards Integration Exposome Data and Personal Health Records in the Age of IoT.* In: 11th ICT Innovations Conference 2019, 17-19 October, Ohrid, Republic of Macedonia.pp.237-246

[10] Dimitrievski, A., et al. A survey of Ambient Assisted Living systems: Challenges and opportunities. In 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP) pp. 49-53. (2016).

[11] Sun, H., et al. Promises and challenges of ambient assisted living Systems. In: Proc. of International Conference on In-formation Technology: New Generations (ITNG), pp.1201 –1207. (2009).

[12] Mountford, N., et al. Connected Health in Europe: Where are we today? University College Dublin. (2016).

[13] Kaloyanova K., Krastev E., Mitreva E., Extracting Data from General Practitioners' XML Reports in Bulgarian Healthcare to Comply with ISO/EN 13606. BCI'19: Proceedings of the 9th Balkan Conference on Informatics•September 2019 •Article No.: 3, Pages 1–5, https://doi.org/10.1145/3351556.3351578

[14] Xu, L., et al. What clinics are expecting from data scientists? A review on data-oriented studies through qualitative and quantitative approaches. IEEE Access, 7, pp. 641-654. (2018).

[15] Knaup, P., Schöpe, L., Using Data from Ambient Assisted Living and Smart Homes in Electronic Health Records, Methods InfMed. 53, pp,149–151. (2014).

[16] Lameski, P., et al., Suppression of intensive care unit false alarms based on the arterial blood pressure signal. IEEE Access, 5, pp. 5829-5836. (2017).

[17] Dimitrievski, A., et al. Data Processing within Ambient Assisted Living System. AIIT 2019 Conference, (2019).

[18] Vlahu-Gjorgievska, E., et al. Connected-health algorithm: development and evaluation. Journal of medical systems, 40(4), pp. 109-121. (2016).

[19] Caulfield B., and Donnelly S., `What is Connected Health and why will it change your practice?' QJM, Int. J. Med., vol. 106, no. 8, pp. 703-707. (2013).

[20] Dimitrievski, A., et al. Towards application of non-invasive environmental sensors for risks and activity detection. In 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP) pp. 27-33. (2016).

[21] Sun Y., et al. Internet of Things and big data analytics for smart and connected communities, IEEE Access, vol. 4, pp. 766-773. (2016).

[22] Loncar-Turukalo, T., et al. Literature on Wearable Technology for Connected Health: Scoping Review of Research Trends, Advances, and Barriers. Journal of medical Internet research, 21(9), e14017. (2019).

[23] Kvedar J., et al. Connected health: A review of technologies and strategies to improve patient care with telemedicine and telehealth,' Health Affairs, vol. 33, no. 2, pp. 194-199. (2014).

[24] Conceic¸ A.F., et al. Electronic health records using blockchain technology. arXiv preprint arXiv:1804.10078 (2018)

[25] Bell, L., Buchanan, W.J., Cameron, J., Lo, O. Applications of blockchain within healthcare. Blockchain in Healthcare Today (2018)

[26] Peterson, K., et al. A blockchain-based approach to health information exchange networks. In Proc. NIST Workshop Blockchain Healthcare, Vol. 1, pp. 1-10. (2016).

[27] Khezr, S., et al. Blockchain technology in healthcare: A comprehensive review and directions for future research. Applied Sciences, vol. 9 No.9, pp.1736-1764. (2019).

[28] Gavrilov, G., et al. Blockchain-based model for authentication, authorization, and immutability of healthcare data in the referrals process., In Proc of CIIT 2020 Conference, (2020).

[29] Dodevski, Z., Trajkovik, V. Decentrilizing the health information exchange systems: A blockchain based approach. In Proc of CIIT 2018 Conference, (2018).

Web-Based Educational Software for Teaching Distributed Information Systems with Partitioned Database Recovery

Lj. Kazi*, A. Kansara**, B. Radulović*, I. Berković*, G. Gecin*, S. Nadrljanski*, V. Šašić*
* University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
** Parth Systems LTD, Navsari, Gujarat, India
ljubica.kazi@gmail.com, amar.kansara@gmail.com, biljana.radulovic66@gmail.com, ivana.berkovic62@gmail.com, gorangecin@gmail.com, sloba92ki@gmail.com, sasic.vladimir2@gmail.com

Abstract –Aim of this paper is to present practical part of teaching at course Distributed Information Systems, within Information Technology master studies at University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia. The paper specially highlights created educational software as a teaching tool. The software is created as a model that students could use in their practical work during learning and examination. The software is created as an n-tier web application that enables distributed data processing with web services, while working with partitioned databases. Particular focus in this paper has been put on the recent enhancement of the software, by developing a module that enables recovery of data in partitioned databases, by using time stamp.

I. Introduction

There are many teaching methods and tools, designed to encourage students' interaction and active role in the teaching and learning process. "*Learning by doing"* has been a paradigm that directs students to be engaged in the working process, where being active in the creation or transformation leads to better understanding of underlying theoretical concepts. Therefore, in higher education, in aim to bring students closer to their future working role (in industry, public services etc) [1], courses regularly include both theoretical and practical classes.

Aim of this paper is to present an educational software created and used in Information Technologies master studies course Distributed information systems (DIS) at University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia (@TFZR). The educational software was created by Ljubica Kazi in 2011, for the purpose of practical classes (the software name: DIS@TFZR). It provides a model of a solution that students could use in their practical work – mid-term exams and projects. The educational software is created as a web application, that integrates web services, object-oriented n-tier architecture and partitioned databases. This paper specially emphasizes the new segment of the software that has been developed, in interaction of Ljubica Kazi with students, during school year 2019/20. It enables recovery of partitioned databases by using timestamp data.

The rest of the paper is organized as follows: second section presents theoretical background by description of basic terms, third section presents short literature review of related work, fourth section describes basic elements of the developed educational software, section five describes the implementation of database partitioning, section six describes the database recovery implemented in this software and the final section provides conclusion and future work directions.

II. Theoretical Background

Educational software could be considered as specially created software to support the process of teaching and learning. There are many categories of educational software, such as Content Management Systems, Simulation tools, Multimedia Presentations, Interactive games etc. Educational software could be used for both teaching of theoretical concepts or practical learning with illustrations, experiments, simulations, transformations etc. Special area of interest for creating educational software is focused on information technology teaching and learning, particularly in higher education.

Contemporary software development is usually based on distributed software architectures, including remote access to data, distant processing and integration of services, components and layers. Distributed information systems represent a sub-type of "*Distributed systems*" which place special emphasis on distributed software components (distributed processing) and distributed data (distributed databases, distributed files and formats) structures and processing.

Techniques with distributed databases include database partition, distributed databases catalog processing, data synchronization, data replication, data recovery, distributed transactions and others. Research [2] presents the most popular relational database management systems (RDBMS), such as MySQL, MS SQL and Oracle support to all essential techniques needed for distributed databases processing.

Database partition is a physical separation of a database table structure or data in multiple databases, while preserving logical completeness. There are two commonly used partitioning types:

- Vertical partitioning - A relational database could have a table with complex structure (large number of data fields) and the structure could be divided into logical units, physically located in different databases. These units have the same table name and the same primary key field, while other fields are grouped according to the criteria for logical coherence of fields.

- Horizontal partitioning – each partition have the same structure of tables and equal table names, but the data records are different. In the data entry, according to certain criteria, data record is categorized and sent for saving into appropriate horizontal partition.

In aim to manage all partitions, special catalog is formed to keep and maintain data about the partitioned databases, their locations, names and type of partition and data stored. The catalog could be part of the DBMS i.e. part of special DBMS created for distributed databases (Distributed DBMS – DDBMS) or a separate file that keeps all previously mentioned relevant data.

Data replication performs copying the data structure and data records, in aim to provide backup data. Database partitions should have synchronized data, i.e. there should not be missing or duplicate data. Important technique that enables data alignment between database partitions is distributed transactions, having two-phase commit. This provides any data entry/update/delete operation to be performed at all partitions and only if all partitions have successful transaction completed (first partial commit), global transaction commit (second and final commit) is finalized.

Data recovery enables retrieving missing data in a database, which is particularly important in alignment of data between database partitions. In aim to enable tracking of data records (their existence, data of insert/update...), there are several techniques, but most often used is including time stamp data field in the structure of a data table. Every time a data record is saved, DBMS creates timestamp value and inserts in appropriate timestamp field. If a data record is missing or deleted, there will be no data record with appropriate value of time stamp. This could be used for determining which records should be used for database recovery.

III. RELATED WORK

A. Distributed Information Systems as a Higher Education Infrastructural Element

Information systems represent one of organizational infrastructure components, supporting data flow and exchange, collection, storage and processing, in aim to enable operative and management activities. There are many areas of application and academic institutions are one of them [3]. "The use of information systems in higher educational institutions such as universities has grown in recent years because of the enhanced capabilities of information systems in handling complex information in a sophisticated manner for the purpose of querying, retrieving and presenting information to advanced users. These academic information systems are reshaping the dialogue on knowledge sharing and creation in institutes of higher education." [3]

Distributed systems are used as a basis for distance learning. [4] "A distributed system is a system designed to support the development of application and services which can exploit a physical architecture consisting of multiple, autonomous processing elements that do not share primary memory but cooperate by sending asynchronous messages over a communication network." [4][5] There are several reasons for using distributed systems in distance education, but also in other applications as well. According to [4], these are: resource sharing, availability of content and computing resources, extensibility (scalability) to enable availability and improve performance.

B. Distributed information systems higher education – teaching content, methods and tools

Since information system (IS) has a complex structure (consisting of software, hardware, orgware and lifeware), it is of a special interest to include IS in curriculums, starting with bachelor studies and supporting it also at higher level of studies. Very often, bachelor level of studies include courses named "Management Information Systems" (MIS) [6], emphasizing that not only data collecting, storing and exchange are important, but also data processing, as a support to management, i.e. making decisions.

Distributed information systems-related courses at higher education have drawn significant attention in recent years [7]. In aim to explore how increasing interactivity with students could improve students learning outcome, Cano [7] investigates the effect on students of moving from a traditional learning process, based on lectures and laboratory work, to an approach closer to continuous evaluation. It was based on various types of weekly assessments included in a Distributed Information Systems course. Analysis of the results (7 years, 750 students in the research sample) shows that most students preferred to participate in the course following the new methodology, the pass rate and the final students' grades improved, the percentage of students dropping out of the course increased. At teacher's evaluation, scores were slightly lower with the new methodology, even though the students' generally performed better. [7]

Modern industrial environment influences utilization of contemporary technologies into teaching and learning process. Recent technologies in distributed information systems include cloud-based solutions, as well as using blockchain technology. [8] "The authors examine conditions for implementing distributed information systems based on blockchains in higher educational establishments. The authors also define the types of educational services which can be automated by using blockchains." [8]

In recent period (for more than decade), there are many improvements made in the teaching methods and tools generally, by including information technologies. Special types of software have been created to help learning and teaching process – educational software. E-learning as a paradigm is deployed not only with specially created educational software, but with general communication platforms, most commonly utilized in business environments. Particular research focus in MIS learning support has been put on e-Learning software solutions, as presented in [6].

There are many initiatives for maximizing benefits of using Internet and web applications in teaching and learning in higher education. The expanding population of students and increased demand for flexible education direct development of suitable web-based educational software. Web-based, interactive simulations are created for the educational environment and presented in research [9].

Using remote emulated laboratory in Distributed Information Systems Course (included in the Telematics Engineering degree) has been proposed and described in [10]. The benefits and weaknesses of the remote emulated laboratory experience are assessed by the students' feedback through anonymous questionnaires, and by the students' final grades. Results show students are highly motivated to use this method/tool for learning and they did not find any difficulties in the process. Students' attitudes towards new lab approaches were very positive, and success in the final written exam improved almost 30%. Nevertheless, remote emulated labs do not completely cover all benefits of the hands-on labs, so both types should be integrated. [10]

Higher education is mostly oriented towards teaching students in aim to develop their knowledge and skills needed for their professional environment. There is also a significant segment of adult teaching, where special professional courses are designed for the employed personnel from some institutions or industry, to enable their knowledge and skills advancements, according to new technologies. In paper [11], special software tools were designed in aim to assist adult courses attendees in learning new distributed information systems-related contents.

C. Distributed Database Recovery

The problem of database recovery has been addressed for many decades in professional and educational environments [12]. Initial techniques for database recoveryt included transactions processing [13]. Having data retrieval performances in focus, one of solutions directed development towards keeping the whole database in a main memory, which introduces the problem of database recovery at memory level [14]. With advancements of information technologies, client/server [15] and online [16] systems brought new insights in the problem. Distributed data processing introduced the concept of distributed databases, with partitioning. Keeping all database partitions up-to-date was put in the context of databases recovery[17] [18]. Using and maintenance of database catalog (journal) [19], which keeps record on all partitions, came into the focus of databases recovery, as well.

Other approaches to database recovery were directed towards methodology-related issues, i.e. how to perform the recovery (e.g. by using online mirror database [20]). performance metrics [21] and acceleration [22], formalization and automatization of the recovery process [23], as well as recovery cost estimation [24].

Finally, with the introduction of mobile technologies, it is of a special interest to support database recovery techniques for working environments, such as Android [25].

IV. Educational software DIS@TFZR

A. Functional Design

Educational web-based software DIS@TFZR functions are presented at use-case diagram (Figure 1) and they include:

- *Data entry* at web form, where data validations are performed (basic validations of presentation logic and business-rules enforcement), as well as data storing in partitioned databases, according to the partitions defined in the distributed databases catalog. It also includes processing database transactions upon all partitioned databases, in case of complex database structure (more than 1 data table in a database).
- *Tabular data presentation* at web form, where partitioned data are merged into a single memory-level data collection, in aim to have data presented to a user as if from a single centralized database,
- *Data checking* i.e. control of all database partitions for the alignment of data. All necessary data should be in appropriate database partition and there should not be missing data records.
- *Data recovery*, with return of missing data records in partitioned databases that had temporary non-functional periods and did not receive all data records, as other database partitions.

Particular domain that has been supported in this example is data recording and processing related to preschool children.

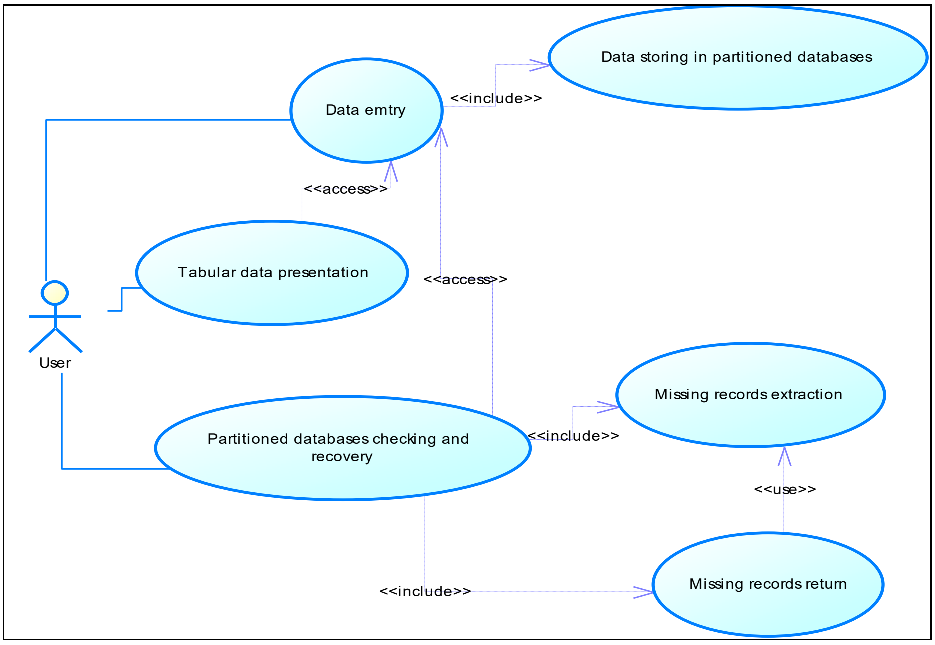


Figure 1. Use case diagram with software functions of the educational software DIS@TFZR

B. *User Interface Design*

Web application has been designed as a typical ASP.NET web-forms application, with a Site.Master page containing main menu a design model for each web page containing the same master page. Figure 2. presents web page for data entry.

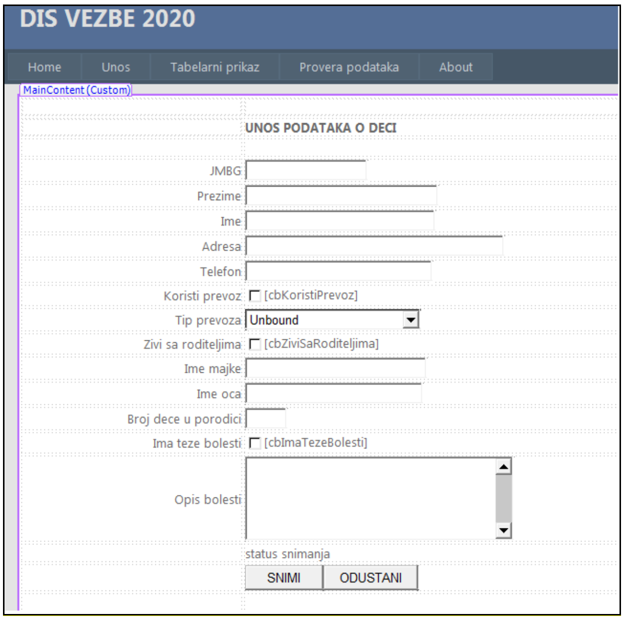


Figure 2. Data entry web form

Figure 3 presents web page for tabular presentation of integrated data from horizontal data partitions

Third most important part is data checking and recovery web page. It has been added in 2019/20 to the DIS@TFZR software and will be presented in a separate section in this paper.

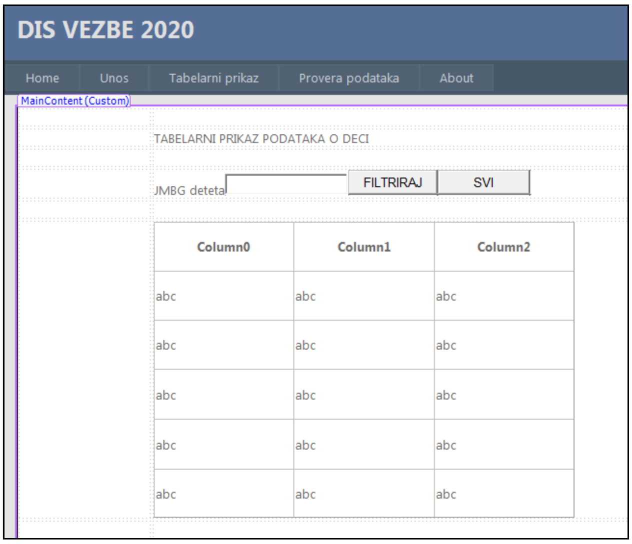


Figure 3. Tabular presentation of integrated data – user interface design of a web form

C. *Implementation*

Aim of creating the software DIS@TFZR is to illustrate different elements and functionalities of a distributed information system, as well as to present different software architectures and their integration: layer-oriented, service-oriented etc. Figure 4. presents a component diagram of DIS@TFZR software.

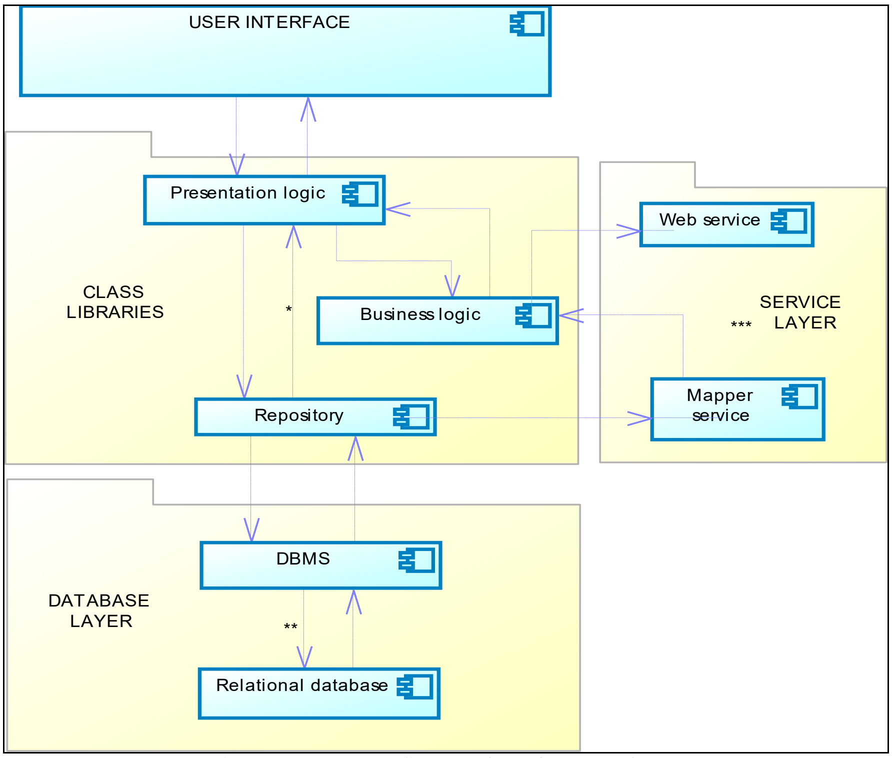


Figure 4. Component diagram of DIS@TFZR software

Components of DIS@TFZR include user interface (ASP.NET project), class libraries projects, service layer with two type of projects (web service – service as additional functionality and mapper class library - service as an integrating layer for mapping data between other layers). Class libraries have separate duties: presentation logic (to prepare data for presentation in the user interface or to validate data taken from user interface), business logic (to implement business rules that are enforced to be checked from presentation logic), and repository classes (to implement CRUD operations over database).

V. Database Partitioning in DIS@TFZR Software

Crucial elements for implementation of data partitioning are presented at Figure 5. Web.config file keeps all connection strings for all partitioned databases, with an assigned alias name. The connection string alias name, together with the type of partition, is stored in catalog.XML file. This file is loaded in the software when the data storing, checking and recovery is performed.

Within the implementation of database partitioning, there should be defined the criteria for the data separation, during data saving in particular database partition. In this educational example, criteria are for:

1. Horizontal partitioning – healthy children, children with health-related problems (only one partition gets the whole record)
2. Vertical partitioning – personal data, family-related data, accommodation data, travel data. (Each record is saved, but it is divided by logical units into vertically partitioned databases and could be combined into a single record, since they all have the same primary key value. In this case, it is personal ID, i.e. JMBG).

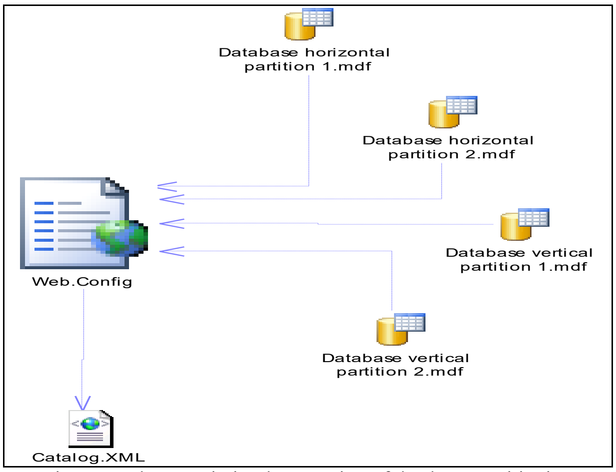


Figure 5. Elements in implementation of database partitioning

Tabular presentation of data is implemented as merging data from different horizontal partitions into a single memory-based data collection. Crucial code for the merging is based on creating a central DataSet, performing copy operation to get the structure of a dataset (taking the structure of any of the horizontal partitions) and merging data from both (all) horizontal partitions.

```
System.Data.DataSet dsCentral;
dsCentral = ds1.Copy();
dsCentral.Merge(ds2);
```

Listing 1. Key code segment for merging data from different data sources - horizontal partitioned databases

VI. Database Recovery in DIS@TFZR Software

Database recovery implemented in this example software was based on the central database that contains all data successfully saved in the system. At the same time, parallelly with the central database, data are stored into horizontal and vertical partitioned databases as well.

An algorithm of checking and recovery is described as follows:

1. Determines which databases do not contain all necessary records and which these missing records are. Time stamp data help in this extraction. Every time a record is saved in the central database, time stamp is automatically assigned to a record. Other, partitioned databases get the same number into a simple integer field, intended to be paired with the timestamp filed in the central database. The extraction part of algorithm counts the total number of records and compares it for each database (central and partitioned databases). Then, it finds the last value of time stamp for central and partitioned databases and compares them as well. Final result of extracting necessary data to be added is creating list of strings having insert into queries to be processed in aim to recover databases. Insert into query data include the missing rows having time stamp values that are not present in the database.
2. Presents a report which databases have missing data and how many records are missing.
3. Recovers partitioned databases, by adding appropriate data. It is performed by transaction that uses a list of SQL insert into queries and processes them as a package. For each partitioned databases there could be one or many records missing. Therefore, for each database there could be a list of insert into queries that will recover missing records.

Example of using the checking and recovery is presented at following figures. Checking the consistency of all partitions with the central database, when all partitioned databases have all expected data, according to the content of central database is presented at Figure 6.

In experimenting with the software, intentional removal of some records in selected partitioned databases is performed, as presented at Figure 7.

In case of having any (one or more) partitioned database not consistent with the central database, there will be report about the problems and, immediately after that, the software will automatically fix these problems by adding the missing records. After checking and recovery of data, report is presented at Figure 8.

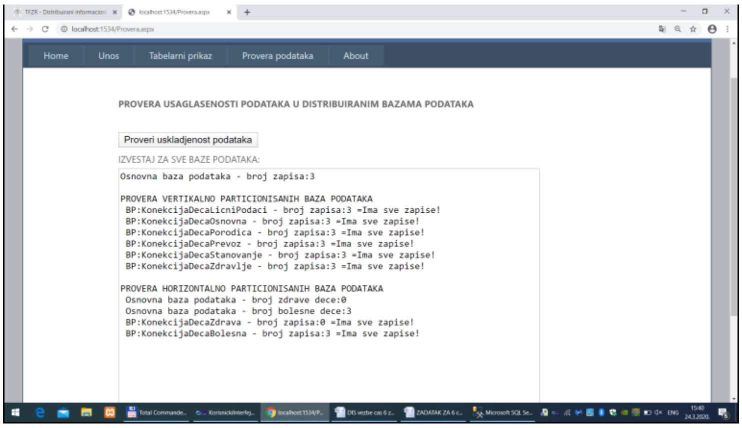


Figure 6. Successful checking report

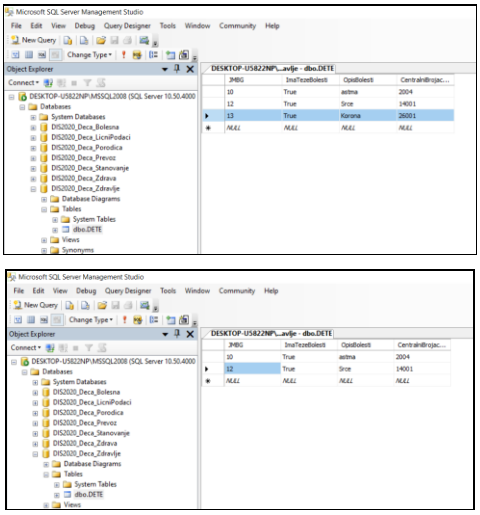


Figure 7. Intentional removal of data in database (before and after deleting a data record)

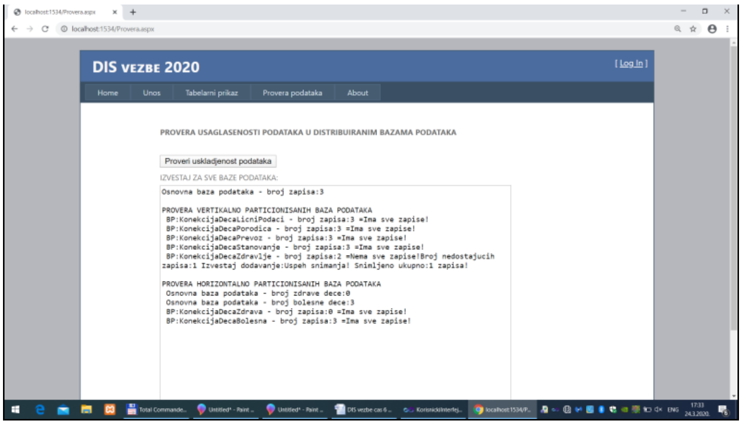


Figure 8. Report after checking and recovery of partitioned databases

VII. Conclusion

In aim to support students practical activities in teaching process of Distributed information systems course (at master level studies of Information Technologies at University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin., Serbia), an educational web-based software was created by Ljubica Kazi. It has been used for several years as a model in the classroom work within practical classes, as well as within examination (at mid-term practical knowledge evaluation and in creating students' projects).

Aim of this paper was to present the concept of the created software solution, with functional and structure aspect description. Particular emphasis has been put on new functional elements implemented during teaching process in 2019/20 year.

It is interesting that the new module, related to partitioned databases mutual consistency checking and recovery, has been developed partially at regular classes (by mentor/teaching staff member Ljubica Kazi in interaction with master level students), while finalizing the software was performed by Ljubica Kazi, preparing the material for on-line lecturing during Covid-19 lockdown.

Created educational software DIS@TFZR will remain the model of the software that students could use in future learning process at the same course, but also it could be used for other courses at undergraduate and master level.

The software could be enhanced with additional functionality, such as: support to updating and deleting of data and appropriate checking/recovery of partitioned databases, as well as more complex business logic (having business objects implement different kinds of business entities, such as documents, procedures, rules). It could be also used to develop different versions, with refactoring the solution to other types of software architectures, creating or using design patterns and frameworks etc. Currently, the focus of the course Distributed information systems has been targeted and the goal has been reached. Future improvements and work will direct the creating other versions of software, which could be a beneficial model for other courses as well.

References

[1] Bourn D, Neal I: "*The Global Engineer – Incorporating global skills within UK higher education of engineers*", Institute of Education, University of London, 2008.

[2] Hrnjaz M: "*Distributed Databases Technologies*", Master thesis, menthor Ljubica Kazi, University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Serbia, 2017.

[3] Alyoubi, B.A, Alyoubi, A.A: "*Impact of Academic Information Systems on the Performance of the Higher Education Institutions*", International Journal of Science and Applied Information Technology (IJSAIT), Vol.2 , No.2, Pages : 32-37 (2013) Special Issue of ICET4E 2013 - Held during 11-12 March, 2013 in Hotel Crowne Plaza, Dubai.

[4] Prasolova-Forland E: DIF8914 Distributed Information Systems, Distance Learning: overview and design issues, The Norvegian University of Science and Technology, 2003.

[5] Blair, G. S, Stefani, J. B. *Open distributed processing and multimedia*. Addison Wesley Longman Ltd, 1998.

[6] Aleksic-Maslac K, Vasic D, Korican M: "*Student Learning Contribution through E-learning Dimension at Course 'Management Information Systems*;", WSEAS Transactions on Information Science and Applications, ISSN 1790-0832, Issue 3, Volume 7, March 2010.

[7] Cano M-D: "*Students' Involvement in Continuous Assessment Methodologies: A Case Study for a Distributed Information Systems Course*", IEEE Transactions on Education (Volume: 54, Issue: 3, Aug. 2011), p. 442 – 451.

[8] Zaiarnyi O.A., Zaiarna I. S.: "*The Application of Distributed Information Systems Based on Blockchain Technologies in Higher Education: Benefits and Challenges for Ukraine*", 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT)

[9] Masson, A.M.: "*Web-based simulations for computer-assisted learning in the higher education sector*", Engineering Science & Education Journal,Volume 8, Issue 3, June 1999, p. 107 – 114.

[10] Cano M-D: "Using Remote Emulated Laboratory in a Distributed Information Systems Course", 40th ASEE/IEEE Frontiers in Education Conference, October 27-30, 2010. Washington, DC

[11] Karlovic G.V, Grigoryev V. K: "Advanced Training Technology of Mass Professional Users of Distributed Information Systems" (in Russian), Informatization of education and science, No 4, (16), 2012, pp. 183-195.

[12] Kumar V, Son S.H: Database Recovery, Springer Science and Business Media, 1998.

[13] Hearder T, Reuter A: "Principles of Transaction-Oriented Database Recovery", Computing Surveys, Vol 15, No 4, 1983.

[14] Eich M. H: "A classification and comparison of main memory database recovery techniques", IEEE Third International conference on Data Engineering, 1987.

[15] Mohan C, Narang I: "ARIES/CSA: a method for database recovery in client-server architectures", SIGMOD '94: Proceedings of the 1994 ACM SIGMOD international conference on Management of data, May 1994, pp. 55–66.

[16] Moore D.W et al: "Database recovery to any point in time in an online environment utilizing disaster recovery technology". US Patent - US 6.732,123 B1, 2004.

[17] Hsiao H-I, Chang A, "Method for providing database recovery accross multiple nodes", US Patent - US 6.247.023 B1, 2001.

[18] Dhamankar et al: "Consistent database recovery accross constituent segments", US Patent – US 8.671.085 B2, 2014.

[19] Nishikawa et al: "Database recovery method applying update journal and database log", US Patent – US 7.471.139.B2, 2008.

[20] Hart, D.R: "System and method for database recovery using a mirrored snapshot of an online database", US Patent – US 6.983.295 B1, 2006.

[21] Pommerenk et al: "Performance metric-based selection of one or more database server instances to perform database recovery", US Patent – US 7.809.690 B2, 2010.

[22] Stagg M.R: "Apparatus and method for acceperating database recovery", US Patent – US 7.865.471 B1, 2011.

[23] Gurevich Y, Soparkar N, Wallace C: "Formalizing database recovery", Journal of Universal Computer Science, vol.3, no 4, 1997.

[24] Sauer C, Graefe G, Harder T: "An empirical analysis of database recovery costs", RDSS, ACM, 14. June 2014, Snowbird, UT, USA

[25] Hu X, Wu H: "Database management strategy and recovery methods of Android", IEEE 5th International Conference on Software Engineering and Service Science, 2014.

Functional Usability of Web Services Available on Public Repositories

J. Adamov, Lj. Kazi and D. Radosav
University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
jelena.adamov1586@gmail.com, ljubica.kazi@gmail.com, radosav@tfzr.uns.ac.rs

Abstract –Web Services are available over the Internet and anyone can access and use them, by integrating their functions into other web applications. Special web sites host links to publicly available web services and are addressed as public web services repositories. This paper presents results of the research of functional usability of Web Services that are available on public repositories. Particularly, we will analyze the ratio of valid .asmx links comparing to those that are not functional (not working). Special emphasize has been given to application domains of functional Web services.

I. INTRODUCTION

Web services are a natural consequence of the evolution of the Web into an open medium, that facilities complex business and scientific application interactions. Web services enable systematic application-to-application interactions over the Web, and the integration of the existing network computer infrastructure into the Web. [4]

Web services represent a new paradigm in application architecture and development. The importance of Web services is not that they are new, but this technology addresses the needs of modular application development, interoperability and interinstitutional integration. [1] [2]

They are available via the World Wide Web and allow anyone to access it. They allow various applications to "talk to each other" and share data and services among themselves. Web services are used to make the application platform and technology independent. [1]

Aim of this paper is to present results in empirical study of public web services availability. By accessing public web services repositories, it is essential to test whether the web service URL provides functional (working) web service or it does not exist or have malfunction.

The rest of the paper is organized as sections: theoretical background, related work, research methodology, results, and conclusion.

II. THEORETICAL BACKGROUND

A. *Web service*

"A Web service is an interoperable machine-to-machine software function that is hosted at a network addressable location. A Web service is described using a standard, formal XML notation, called its service description. It covers all the details necessary to interact with the service, including message formats, transport protocols and location." [3]

"It has hidden implementation details so that it can be used independently of the hardware or software platform on which it is implemented, and independently of the programming language in which it is written." It can be used alone or with other Web services to carry out a complex aggregation or a business transaction. [3]

B. *ASMX file*

Active Server Method Extended - ASMX is a file extension of Web services due to which Web services are also called ASMX Web services.

ASMX Web services send messages using the Simple Object Access Protocol (SOAP). SOAP is a platform-independent and language-independent protocol for building and accessing Web services. Anyone who uses ASMX service does not need to know how to send and receive SOAP messages, nothing about the platform, object model or programming language used to implement the service.

An ASMX file servers are the end point for an ASP.NET Web service. XML (Extensible Markup Language) is used as a specific data format that formulates the result of Web Service processing. This way the user software can get and "understand" the result and easily integrate it within internal functionality. An ASMX service can only operate over HTTP. They can be run using a Windows-based Web server with .NET framework. [5]

```
[WebMethod]
public int Add(int a, int b)
{
  return(a + b);
}
[WebMethod]
public System.Single Subtract(System.Single A, System.Single B)
{
  return (A - B);
}
[WebMethod]
public System.Single Multiply(System.Single A, System.Single B)
{
  return A * B;
}
[WebMethod]
public System.Single Divide(System.Single A, System.Single B)
{
  if(B == 0) return -1;
  return Convert.ToSingle(A / B);
}
```

Figure 1. Example of simple methods in Web service [13]

C. *Public repository*

A repository is like a folder for a project. Project's repository contains all of project's files and stores each file's history. It allows discussing and managing project's work within the repository.

Repositories can be owned individually or ownership can be shared with other people. Access of a repository can be restricted by choosing the repository's visibility – public or private. Anyone can access public repositories on the Internet. [6]

III. RELATED WORK

Web services are gradually becoming a dominant form of granular components used to build applications in today's business IT systems. If any involved Web services become unavailable for some reason, the whole business process appears unavailable. Sustainable success relies on capabilities to manage possible service failures. [7]

Authors of research [8] claim that there are two kinds of requirements that are crucial to Web service composition, which are functional and non-functional requirements. Functional requirements focus on functionality of the composed service and nonfunctional requirements are concerned with the quality of service. Nonfunctional requirements are important to Web service composition, and are often an important clause in service-level agreements. Even though the functional requirements are satisfied, a slow or unreliable service may still not be adopted. Because of that, they proposed an automated approach to verify combined functional and nonfunctional requirements directly based on the semantics of Web services composition. [8]

Special aspect of web services quality is related to their availability. Little work has been reported on highly available Web services which are essential for mission critical applications, because companies are not willing to make their business critical applications dependent on external Web services. Important non-functional requirements related to security, reliability and availability are addressed. Not much has been written about how to ensure that a Web service is available. In paper [9], authors proposed a solution for enhancing Web services availability. The central idea they suggested was the enhancement of Web services by the introduction of a central hub. [9]

Availability is the probability that a Web service is in functioning condition at a specific time. In paper [10], authors study how to sustain the high availability and quality of Web services by using communities of Web services. Communities of Web services gather Web services that provide the same functionality, but not necessary with the same quality. An available Web service can be selected or even substituted by another one when it fails. [10]

One study about bioinformatics Web services has shown that out of all available services at that moment, 45% had positively confirmed their functionality, while 9% of services were completely unavailable. Next, 33% of them could not be tested because there was no example data or related problem and 13% were not working as expected. Additionally, a survey conducted among 872 Web Server Issue corresponding authors and 274 of them replied. 78% of all respondents indicated that their services have been developed solely by students and researchers without a permanent position. These services are in danger to disappear (or malfunction due to bad or missing maintenance) after the original developers move to another institution, and for 24% of services there is no plan to maintenance, according to respondents. [11]

High availability of Web services is a significant challenge due to the varying number of invocation requests the Web services have to handle at a time, as well as the dynamic nature of the Web. The issue becomes even more challenging for composite Web services in the sense that their availability is inevitably dependent on corresponding component Web services. In paper [12], these problems are tackled by exploiting particle filtering-based techniques. In particular, algorithms to accurately predict the availability of Web services and dynamically maintain a subset of Web services with higher availability ready to join service compositions. Web services can be always selected from this smaller space, thereby ensuring good performance in service compositions.[12]

In a highly dynamic system environment, services that can produce similar or complementing results need to be quickly identified in a number of scenarios. To substitute a failed service with some other equivalent service is unavoidable in recovering a suspended application due to failure of a constituent service. In paper [7], authors report a rule based approach to Web service substitution in order to secure availability of services. Availability provides delivery assurance for each Web service so that Simple Object Access Protocol (SOAP) messages cannot be lost undetectably, especially in a Web service composition. The rules are written in Semantic Web Rule Language. The rules are a formal representation of a categorization-based scheme to identify exchangeable Web services. Experiments on service substitution based on the proposed framework achieve a best precision of 85%. [7]

IV. Research Metodology

Research objective is to determine the level of functional Web services available on public repositories.

Research hypothesis is: "More than 70% of existing ASMX Web services on public repositories are not functional."

Research sample consists of publicly available web services of ASPX type. The research is conducted through testing and validation of randomly selected .asmx links from public repositories. ASMX URL links were collected from 3 public repositories [14] [15] [16]. After collecting URL links from public repositories, research sample consists of 100 randomly selected links.

Research method is applied upon publicly exposed URLs from public web services repositories by: collecting URL links, categorization of URL links according to description of the web service use, testing the functionality of the link by clicking on the URL and categorization of testing results (functional, not existing/not functional), further testing of functional URLs by invoking their methods.

V. Results

Validation testing included 100 random selected links. Throughout testing links, numbers .asmx links from various areas were included: Stock market, Banking, Advertising, Databases, Communication, Internet services, Zip codes and Address validation, Weather Info, Finders and Locators, and Search Engines. The most percentage of tested links are from the field of Communication – 17%, and the least percentage of tested links are from the field of Weather Info – 3%. In the figure down below are represented detail information about the ratio of tested links from all the fields.

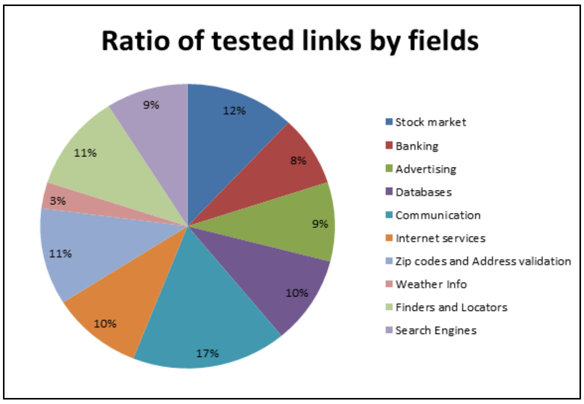


Figure 2. Ratio of tested .asmx links by fields

After testing all of 100 links, the results of research confirms an earlier hypothesis. The number of valid .asmx links is very small, only 9% of all tested links, which means that there is over 70%, more precisely 91%, of not functional Web services on public repositories.

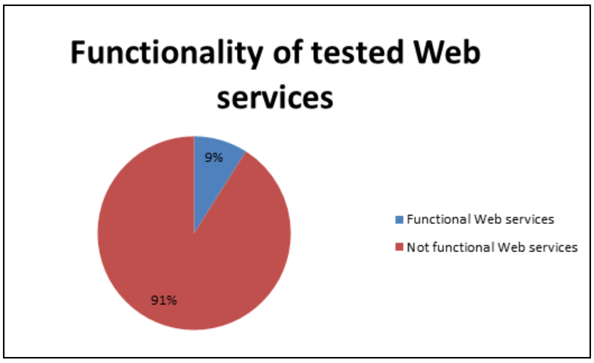


Figure 3. Results of tested Web services

Only functional .asmx links from random selection are from these fields: Advertising – 2 links, Communication – 3 links, Internet services – 3 links, and Zip codes and Address validation – 1 link. That is 9 out of 100 random selected links.

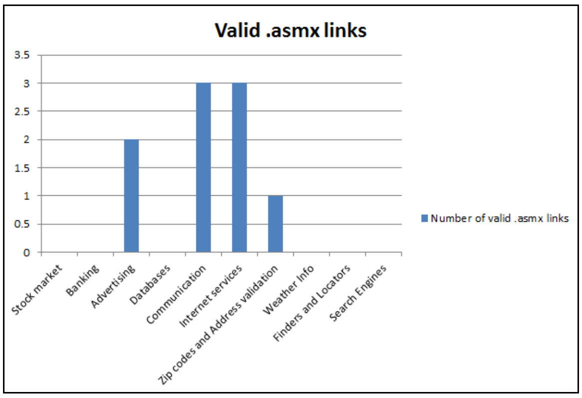


Figure 4. Valid .asmx links

Figure 5 shows an example of a functional Web service (in this case: a Calculator).

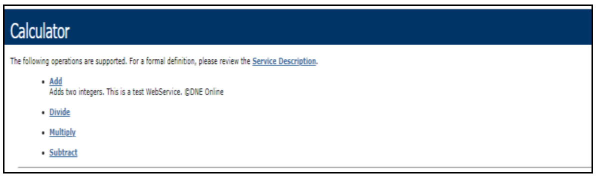


Figure 5. Functional Web service

Non-functional .asmx links lead to nonexistent pages and message: “Page not found”. It is shown in a figure down below.

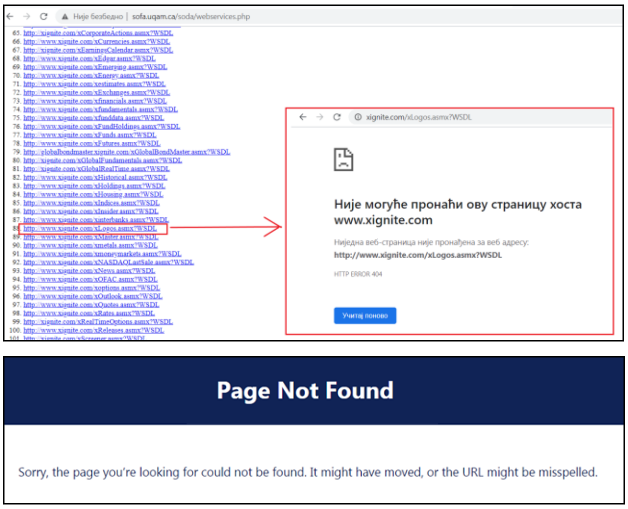


Figure 6. Non-functional Web service (“Page not found” response)

VI. Conclusion

Web services are available via the Web and anyone can access them. They allow applications to share data and services among themselves. There is no need to know anything about programming languages or platforms to use them.

This research of 100 randomly selected links of Web services, available on public repositories, showed that small number of them is functional. Only 9% of them is valid and can be used, while 91% of links are unavailable.

Although, most of tested links are from the field of communication (17), only 3 of them are valid. The same is with Internet services – 3 valid links, followed by Advertising – 2 valid links, and Zip codes and Address validation – 1 valid link.

The set hypothesis was confirmed, that over 70% of existing Web services on public repositories are not available.

Future research in this field could be directed towards expanding areas of included links, or towards one specific area (communication, for example), and analysis of available Web services on public repositories that belong to that category.

References

[1] S. Chatterjee, J. Webber: Developing Enterprise Web Services: An Architect’s Guide, 2004.

[2] Lj. Kazi andcoauthors: Fundamentals of Industrial Software development, Technical Faculty “Mihajlo Pupin” Zrenjanin, 2018.

[3] IBM, What is a Web service, URL: http://www.ibm.com/support/knowledgecenter/SSGMCP_5.2.0/com.ibm.cics.ts.webservices.doc/concepts/dfhws_definition.html [accessed: June 20, 2020].

[4] F. Curbera, W. A. Nagy, S. Weerawarana: Web Services: Why and How, 2001.

[5] Consume an ASP.NET Web Service, URL: http://docs.microsoft.com/en-us/xamarin/xamarin-forms/data-cloud/web-services/asmx [accessed: June 20, 2020].

[6] GitHub, About repositories, URL: http://docs.github.com/en/github/creating-cloning-and-archiving-repositories/about-repositories [accessed: July 18, 2020].

[7] Q. A. Liang, H. Lam, L. Narupiyakul, P. C. K. Hung: A Rule-Based Approach for Availability of Web Service, 2008.

[8] Chen M., Tan T.H., Sun J., Liu Y., Pang J., Li X. (2013) Verification of Functional and Non-functional Requirements of Web Service Composition. In: Groves L., Sun J. (eds) Formal Methods and Software Engineering. ICFEM 2013. Lecture Notes in Computer Science, vol 8144. Springer, Berlin, Heidelberg

[9] S. Abraham, M. Thomas, J. Thomas: Enhancing Web services availability, IEEE International Conference on e-Business Engineering (ICEBE'05), 2005.

[10] E. Lim, P. Thiran: Sustaining High-Availability and Quality of Web Services, 2010.

[11] S. J. Schultheiss, M. C. Munch, G. D. Andreeva, G. Ratsch: Persistance and Availability of Web Services in Computational Biology, 2011.

[12] L. Yao, Q. Z. Sheng, Z. Maamar: Achieving High Availability of Web Services Based on a Particle Filtering Approach, ACM, 2012.

[13] Microsoft, Write a Web Sevice, URL: http://docs.microsoft.com/en-gb/troubleshoot/dotnet/csharp/write-web-service [accessed: June 20, 2020].

[14] A summary of public web services, URL: http://www.ai.soc.i.kyoto-u.ac.jp/sw/Web_Services_List.html [accessed: June 27, 2020].

[15] API Directory, URL: http://www.programmableweb.com/category/all/apis [accessed: June 27, 2020].

[16] The List of Web Services, URL: http://sofa.uqam.ca/soda/webservices.php [accessed: June 27, 2020]

Channel Capacity of Macrodiversity System in Gamma Shadowed k-μ Fading Environment

Dragana S. Krstic*, Suad N. Suljovic*, Nenad N. Petrovic*, Zoran J. Popovic** and Sinisa Minic***
* Faculty of Electronic Engineering, University of Nis, Nis, Serbia
** Name of Institution/Department, Higher Technical School of Professional Studies, Zvecan, Serbia
*** Department of Informatics and Mathematics, Faculty of Teacher Education in Prizren-Leposavic, University of Pristina-Kosovska Mitrovica, Serbia
E-mails: dragana.krstic@elfak.ni.ac.rs, suadsara@gmail.com, nenad.petrovic@elfak.ni.ac.rs, z.popovix@live.com, sinisa.minic@pr.ac.rs

Abstract - In this article, the channel capacity (CC) of the macrodiversity (MD) system consisting of selection combining (SC) macrodiversity receiver and two maximal ratio combining (MRC) microdiversity (mD) receivers in Gamma shadowed k-μ fading environment is determined. Both microdiversity receivers are with *L* input branches. The analytical expressions for the probability density functions (PDF) of the signal to noise ratio (SNR) at the output of the mD MRC and MD SC receivers are presented. Then, the channel capacity of the MD SC receiver is calculated for described system model. The results are presented by some graphs to show parameters influence to observed system performance.

I. Introduction

Channel capacity (CC) is among the most important performance measures of the system. The Shannon capacity of a wireless channel defines its theoretical upper bound for the maximum rate of data transmission at an arbitrarily small bit error probability (BEP), without any delay or complexity constraints [1].

In this paper, the CC of selection combining (SC) macrodiversity (MD) system with two *L*-branches maximal ratio combining (MRC) microdiversity (mD) receivers will be derived. Here, MRC mD receiver is used to reduce the influence of multipath fading and SC MD combiner mitigates the effects of shadowing [2]. We will derive for a shadowed *k*-μ fading channel a closed form expression for CC at the MRC mD receivers' outputs and SC MD receiver output. The consideration of κ-μ shadowed distribution is of great interest because it contains other classical distributions like: Rayleigh, Rician, Nakagami-*m*, one-side Gaussian shadowed distributions as special cases.

In practice, different diversity techniques can be used to reduce the impact of fading on system performance. Diversity techniques combine the sent replicas of the same data signal. The most often used diversity techniques are maximum ratio combining (MRC), equal gain combining (EGC), switch and stay combining (SSC) and selection combining (SC) [3].

Maximal-Ratio Combining (MRC) is widely used diversity combining scheme and an optimal combining model. MRC receiver can restore a signal to its original state. Also, MRC is known as ratio-squared combining. MRC requires cognition of all channel fading parameters. Because of that it has the highest complexity and cost.

The next by performance is Equal Gain Combining (EGC) and then selection combining shemes such as Selection Combining (SC) and Switch and Stay Combining (SSC). The last two have poorer performance, but are simpler for implementation and practical realization since do not require signal cophasing and fading envelope estimation. Consequently, they are very often implemented in practice. Therefore, SC receiver chooses the branch with the highest signal-to-noise ratio and processes it further. If the noise power is the same in all SC receiver branches, then SC combiner determines the branch with the strongest signal for prossesing.

In earlier published papers, CC is analyzed for some scenarios. So in [4], moment generating function (MGF) based expressions for effective capacity over generalized fading channels are obtained by utilizing Fox's H function [5], without the need of computing the integrals. These expressions are applicable for a wide range of fading distributions.

The channel capacity of MRC combiner per unit bandwidth in exponentially correlated Nakagami-m fading environment for optimal power and rate adaptation policy was derived in [6] in closed form expressions as finite series. Also, the closed form expressions for outage probability are obtained. An influence of fading severity, level of correlation and diversity order is analyzed via plotted graphics.

The paper [7] deals with the capacity analysis of a dual-branch switch and stay combining (SSC) system operating over k–μ and η–μ fading channels. There, expressions for the CC of these two fading distributions are derived under three adaptive transmission techniques containing infinite series. The corresponding expressions for Nakagami-m, Nakagami-q and Rician fading are presented as special cases of k–μ and η–μ fading.

The effective capacity of communication systems over generalized κ-μ shadowed fading channels is investigated in [8]. An analytical expression for the effective capacity is derived via extended generalized bivariate Meijer's G function [9]. The impact of the system and channel parameters on the effective capacity was seen intuitively through closed-form expressions for the effective capacity in the asymptotically high signal-to-noise ratio regime.

A closed-form expression for the ergodic capacity of channels suffering fading described by the κ-μ shadowed distribution is derived in [10]. These communication channels can be found in systems where a line-of-sight component experiences shadowing, such as land mobile satellite systems, underwater acoustic communications, body centric communications, and other wireless communication applications.

The paper [11] gives analysis of Shannon capacity for SC and MRC diversity systems with independent and identically distributed branches, as well as for the channel without diversity in the presence of generalized α−κ−μ fading. Numerical calculations are graphically presented for different combinations of fading parameters α, κ and μ, and their influence was checked.

These authors introduce in [12] a macrodiversity system consisting of two mD SC receivers and one MD SC receiver. Independent κ-μ fading and correlated slow Gamma fading disturb the inputs of the mD SC combiners. For such structure, analytical expression for the probability density function (PDF) of the signal at the output of the MD SC receiver, as well as the output capacity, are calculated using the well-known Meijer G-function.

Unlike that paper, our macrodiversity system has SC MD receiver with two mD MRC receivers, each with L branches. The MRC combiner was chosen as the optimal, and also general case with an arbitrary number of branches. We will obtain here the observed performance (CC) in closed form and give numerical results for illustration purposes.

Organization of this paper is through six sections. In the Section II, we introduce the system model. Following, we give PDF at the outputs of micro and macrodiversity receivers. In Section III, the derivation of channel capacity is done; further, in Section IV, some graphs show the impact of fading parameters to CC. The simulation and network planning tools are presented in Section V and some concluding remarks are given in the last section.

II. Probability Density Function of Micro and Macrodiversity System

A macrodiversity system with one SC receiver and two L-branches microdiversity MRC receivers is observed. System model is presented in Fig. 1 [13]. The input and output signals of mD and MD receivers are marked on the picture. Here, MD combiner chooses the mD with higher output signal.

This system is under the influence of Gamma-shadowed k-μ multipath fading. Physical model of the k-μ fading and the derivation of this distribution are presented in [14]. This fading is characterized by two parameters, Rician factor k and number of clusters μ.

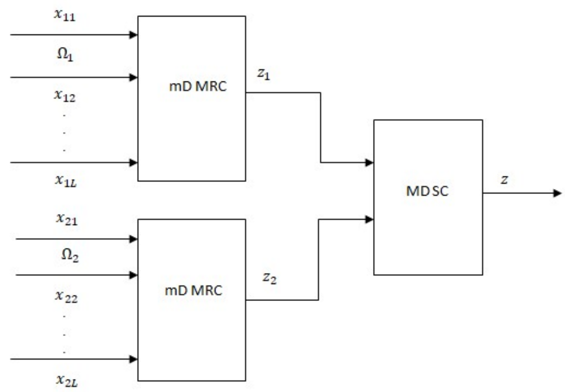


Fig. 1. Considered macrodiversity (MD) system

The PDF of the signal envelope x_{ij}, modeled by k-μ distribution is given by [15, Eq. (1)]:

$$p_{x_{ij}}\left(x_{ij}\right)=\frac{2\mu x_{ij}{}^{\mu}}{k^{(\mu-1)/2}e^{\mu k}}\left(\frac{1+k}{\Omega}\right)^{(\mu+1)/2}\cdot$$

$$\cdot e^{-\frac{\mu(1+k)}{\Omega}x_{ij}{}^{2}} I_{\mu-1}\left(2\mu x_{ij}\right)\sqrt{\frac{k\left(1+k\right)}{\Omega}}\ . \tag{1}$$

Here, k>0 is Rician factor defined as the ratio of the total power of the dominant components and the total power of the scattered waves; μ>0 is the number of clusters, Ω is the signal power mean value. $I_{\nu}(\cdot)$ is the modified Bessel function of the first kind and order ν and can be expressed by the expression [16, Eq. 9.6.20] [17, eq. (17.7.1.1)]:

$$I_{\nu}\left(x\right)=\sum_{k=0}^{\infty}\frac{\left(x/2\right)^{\nu+2k}}{k!\Gamma\left(\nu+k+1\right)}. \tag{2}$$

The MRC combiner's output SNRs z_i from i-th base station are [18]:

$$z_i = x_{11}^2+x_{12}^2+....+x_{1L}^2 = x_{21}{}^2+x_{22}{}^2+...+x_{2L}{}^2 =$$

$$=\sum_{j=1}^{L}x_{ij}{}^{2} =, i=1,2., j=1,2...L\ . \tag{3}$$

PDF of the SNR z_i is derived in [13, Eq. (5)] as:

$$p_{z_i}\left(z_i\right)=\frac{e^{-\frac{\mu(1+k)}{\Omega}z_i}}{e^{L\mu k}}\sum_{i=0}^{\infty}\frac{\mu^{2i+L\mu}\left(LK\right)^{i}z_i^{\,i+L\mu-1}}{i!\Gamma\left(L\mu+i\right)}\left(\frac{1+K}{\Omega}\right)^{i+L\mu},$$

$$z_i\ge 0\ . \tag{4}$$

The joint probability density function (JPDF) of the signal envelope average powers Ω_1 and Ω_2 at the MD SC receiver's inputs is [19]:

$$p_{\Omega_1\Omega_2}\left(\Omega_1,\Omega_2,....\Omega_n\right)=\frac{\left(\Omega_1\Omega_n\right)^{\frac{c-1}{2}}\rho^{-\frac{(n-1)(c-1)}{2}}}{\Gamma\left(c\right)\left(1-\rho\right)^{n-1}\Omega_0{}^{n+c-1}}\cdot$$

$$\cdot e^{-\frac{1}{\Omega_0(1-\rho)}\left[\Omega_1+\Omega_n+(1+\rho)\sum_{i=2}^{n-1}\Omega_i\right]}\prod_{i=1}^{n-1} I_{c-1}\left(\frac{2\sqrt{\rho\Omega_i\Omega_{i+1}}}{\Omega_0(1-\rho)}\right), \tag{5}$$

where c is order of Gamma distribution, Ω_0 is average value of Ω_1 and Ω_2, and ρ denotes correlation coefficient. The final PDF of Ω_1 and Ω_2 is [13, Eq. (12)]:

$$p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) = \frac{1}{\Gamma(c)} e^{-\frac{\Omega_1+\Omega_2}{\Omega_0(1-\rho)}} \cdot$$

$$\cdot \sum_{i=0}^{\infty} \frac{\rho^i (\Omega_1\Omega_2)^{i+c-1}}{i!\Gamma(i+c)\Omega_0^{2i+2c}(1-\rho)^{2i+c}} . \tag{6}$$

MD SC receiver selects mD MRC with higher signal level. Based on this working algorithm, PDF of SNR at output of MD SC receiver is [20]:

$$p_z(z) = \int_0^{+\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2 p_{z_1|\Omega_1}(z_1) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) +$$

$$+ \int_0^{+\infty} d\Omega_2 \int_0^{\Omega_2} d\Omega_1 p_{z_2|\Omega_2}(z_2) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) =$$

$$= 2\int_0^{+\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2 p_{z_1|\Omega_1}(z_1) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2). \tag{7}$$

By substitutions of (4) and (6) in (7), PDF is:

$$p_z(z) = 2\int_0^{+\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2 p_{z_1|\Omega_1}(z_1) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) = \frac{2}{\Gamma(c)e^{L\mu k}} \cdot$$

$$\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} \frac{\mu^{2i_2+L\mu}\rho^{i_1}(LK)^{i_2}(1+K)^{i_2+L\mu} z^{i_2+L\mu-1}}{i_1!i_2!\Gamma(i_1+c)\Gamma(i_2+L\mu)\Omega_0^{2i_1+2c}(1-\rho)^{2i_1+c}} \int_0^{+\infty} d\Omega_1$$

$$\Omega_1^{i_1-i_2+c-L\mu-1} e^{-\frac{\Omega_1}{\Omega_0(1-\rho)}} e^{-\frac{\mu(1+k)}{\Omega_1}z} \int_0^{\Omega_1} d\Omega_2 \Omega_2^{i_1+c-1} e^{-\frac{\Omega_2}{\Omega_0(1-\rho)}} . \tag{8}$$

Integral I_1 from (8) is [13, Eq. (A2)]:

$$I_1 = \int_0^{+\infty} d\Omega_2 \Omega_2^{i_1+c-1} e^{-\frac{\Omega_2}{\Omega_0(1-\rho)}} =$$

$$= (\Omega_0(1-\rho))^{i_1+c}\,\Gamma(i_1+c). \tag{9}$$

By substituting the expression (9) in (8), the PDF of the SNR is obtained as:

$$p_z(z) = 2\int_0^{+\infty} d\Omega_1 \int_0^{\Omega_1} d\Omega_2 p_{z_1|\Omega_1}(z_1) p_{\Omega_1\Omega_2}(\Omega_1\Omega_2) =$$

$$= \frac{2}{\Gamma(c)e^{L\mu k}} \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} \frac{\mu^{2i_2+L\mu}\rho^{i_1}(LK)^{i_2}(1+K)^{i_2+L\mu} z^{i_2+L\mu-1}}{i_1!i_2!\Gamma(i_2+L\mu)\Omega_0^{i_1+c}(1-\rho)^{i_1}}$$

$$\int_0^{+\infty} d\Omega_1 \Omega_1^{i_1-i_2+c-L\mu-1} e^{-\frac{\Omega_1}{\Omega_0(1-\rho)}} e^{-\frac{\mu(1+k)}{\Omega_1}z} . \tag{10}$$

Using the form [21; 3.471]:

$$\int_0^{\infty} x^{v-1} \mathrm{e}^{-\frac{\beta}{x}-\gamma x} \mathrm{d}x = 2\left(\frac{\beta}{\gamma}\right)^{\frac{v}{2}} K_v\left(2\sqrt{\beta\gamma}\right), \tag{11}$$

the integral I_2 from (10) becomes [13, Eq. (A5)]:

$$I_2 = \int_0^{+\infty} d\Omega_1 \Omega_1^{i_1-i_2+c-L\mu-1} e^{-\frac{\Omega_1}{\Omega_0(1-\rho)}} e^{-\frac{\mu(1+k)}{\Omega_1}z} =$$

$$= 2\left(\mu z \Omega_0 (1+k)(1-\rho)\right)^{\frac{i_1-i_2+c-L\mu}{2}}$$

$$K_{i_1-i_2+c-L\mu}\left(2\sqrt{\frac{\mu z(1+k)}{\Omega_0(1-\rho)}}\right). \tag{12}$$

Substituting expressions (12) into expression (10) gives PDF of SNR at output of SC MD receiver:

$$p_z(z) = \frac{4}{\Gamma(c)e^{L\mu k}} \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} \frac{\rho^{i_1}(LK)^{i_2}\mu^{\frac{i_1+3i_2+L\mu+c}{2}} z^{\frac{i_1+i_2+L\mu+c-2}{2}}}{i_1!i_2!\Gamma(i_2+L\mu)(1-\rho)^{\frac{i_1+i_2+L\mu-c}{2}}}$$

$$\left(\frac{1+K}{\Omega_0}\right)^{\frac{i_1+i_2+L\mu+c}{2}} K_{i_1-i_2+c-L\mu}\left(2\sqrt{\frac{\mu(1+k)z}{\Omega_0(1-\rho)}}\right). \tag{13}$$

III. Channel Capacity of Macrodiversity System

After we got the expression for the joint PDF, we can calculate channel capacity at the output of the macrodiversity system shown in Fig 1. The maximum data rate can be achieved after the chanel has expressed as follows in the unit of bits per second, where B denotes channel bandwith expressed in Hz. So, CC is [22], [23]:

$$C = B\int_0^{\infty} \log_2(1+z) p_z(z)\, dz . \tag{14}$$

By substituting expression (13) into expression (14), we obtain that the CC as:

$$\frac{C}{B} = \frac{4}{\Gamma(c)e^{L\mu k}} \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} \frac{\rho^{i_1}(LK)^{i_2}}{i_1!i_2!\Gamma(i_2+L\mu)}$$

$$\frac{\mu^{\frac{i_1+3i_2+L\mu+c}{2}}\left((1+K)/\Omega_0\right)^{\frac{i_1+i_2+L\mu+c}{2}}}{(1-\rho)^{\frac{i_1+i_2+L\mu-c}{2}}}$$

$$\int_0^{\infty} \log_2(1+z) z^{\frac{i_1+i_2+L\mu+c-2}{2}} K_{i_1-i_2+c-L\mu}\left(2\sqrt{\frac{\mu(1+k)z}{\Omega_0(1-\rho)}}\right) dz . \tag{15}$$

Applying the rule for the second-order Bessel function $K_v(x)$ [22, eq. (9.6.9)]:

$$K_v(z) \sim \frac{1}{2}\Gamma(v)\left(\frac{1}{2}z\right)^{-v}, \tag{16}$$

in expression (15), we get that the CC at the output of the SC combiner is:

$$\frac{C}{B} \approx \frac{2}{\Gamma(c)e^{L\mu k}\ln 2}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\frac{\rho^{i_1}(LK)^{i_2}\mu^{2i_2+L\mu}(1+K)^{i_2+L\mu}}{i_1!i_2!\Gamma(i_2+L\mu)(1-\rho)^{i_2+L\mu-c}\Omega_0{}^{i_2+L\mu}}$$

$$\Gamma(i_1-i_2+c-L\mu)\int_0^{\infty}\ln(1+z)z^{i_2+L\mu-1}dz \qquad (17)$$

Using the form [16; Eq.3, 4.293]:

$$\int_0^{\infty}\ln(1+x)x^{\mu-1}dx = \frac{\pi}{\mu\sin\mu\pi} \qquad (18)$$

we have the channel capacity at the output of the SC combiner as:

$$\frac{C}{B} \approx \frac{2\pi}{\Gamma(c)e^{L\mu k}\ln 2}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\frac{\rho^{i_1}(LK)^{i_2}}{\Gamma(i_2+L\mu)}\left(\frac{1+K}{\Omega_0}\right)^{i_2+L\mu}$$

$$\frac{\mu^{2i_2+L\mu}\Gamma(i_1-i_2+c-L\mu)}{i_1!i_2!(1-\rho)^{i_2+L\mu-c}(i_2+L\mu)\sin((i_2+L\mu)\pi)}. \qquad (19)$$

Graphs for CC, depending on the mean power Ω_0, were obtained using the software packages Mathematics and Origin.

IV. Obtained Graphs

In this section, some graphs of the system's channel capacity are presented, in order to examine the impact of fading parameters: fading and shadowing severity, number of diversity branches at the input of mD receiver, number of clusters μ and correlation coefficient ρ on the observed performance.

Figs. 2 and 3 show the channel capacity normalized by channel bandwith B, depending on the SNR (dB). Some parameters are changing, and others are remaining constant. It is possible to see from these figures that channel capacity is growing with increasing of power Ω_0, decreasing of k, μ and ρ.

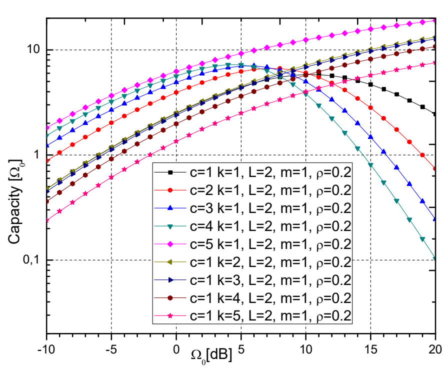


Figure 2. The C/B at the output of the MD SC combiner when the parameters c and k change

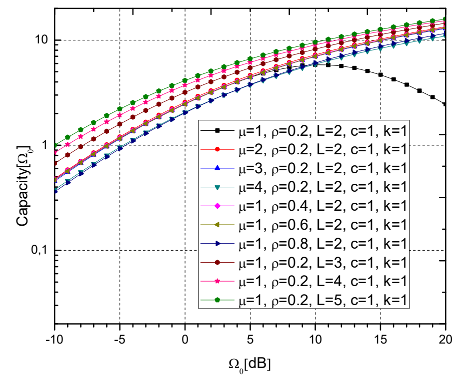


Figure 3. The C/B at the output of the MD SC combiner when the parameters μ, ρ and L change

V. Simulation and Network Planning Tool

The derived expression is further leveraged inside the software aiming modeling and simulation of mobile network smart city scenarios. The tool is accessed via web browser. It is developed relying on JavaScript (Three.js library for 3D graphics) and HTML on the front end, while Node.js is responsible for the back-end capabilities. Its architecture builds upon the previous work on simulation tools considering energy efficiency [24], fog computing [25] and resource planning in pandemic crisis [26].

In the first step, user creates a model of a smart city mobile network model via 3D graphical environment run in a web browser. Various aspects regarding the network infrastructure (base stations), environment (terrain and obstacles), communication channel properties and consumers are covered [27]. Moreover, the calculation of Quality of Service (QoS) parameters at different city locations is performed by leveraging the P_{out} and C/B values. Furthermore, the user-defined model together with assigned QoS and capacity values is used as input of linear optimization process with respect to the linear program from [27]. Finally, the output of optimization is optimal base station configuration, which can be further leveraged and translated to Software Defined Radio (SDR) commands. Fig. 4 gives an overview of the proposed modeling and simulation environment for network planning is given. However, in this paper, we focus on GPU-enabled fading calculation.

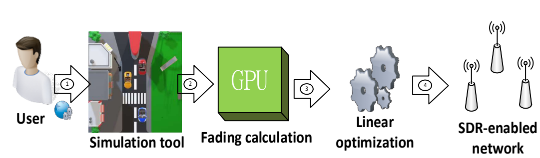


Figure 4. Smart city network simulation and planning environment: 1-Drawing network model 2-Model 3-QoS values based on fading value 4-Optimal SDR network configuration

The goal of general-purpose GPU (GPGPU) programming is to accelerate the calculations performed against huge data streams in case of non-graphic applications leveraging the graphics processing unit (GPU). According to the results in scientific literature [28, 29], this approach was approved as quite beneficial when it comes to fading calculation. In this paper, we adopt NVIDIA CUDA-enabled C program for calculation of P_{out} [30] value further used for determination of QoS value at a given location (*L*) at time point *t+1*, with respect to the formula:

$$Q_0S(L,\ t+1) = (1-P_{out})Q_0S_L(L,\ t)\,. \tag{20}$$

The structure of CUDA kernel used for QoS calculation is illustrated as a pseudo-code in Fig. 5. The capacity at each location is calculated leveraging GPU kernel in a similar way, as shown in Fig. 6.

According to the results achieved for system analyzed in this paper, calculation was around 58 times faster on GPU than the Origin-based program and 32 times compared to our previous analysis related to k-μ fading [30].

Once the QoS and channel capacity values are known, the linear optimization process can be executed. Its model's objective function [27] aims to minimize the energy consumption, distribution cost, quality of service and capacity drop for base stations at their locations, given as:

$$minimize \sum_{i\in BS,\, j\in L} x[i,j]energy\_c[i]distribution\_c[i,j]q\_ratio[i,j]c\_ratio[i,j] \tag{21}$$

According to the results achieved for system analyzed in this paper, calculation was around 58 times faster on GPU than the Origin-based program and 32 times compared to our previous analysis related to k-μ fading [30].

```
__global__ void calculate_qos(Location *l, float *qos_current)
{
    int tid = threadIdx.x + blockIdx.x * blockDim.x;
    while (tid < N)
    {
        qos_current[tid] = (1-Pout(l[tid])*l->qos_current[tid];
        tid += blockDim.x * gridDim.x;
    }
}
```

Figure 5. CUDA kernel for QoS factor determination based on P_{out} value

```
__global__ void calculate_capacity(Location *l, float *capacity)
{
    int tid = threadIdx.x + blockIdx.x * blockDim.x;
    while (tid < N)
    {
        capacity[tid] = C_B(l[tid])/C_NORM;
        tid += blockDim.x * gridDim.x;
    }
}
```

Figure 6. CUDA kernel for QoS factor determination based on C/B value.

Once the QoS values are known, the linear optimization process can be executed. Its model's objective function [27] aims to minimize the energy consumption, distribution cost, quality of service and capacity drop for base stations at their locations, given as:

$$minimize \sum_{i\in BS,\, j\in L} x[i,j]energy\_c[i]distribution\_c[i,j]q\_ratio[i,j]c\_ratio[i,j] \tag{22}$$

In expression (22), *x[i,j]* represents a decision variable, *energy_ct[i]* is the consumption of energy of a given *BaseStation[i]*, *distribution_c[i,j]* is an energy distribution cost at *Location[j]*. Moreover, *q_ratio[i,j]* is a ratio between the hypothetical QoS and the value affected by fading at *Location[j]*, while c_ratio[i,j] refers to ratio between maximal capacity and fading-affected capacity at a given location. When it comes to model constraints, at each Location[j] we can place exactly one *BaseStation[i]*, while each *BaseStation[i]* can be placed at only one location at the moment. Finally, the *capacity[i,j]* of a *BaseStation[i]* at *Location[j]* must be enough to satisfy the customer demand and minimal QoS value for a given location, given as:

$$\sum_{i\in Bs} (x[i,j]capacity[i,j]) \ge demand[j],\ j\in L\,. \tag{23}$$

VI. Conclusion

In this paper, the expression for channel capacity of the macrodiversity system made of SC macrodiversity receiver and two *L*-branches MRC microdiversity receivers in Gamma shadowed k-μ fading environment is determined. The graphically presented results display the impact of fading parameters: Rician *k* factor, the number of clusters μ, the shadowing severity *c*, and correlation coefficient ρ on the channel capacity at the output of observed macrodiversity system. Based on this analysis, it is possible to estimate the behavior of real macrodiversity system in the presence of shadowed *k*-μ fading.

These results for the *k*-μ fading can be used when other fading are present in the channel, such as: Nakagami-*m*, Nakagami-*n* (Rician) fading, and their special cases: Rayleigh and one-sided Gaussian fading because *k*-μ fading is general fading distribution. By putting special values for parameters *k* and μ, all special cases are covered by analysis given in this paper.

The adoption of GPGPU approach to fading calculation is highly beneficial, due to significant speed up which is of the most importance for simulation tools and calculations. To the best of authors' knowledge, similar analysis has not been presented in literature which increases significant the contribution of this paper.

Acknowledgment

The projects III44006, TR33035 and III47003 of the Ministry of Education, Science and Technological Development, Republic of Serbia partially support this paper.

REFERENCES

[1] M. Ibnkahla, Signal Processing for Mobile Communications, Handbook. 2005. 872.

[2] S. Mukheriee, D. Avidor, "Effect of micro diversity and correlated macro diversity on outages in a cellular system", IEEE Transactions on Wireless Technology, vol. 2, no. 1, pp. 50-59, 2003. doi:10.1109/twc.2002.806363

[3] G. L. Stüber, Principles of Mobile Communication, 2nd ed. New York: Kluwer Academic Publishers, 2002.

[4] Z. Ji, C. Dong, Y. Wang, and J. Lu, "On the analysis of effective capacity over generalized fading channels", IEEE International Conference on Communications (ICC 2014) - Communications Theory, pp. 1977-1983. doi:10.1109/icc.2014.6883613

[5] F. Mainardi, G. Pagnini, R. K. Saxena, "Fox H functions in fractional diffusion", Journal of Computational and Applied Mathematics, Vol. 178, Issues 1–2, 1 June 2005, Pages 321-331, https://doi.org/10.1016/j.cam.2004.08.006

[6] M.C. Stefanovic, D. S. Krstic, B. Z. Milosevic, J. A. Anastasov, S. R. Panic, "Channel capacity of maximal-ratio combining over correlated Nakagami-m fading channels", 9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services, Nis, Serbia, October 7-9, 2009, doi:10.1109/telsks.2009.5339434

[7] S. Khatalin, "On the channel capacity of SSC diversity system in η–μ and κ–μ fading environments", AEU - International Journal of Electronics and Communications, 69(11), 2015, pp. 1683–1699. doi:10.1016/j.aeue.2015.08.002

[8] J. Zhang, L. Dai, W. H. Gerstacker, Z. Wang, "Effective capacity of communication systems over κ–μ shadowed fading channels", Electronics Letters, 51(19), July 2015, DOI: 10.1049/el.2015.1032

[9] https://mathworld.wolfram.com/MeijerG-Function.html

[10] C. García-Corrales, F. J. Cañete, J. F. Paris, "Capacity of κ-μ shadowed fading channels", International Journal of Antennas and Propagation, vol. 2014, Article ID 975109, 8 pages, 2014. https://doi.org/10.1155/2014/975109

[11] M. Savic, M. Smilic, B. Jakšic, "Analysis of Shannon capacity for SC and MRC diversity systems in α–κ–μ fading channel", University Thought, Publication in Natural Sciences, Vol.8, No.2, 2018. doi: 10.5937/univtho8-19491

[12] M. M. Smilić, B. S. Jakšić, D. N. Milić, S. R. Panić, and P. Ć. Spalević, "Channel capacity of the macrodiversity SC system in the presence of kappa-mu fading and correlated slow gamma fading", Facta Universitatis, Series: Electronics and Energetics Vol. 31, No 3, September 2018, pp. 447-460, https://doi.org/10.2298/FUEE1803447S

[13] D. S. Krstic, S. Vasic, S. Suljovic, E. Biberovic, "Performance of macrodiversity system with selection combining and two microdiversity MRC receivers in the presence of k-μ fading", IEICE Information and Communication Technology Forum ICTF 2020, 10th - 12th of September 2020, Niš, Serbia

[14] M. D. Yacoub, "The κ − μ distribution and the η − μ distribution", IEEE Antennas and Propagation Magazine, vol. 49, no. 1, pp. 68–81, 2007.

[15] G. V. Milovanović, S. Suljović, S. R. Panić, I. Kalčo, and M. H. Stefanović, "Efficient numerical methods for analysis of square ratio of κ-μ and η-μ random processes with their applications in telecommunications", Mathematical Problems in Engineering, vol. 2018, Article ID 4967613, 9 pages, 2018.

[16] I. S. Gradshteyn and I. M. Ryzhik, Tables of Integrals, Series and Products, Academic. New York: 1980.

[17] J. Alan and D. Hui-Hui, Handbook of Mathematical Formulas and Integrals, San Diego: 4th edition, 2008.

[18] D. Brennan, "Linear diversity combining techniques", Proceedings of Institute of Radio Engineers, vol. 47, pp. 1075-1102, 1959.

[19] P. M. Shankar, "Macrodiversity and microdiversity in correlated shadowed fading channels", IEEE Transactions on Vehicular Technology, vol. 58, No. 2, February 2009.

[20] D. Krstić, S. Suljović, D. Milić, S. Panić, and M. Stefanović, "Outage probability of macro-diversity reception in the presence of Gamma long-term fading, Rayleigh short-term fading and Rician co-channel interference", Annals des Telecommunications=Annals of Telecommunications, vol. 73, Issue 5–6, pp. 329-339, June 2018.

[21] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Function with Formulas, Graphs and Mathematical Tables, National Bureau of Standards, Applied Mathematics Series – 55, Issued June 1964, Tenth Printing, December 1972, with corrections.

[22] M. S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," IEEE Trans. Veh. Technol., vol. 48, no. 4, pp. 1165–1181, 1999.

[23] N. Y. Ermolova, "Capacity analysis of two-wave with diffuse power fading channels using a mixture of Gamma distributions," IEEE Commun. Lett., vol. 20, no. 11, pp. 2245–2248, 2016.

[24] N. Petrović, Đ. Kocić, "Data-driven framework for energy-efficient smart cities", Serbian Journal of Electrical Engineering, Vol. 17, No. 1, Feb. 2020, pp. 41-63, 2020. https://doi.org/10.2298/SJEE2001041P.

[25] N. Petrovic, M. Tosic, "SMADA-fog: semantic model driven approach to deployment and adaptivity in fog computing", Simulation Modelling Practice and Theory, 102033, pp. 1-25, 2019. https://doi.org/10.1016/j.simpat.2019.102033

[26] N. Petrovic, Dj. Kocic, "Framework for efficient resource planning in pandemic crisis", CIIT 2020, pp. 1-6, 2020.

[27] N. Petrović, S. Koničanin, D. Milić, S. Suljović, and S. Panić, "GPU-enabled framework for modelling, simulation and planning of mobile networks in smart sities", ZINC 2020, pp. 1-6, 2020, https://doi.org/10.1109/ZINC50678.2020.9161773

[28] A. F. Abdelrazek, M. Kaschub, C. Blankenhorn, M. C. Necker, "A novel architecture using NVIDIA CUDA to speed up simulation of multi-path fast fading channels", VTC Spring 2009 - IEEE 69th Vehicular Technology Conference, pp. 1-5, 2009. doi:10.1109/vetecs.2009.5073317

[29] R. Carrasco-Alvarez, J. V. Castillo, A. C. Atoche, J. O. Aguilar, "A fading channel simulator implementation based on GPU computing techniques", Mathematical Problems in Engineering Vol. 2015, pp. 1-8, 2015.

[30] D. Milić, S. Suljović, N. Petrović, S. Koničanin, S. Panić, "Software environment for performance of relay signal by DF technique influenced by k-μ fading", INFOTEH 2020, 18-20 March 2020, East Sarajevo, Bosnia and Herzegovina, pp. 1-4, 2020. DOI: 10.1109/INFOTEH48170.2020.9066304

LoRaWAN Technology Mapping to Layered IoT Architecture

F. Flammini*, D. Dobrilović**, A. Gaglione*** and D. Tokody****
* School of Innovation, Design and Engineering, Division of Product Realisation, Mälardalen University, Eskilstuna, Sweden
** Technical Faculty "Mihajlo Pupin" Zrenjanin, University of Novi Sad, Zrenjanin, Serbia
*** Brit Insurance, London, UK
**** Doctoral School on Safety and Security Sciences Óbuda University, Budapest, Hungary
francesco.flammini@mdh.se; dalibor.dobrilovic@tfzr.rs; andrea.gaglione@britinsurance.com; tokody.daniel@nexttechnologies.hu

Abstract - The rapid expansion of the Internet of Things (IoT) systems and its growing usage, directly affected the expansion of the research interests in the area of IoT system layered architectures. The justification for the research in this area is connected with the complexity of IoT systems. Considering the complexity of IoT systems, it is crucial to define precise and detailed IoT system architecture to facilitate the development and implementation of these systems. The detailed IoT system architecture, in this approach, assumes the exact definition of the layers, their roles, and functional description.

This paper focuses on positioning LoRa/LoRaWAN technologies in existing IoT architecture models. The positioning of LoRa/LoRaWAN technologies is presented in the example of two IoT traffic application systems, one designed for the regional coverage, and one designed for urban coverage. The regional IoT traffic application is designed for the cooperation of two independent LoRaWAN networks. The urban IoT traffic system is designed as part of the Smart City or Smart transport application. The contribution of the paper is the proposed approach in positioning LoRaWAN technology in those systems and its mapping to IoT system architecture. These findings can contribute to an easier and more efficient implementation of LoRaWAN technology in future IoT systems.

I. Introduction

As one of the substantial characteristics of a modern word is the growing usage of IoT systems and the rapid evolution of supporting technologies for these systems. Together with technology consecution, the understanding of IoT system changes as well. These factors affect the increasing research interests in the area of IoT systems. To precisely define the structure of IoT systems, the layered architecture appears to be highly suitable.

The layered architecture approach in the description of IoT systems is justified with the complexity of IoT systems. Considering their complexity, it is crucial to precisely define the detailed IoT system architecture to facilitate the process of design and implementation of those systems. The detailed IoT system architecture, in this approach, assumes the exact definition of the layers, their roles, functional description, and their relation to other layers.

This paper presents the approach of positioning the LoRa and LoRaWAN technologies in existing IoT architecture models. The LoRaWAN is selected as a very popular and widely used technology in IoT systems today. Its popularity grows, and according to analytics, it will be one of the dominant technologies in the future together with NB-IoT.

The exact role of LoRa/LoRaWAN technologies in IoT is presented in this paper with the usage of two examples of IoT traffic application systems. One IoT Traffic system is designed for regional coverage, and one is designed for urban coverage.

The IoT Traffic application with the regional coverage is designed for the cooperation of two independent LoRaWAN networks in the propriety of shipping companies. The purpose of this application is to enable roaming between the networks as a support for shipment tracking. The IoT traffic system with urban coverage is designed as part of the Smart City or Smart transport application and presents the prototype application for monitoring noise and air pollution in the cities caused by the city traffic.

The contribution of the paper is the approach in positioning LoRaWAN in those systems and its mapping to widely accepted IoT system architecture. The finding of this research should facilitate and accelerate the design, development, and implementation of LoRaWAN networks in future IoT systems.

This paper is motivated with the research [1] where various protocol stacks, such as ZigBee, Bluetooth Low Energy, and LoRaWAN are compared to TCP/IP and proposed IoT protocol model. Another motivating research is presented in [2] where LoRaWAN components are mapped to the OSI reference model. Finally, the third motivational research [3] gives the mapping of modern communication technologies and protocols to the three layers of IoT architecture.

The paper is structured as follows. After the introduction, the related work is presented in the form of two sections of the paper, where the ITU-T IoT reference model is presented in one section, and four multilayered models are presented in the next section. In Section IV, the details of LoRa and LoRaWAN technologies, as well as the architecture of LoRaWAN networks, are presented. In Section V LoRaWAN technology is mapped to the selected multilayered architecture of IoT systems. The mapping is explained with the example of two different IoT traffic applications. In the end, the conclusion and further research directions are presented.

II. ITU-T IoT Reference Model

The first IoT reference model can be found in ITU Recommendation ITU-T Y.2060. This recommendation provides an overview of the Internet of things (IoT) with the clarification of its concept and scope. Additionally, it identifies the IoT fundamental characteristics and high-level requirements. One of the parts of the recommendation, containing the IoT reference model, is presented in Fig. 1. [4]

IoT reference model is a four-layer model with associated management and security capabilities. The four layers of the model are as follows: application layer, service support and application support layer, network layer, and the device layer.

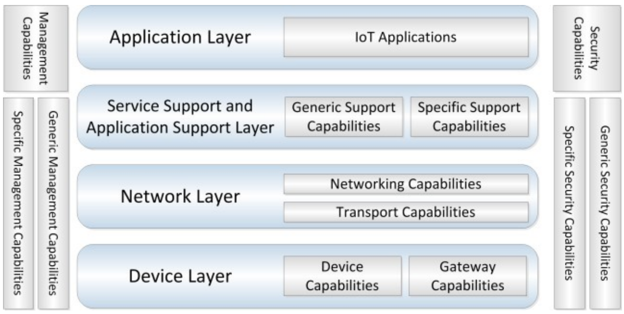


Figure 1. ITU-T IoT reference model

The *Application Layer* is the simplest layer to explain. This layer contains the IoT applications. Those IoT applications depend on IoT system and its end users (e.g. smart transport, e-health, smart agriculture, etc.)

The *Service Support and Application Support Layer* consists of two capability groups: Generic and Specific support capabilities. Generic support capabilities are common capabilities that can be used by different IoT applications, e.g. data processing or data storage. Specific support capabilities are particular capabilities that catered for the requirements of specific applications.

The *Network Layer* also has two following types of capabilities: Networking and Transport capabilities. Networking capabilities provide access and transport resource control functions, mobility management or authentication, authorization, and accounting (AAA). Transport capabilities provide connectivity for the transport of service and application-specific data, as well as the transport of IoT-related control and management information.

Same as two previous layers, the *Device Layer* capabilities are categorized into two groups: Device and Gateway capabilities. The device capabilities include direct interaction with the communication network. This is for the simplest form of IoT systems where end devices in IoT systems collect data and upload collected data directly to the communication network. Also, vice versa, the end devices are capable of receiving information (e.g., commands) form the core of the communication network.

In the case of indirect interaction with the communication network, end-devices collect data. The collected data are uploaded with the assistance (mediation) of the gateway devices. The same is with receiving information from the core of the communication network. The devices should support additional features such as ad-hoc networking, sleeping mode, and wake-up functions.

Gateway capabilities include multiple interface support. This feature is needed because end devices can use a variety of short-range technologies for data collections, such as: CAN bus, ZigBee, Bluetooth, Bluetooth Low Energy or Wi-Fi; and long-range technologies for data transfer such as the public switched telephone network (PSTN), 2G/3G/4G/5G mobile networks, long-term evolution networks (LTE), and variety of emerging Low-power wide-area network (LPWAN) technologies such as LoRa, LoRaWAN SigFox, NB-IoT, LTE-M, etc.

One of the important gateway capabilities is protocol conversion. It is needed in the case when communications at the *Device Layer* use different device layer protocols (e.g. ZigBee and Bluetooth) or when communications between the *Device Layer* and *Network Layer* use different protocols (e.g., ZigBee and Wi-Fi).

Besides four horizontal layers, the ITU-T IoT reference model covers two vertical layers: Management and Security capabilities.

IoT management capabilities cover fault management, configuration management, accounting management, performance management, and security management. Again, two groups can be defined. Essential generic management capabilities in the IoT include device management (remote device activation and de-activation, diagnostics, firmware and/or software updating, device working status management), local network topology management (traffic and congestion management). Specific management capabilities are closely related to application-specific requirements, e.g., smart grid power transmission line monitoring requirements, etc.

Security capabilities consist of generic and specific capabilities. Generic security capabilities are independent of applications, and their functions depend on the layer. At the level of *the Application Layer,* they include authorization, authentication, application data confidentiality, and integrity protection, privacy protection, security audit, and anti-virus functions. At the *Network Layer*, they include authorization, authentication, user and signaling data confidentiality, and signaling integrity protection. Finally, at the *Device Layer* capabilities include authentication, authorization, device integrity validation, access control, data confidentiality, and integrity protection).

Specific security capabilities are closely coupled with application-specific requirements, e.g., mobile payment, security requirements. [4]

As a conclusion, it can be outlined that the ITU-T IoT reference model, precisely defines different layers of IoT systems, the detailed description of the roles they have, detailed functional capability description as well as management and security capabilities of the system. The presented IoT reference model can be used as a guideline for the design and development of IoT systems with high efficiency.

III. IoT Multilayered Architecture Models

As an addition to the ITU-T reference model, many authors define their own multilayered IoT architectures. All these architecture models are more or less based on the ITU-T model. The most recognized architectures today are *Three-Layer* and *Five-Layer* [3, 5, 6].

A. *Three-Layer and Five-Layer Architecture*

Three-Layer and *Five-Layer* architectures are the most common IoT architectures today. The basic architecture is *Three-Layer* architecture. Those three layers are: *Perception*, *Network*, and *Application* layers, as it is shown in Fig. 2 (a). [3, 5]

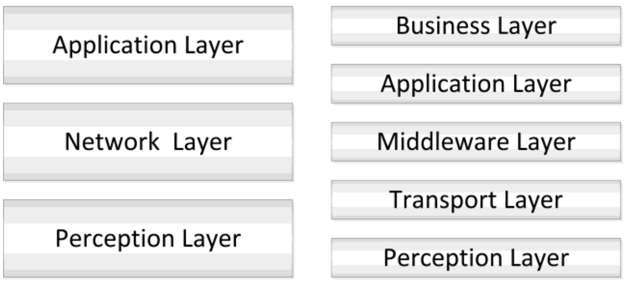


Figure 2. The architecture of IoT (a) Three-Layer (a) Five-Layer

The *Perception Layer* has the same functionalities as the *Physical Layer* of the OSI reference model. Sensor nodes or end devices are placed in this layer. The sensors collect data by sensing the physical parameters of the environment. The *Network Layer* is designed for the connection to the core components of the system (such as network devices) and for transferring collected sensor data to servers and data storage units. *The Application Layer* is responsible for the analytics of collected data, creating reports, and delivering data to the users [5].

Considering the complexity of IoT systems, many authors proposed architectures with additional layers. One, more popular architecture is the *Five-Layered* architecture [5, 6]. This architecture expands the IoT system model with two additional layers. Two additional layers are *Processing* (or *Middleware*) and *Business* layers. The role of the *Perception* and *Application* layers in the *Five-Layered* architecture is similar to in *Three-Layered* architecture. The *Network Layer* is in this model renamed to *Transport Layer*, but it still has similar functions. The *Transport Layer* is responsible for transferring sensor data from the *Perception Layer* to the *Processing Layer* and in the opposite direction. Again, short-range and long-range wireless and wired technologies are utilized for that purpose.

The difference comparing to the previous model is in the additional functions provided with the two new layers.

Two newly introduced layers have roles as follows. The *Processing Layer* or the *Middleware Layer* has the role to store, analyze, and process huge amounts of data. This layer covers a wide range of different technologies such as relational and non-relational databases, cloud computing, and big data, etc.

The *Business Layer* is on the top of the system and provides applications to the users, business and profit models, and users' privacy.

B. *Layered IoT Architecture and Security Issues*

The absence of the layers that are focused on security aspects of the system was the primary motivation for authors to expand the model with additional layers offering security [7]. In the model shown in Fig. 3 the expansion of the *Three-Layered* model is made with the *Support Layer*.

The role of the *Support Layer* is information security, In the process of direct transport of information from *Perception* to the *Network Layer* there are many potential security threats. In architecture presented in Fig. 3 the data from the *Perception Layer* are sent to the *Support Layer* first. After receiving the data, this layer checks the authenticity of the data sender and the authenticity of the data itself. The methods used for that purpose are authentication, pre-shared secrets, keys, and passwords.

After receiving the data from the sensors and their validation, the *Support Layer* forwards data to the *Network Layer* using available wireless and wire-based technologies. The purpose of the layer is to protect the system from DoS attacks, malicious insider attacks, unauthorized access, etc.

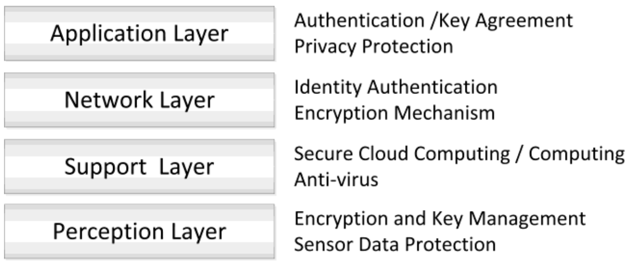


Figure 3. Four-Layer IoT architecture

The same authors [5] considering the difference between IoT and the Internet and telecommunication networks and in order to strengthen the security of the system suggest a new layered architecture of IoT that has six layers. The layers of the new proposed IoT architecture are *Perception Layer*, *Observer Layer*, *Processing Layer*, *Security Layer*, *Network Layer*, and *Application Layer*. *Six-Layered Secure* architecture is shown in Fig. 4.

The *Observer Layer* or *Monitor Layer* receives data from *Perception Layer*. Received data are checked and if there is no sign of threads the data are passed to the *Middleware Layer*. The authentication of the data sender is also preformed in this layer.

The *Security Layer* is designed to provide security for all components of the IoT system. It receives information from the *Processing Layer*, performs encryption by using keys. The only encrypted information is sent further to the network in a form readable and accessible only to authentic users.

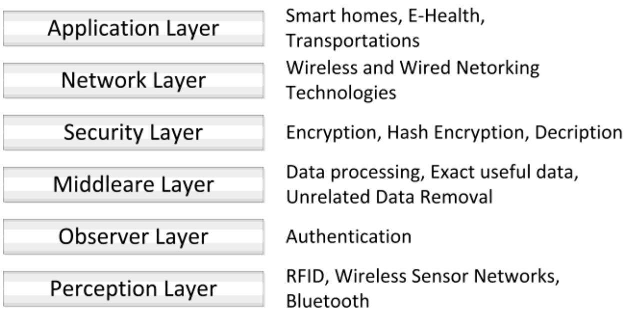


Figure 4. Six-Layer IoT architecture with security features

IV. LORA AND LORAWAN TECHNOLOGIES

The focus of this paper is to describe the possible usage of LoRa and LoRaWAN technologies in layered IoT systems. So, it is important to describe the main characteristics of those technologies, as well as basic examples of LoRaWAN network architectures.

LoRa and LoRaWAN can be classified as Low-power wide-area network (LPWAN) technologies that play a key role in IoT systems deployment. LPWANs are a group of technologies designed for usage in wireless wide area networks. The primary goal of these technologies is to provide long-range communications for IoT systems, with low bit rate and low power consumption. There is a variety of LPWAN technologies such as SigFox, NB-IoT, LTE-M, Weightless, etc.

A. LoRa and LoRaWAN technology basics

LoRa (Long Range) is initially a proprietary standard of Semtech company designed for license-free sub-gigahertz radio frequency bands (169 MHz, 433 MHz, 868 MHz in Europe, and 915 MHz in North America). The range of LoRa is very long and it exceeds 10 or 20 km in rural areas (according to different sources and much more). In the urban areas, it also provides long-range connectivity depending on the terrain configuration, building deployments, base station antenna placement, and receiver location. According to the experience of some authors, in urban areas, good link quality is achieved within a radius of 3 to 5 km. [8]

LoRa makes the foundation of LoRaWAN by enabling the physical layer of the LoRAWAN protocol stack and the communication protocol built upon the LoRa physical layer as it is shown in Fig. 5.

The upper layers of the LoRaWAN protocol stack are covered with LoRaWAN (Long Range Wide Area Network), an open-source communication protocol defined by the LoRa Alliance consortium. To be precise, LoRaWAN defines the communication protocol and system architecture of the network, while the LoRa physical layer enables communication. The first LoRaWAN standard was announced by the LoRa Alliance in June 2015 and LoRaWAN specification 1.1 was released in October 2017 [9]. LoRa Alliance is a non-profit technology alliance with more than 500 member companies.

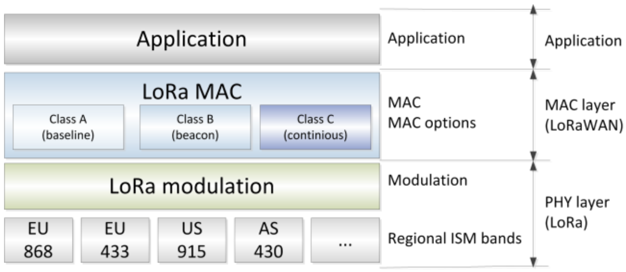


Figure 5. LoRaWAN protocol stack

B. LoRaWAN Network Architectures

LoRaWAN networks use the star and star-of-stars topology. The components of LoRAWAN networks are: end-devices, gateways, network server, application server, and join server. The basic architecture of the LoRaWAN network is shown in Fig. 6. LoRaWAN gateways provide connection and forward messages between end-devices and a central Network Server. The Network Server routes the packets from devices to the associated Application Server. Join Server deals with the storage of the device's root keys and the associated key derivation operations, to prove secure communication between the elements [10].

LoRaWAN End-device is a device with the following elements: sensor or an actuator, and wireless LoRa RF module for connecting to a LoRaWAN network. It is an autonomous device with a battery or solar-powered unit. LoRaWAN Gateways receive messages from End-devices and forward it to the LoRaWAN network server. IP backbone is in the core of the system and it enables a connection between LoRa gateways and other servers. The same end-device can be connected to multiple gateways.

The Network server also performs the management of the network, with dynamical control of the network parameters to adapt the system to the changing conditions. The second role of the network server is to establish secure 128-bit AES connections for the data transfer, to control traffic between End-devices and Network server, and to ensure sensor node authenticity and the integrity of messages [10].

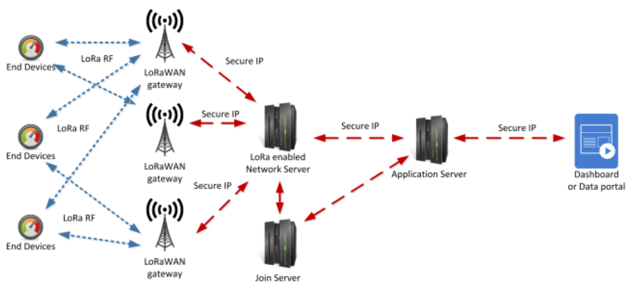


Figure 6. The basic architecture of LoRaWAN network

Two additional components of the LoRaWAN network are Application and Join servers. Application servers deal with secure handling, managing, and interpretation of sensor data. The Join server manages the process of joining end devices to the network and is responsible for signaling to the network server which application server should be connected to the end-device [10]. The complex architecture of the LoRaWAN network is shown in Fig. 7.

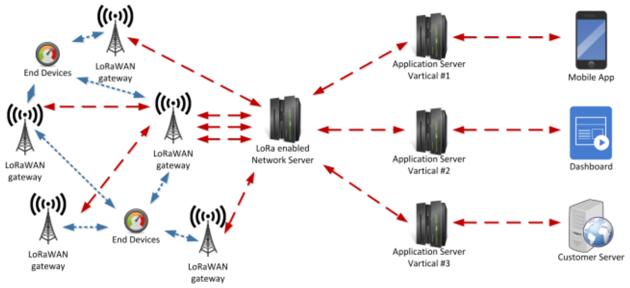


Figure 7. The complex architecture of LoRaWAN network

V. LoRaWAN in IoT Traffic Applications

In this section, LoRaWAN mapping to the multilayered IoT system architecture is presented. The deployment of the LoRaWAN components in multilayered IoT architecture is given according to the *Five-Layered* architecture that was chosen because of its greatest acceptance with the number of researchers. For mapping LoRaWAN components, two example IoT traffic applications based on LoRaWAN technology are used. One application is with regional coverage and one application is with urban coverage.

A. Regional LoRaWAN Roaming Application

This section presents LoRaWAN mapping to *Five-Layered* IoT architecture given on the example of the proposed LoRaWAN roaming model developed for the international shipment tracking, e.g. for transportation between a warehouse in Sweden and a warehouse in Norway. The IoT roaming application is a part of the ongoing research project, and will not be presented in detail in this paper.

For this paper, the relevant feature of the project is the addition of Distribution Servers (DS) to the basic LoRaWAN network basic architecture (Fig. 6). The role of Distribution Servers is to handle the roaming of transport vehicles, and shipment tracking between two LoRaWAN networks of cooperating shipping companies.

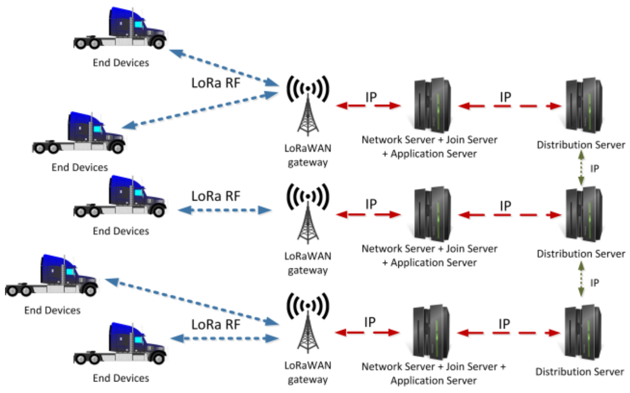


Figure 8. The architecture of Regional LoRaWAN Roaming Network

The assumption is that one operator and its LoRaWAN network covers Sweden, and other operator and its network covers Norway. The connectivity of the Distribution Server with other components of the LoRaWAN network is presented in Fig. 8. The mapping of the LoRaWAN subsystem presented in Fig. 8 to the *Five-Layered* IoT architecture is explained in following subsections and shown in Fig. 10.

B. Urban LoRaWAN Traffic Network

The example of the LoRaWAN network with city area coverage is given for the network designed for urban traffic management. This network is based on the prototype system presented in [11]. In that paper is presented the prototype LoRaWAN based network for urban traffic noise monitoring. The application is designed to propose different traffic routes in the city depending on increasing traffic noise on the particular routes and to keep the noise above certain levels. The similar system architecture can be applied to the system for monitoring traffic caused air pollution, with minor changes of sensor nodes by replacing the sound sensors, with the set of air pollutant sensors.

The deployment of the system components is presented in Fig. 9. LoRaWAN sensor nodes or End-devices (ED) are placed near the selected road routes. The EDs can be classified into several types, but it is not relevant for the research presented in this paper, so all EDs will be considered as the device of the same type. The detailed description of types of EDs is presented in [11]. EDs transfer the data to the core of the network via LoRaWAN Gateways and with the LoRa technology. Software component Semtech UDP Packet Forwarder is installed at GTWs and it is used to forward data, send from the EDs to LoRaWAN GTWs, and further to the core of the LoRaWAN network.

The core of the LoRaWAN network is based on ChirpStack, an open-source LoRaWAN server. ChirpStack has the following components: ChirpStack Gateway Bridge, ChirpStack Network Server, ChirpStack Application server, and ChirpStack Geolocation server. The detailed presentation and description of ChirpStack components are presented in [11]. The ChirpStack server further communicates with the other components of the IoT system such as database, and application, and the web server.

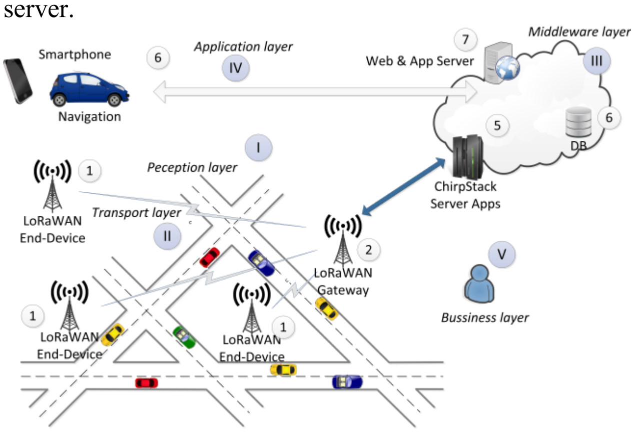


Figure 9. The architecture of Urban LoRaWAN Network

The mapping of the LoRaWAN subsystem of Urban LoRaWAN network presented in Fig. 9 to IoT Architecture is given in the next subsection and in Fig. 10.

C. *LoRaWAN Mapping to IoT Architecture*

Fig. 10 shows the mapping of the LoRaWAN protocol stack to IoT layered architecture. The LoRaWAN protocol stack covers two bottom layers of *Five-Layered* architecture, the *Perception Layer* and the *Transport Layer*. A similar mapping can be applied to the *Three-Layered* architecture (Fig. 2) and ITU-T IoT reference model (Fig. 1), where LoRaWAN covers *Perception* and *Network Layer*, and *Device* and *Network Layer* respectively.

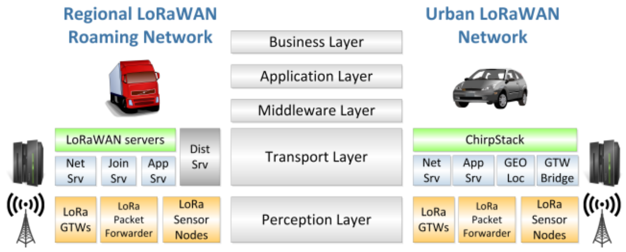


Figure 10. LoRaWAN Mapping to IoT Traffic Applications

The bottom layer (*Perception*) is covered with LoRaWAN hardware components (LoRaWAN enabled microcontroller platforms, LoRa communication modules) and partially with the software components (firmware, LoRa Packet Forwarder, etc.). The *Transport Layer* (or *Network Layer*) is covered with LoRaWAN software components such as Network, Application, Join, and other components such as Distribution and ChirpStack Geolocation servers and ChirpStack Gateway Bridge. The additional components connected to this layer are communication protocols for data transfer between these two layers.

This mapping model can be expanded in the future, by defining the sub-layers of the bottom two layers. The defined sub-layers should have their roles and functionalities also defined. This will lead to the creation of a precise and detailed IoT architecture model directly adopted for the LoRaWAN protocol stack. Considering the increasing popularity of LoRaWAN technology in the future, the proposed and expanded model should be an effective tool for LoRaWAN deployment.

VI. CONCLUSION

The rapid deployment of IoT systems and the rapid development of IoT related technologies established the challenges for the IoT system designers and developers. To tackle these challenges caused by IoT system complexity, significant effort is made in the research community to define IoT architecture, its layers, and layer functionalities.

This paper, in the first part, gives the overview of existing IoT multilayered architectures. In the second part, the paper discusses LoRa/LoRaWAN technology basics and its possible integration in IoT systems. The contribution of this paper is proposal of mapping of LoRaWAN technology to widely recognized IoT architectures. This mapping should facilitate and accelerate the design and development of future IoT systems based on LoRa/LoRaWAN technology.

One of the possible further research directions is the decomposition of the presented layers, and the definition of sub-layer roles and functionalities. This will result in more detailed IoT architecture with focus on LoRaWAN and can lead to the creation of an effective tool for the design and implementation process of LoRaWAN networks to IoT systems.

REFERENCES

[1] J. Tournier, F. Lesueur, F. Le Mouël, L. Guyon, H. Ben-Hassine, “A survey of IoT protocols and their security issues through the lens of a generic IoT stack”, Internet of Things, 2020, 100264, https://doi.org/10.1016/j.iot.2020.100264

[2] M.A. Ertürk, M.A. Aydın, M.T. Büyükakkaşlar, H. A Evirgen, “Survey on LoRaWAN Architecture, Protocol and Technologies”. Future Internet 2019, Vol. 11, Issue 10, 216, 2019.

[3] R. Yugha, S. Chithra, A survey on technologies and security protocols: Reference for future generation IoT, Journal of Network and Computer Applications, Vol. 169, 2020, 102763, https://doi.org/10.1016/j.jnca.2020.102763.

[4] Recommendation ITU-T Y.2060, Telecommunication standardization sector of ITU (06/2012), Series Y: Global information, infrastructure, Internet protocol aspects, and next-generation networks, 2012.

[5] P. Sethi and S. R. Sarangi, “Internet of Things: Architectures, Protocols, and Applications”, Journal of Electrical and Computer Engineering, Vol. 2017, 25 pages, https://doi.org/10.1155/2017/9324035

[6] I. Mashal, O. Alsaryrah, T-Y. Chung, C-Z. Yang, W-H. Kuo, D. P. Agrawal, "Choices for interaction with things on the Internet and underlying issues", Ad Hoc Networks, Vol. 28, 2015, pp 68-90, https://doi.org/10.1016/j.adhoc.2014.12.006.

[7] M. Burhan, R.A. Rehman, B. Khan, B. B.-S. Kim, “IoT Elements, Layered Architectures and Security Issues: A Comprehensive Survey”. Sensors 2018, Vol. 18, 2796, 2018.

[8] D. Dobrilović, M. Malić, D. Malić, S. Sladojević, “Analyses and optimization of Lee propagation model for LoRa 868 MHz network deployments in urban areas”. Journal of Engineering Management and Competitiveness (JEMC) 7(1), 55-62 (2017).

[9] LoRaWAN® Specification v1.1, 2017, LoRa Alliance Inc., [Online] retrieved in May 2020, https://lora-alliance.org/resource-hub/lorawanr-specification-v11

[10] LoRa® and LoRaWAN®: A Technical Overview, December 2019, Semtech Corporation, [Online] retrieved in May 2020, https://lora-developers.semtech.com/library/tech-papers-and-guides/lora-and-lorawan/

[11] D. Dobrilović, V. Brtka, G. Jotanović, Ž. Stojanov, G. Jauševac, and M. Malić, "Architecture of IoT system for smart monitoring and management of traffic noise", In Proceedings of EAI MMS 2020 - 5th EAI International Conference on Management of Manufacturing Systems (Accepted for publication), October 27-29, Opatija, Croatia, 2020.

Identification of Potentially Hazardous Traffic Situations Using Deep Learning

V. Brtka*, G. Jauševac**, G. Jotanović**, A. Stjepanović**, and M. Stojičić**
* University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia
** University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Herzegovina
vbrtka@tfzr.uns.ac.rs, goran.jausevac@sf.ues.rs.ba, gordana.jotanovic@sf.ues.rs.ba, aleksandar.stjepanovic@sf.ues.rs.ba and mirko.stojicic@sf.ues.rs.ba

Abstract - The paper describes the initial results of the application of Unmanned Aerial Vehicles for the control of traffic intersections or places in a special mode of operation, in urban areas. The method of photo processing is described, and the proposed Deep Learning model evaluates the correctness of the situation based on the processed photo. The focus of the work is on determining the initial architecture of the Artificial Neural Network. It was concluded that the application of the Deep Learning model is possible in the case of small training sets.

I. INTRODUCTION

Applications of machine learning (ML) methods and techniques in various domains result in visible progress. ML methods and techniques in the field of traffic and transport are in the focus of many researchers, and the results unequivocally indicate that this approach is fruitful. ML creates systems that learn from experience, so they are not static. This achieves increased system adaptability, fault tolerance, possible applications in various situations, upscaling, etc. The listed features allow the techniques from the ML domain to be used for control of traffic and identifying situations that have visual representation (image).

The research assumes the following scenario: It is necessary to control hazardous places or intersections in the urban environment. This includes road crossings, roundabouts, places that are congested in traffic jams, places that are in reconstruction or special regimes, etc. Of particular interest are intersections near the green areas, pavements, or places where it is possible to (il)legally park/stop the vehicle. The scenario also includes traffic control at intersections with special or temporary operating regimes, where there is no economic or other justification for installing permanent surveillance cameras. From all the above, it is clear that there must be a visual representation of the traffic situation.

The solution to the problems described in the scenario involves the use of Unmanned Aerial Vehicles (UAVs) - drones. The research is considering the possibility of flying over a place of interest with a drone, equipped with cameras to record the situation. In this way, a visual representation of the situation is enabled. The further procedure expects the analysis of the resulting images by ML techniques and binary conclusion: the situation is correct, the situation is not correct. In case of an incorrect traffic situation that implies hazardous or dangerous conditions, a team is sent to the position, which is obliged to resolve this situation. This can be the traffic police and/or the maintenance team. In a special case, a specially equipped drone or a fleet of drones may be sent.

The use of autonomous drones, in particular, creates additional problems: It is necessary to ensure the smooth passage of the drone, is the trajectory that must be passable. Also, it is necessary to comply with the regulations on the use of drones, and the big problem is the changing and unfavorable weather conditions. Given that some obstacles need to be overcome while using drones, the problem will be solved gradually, and this involves conducting experiments and specific research tasks. This specific research refers to the classification of visual representation of the situation (image) into two classes: correct, incorrect. In this sense, images created with the help of a drone that flew over a specific location are used to train an artificial neural network to classify the situation.

The rest of the paper is organized as follows: Section 2 gives some background theory regarding Deep Learning and Convolutional Neural Nets. Section 3 describes the Drone Project, its aims, and its restrictions. The last section presents the results, conclusions, and future enhancements.

II. THEORETICAL BACKGROUND

Artificial Neural Networks (ANNs) are widely used in many areas including traffic, traffic control, transportation, and traffic systems in general. The benefit of using ANNs for process optimization, prediction, and control was observed and this becomes a crucial property that contributed to the popularity of ANNs. A new era of ML and ANNs starts with the introduction of Convolutional Neural Networks (CNN) and Deep Learning (DL). According to [Wani]: "Deep learning refers to the architectures which contain multiple hidden layers (deep networks) to learn different features with multiple levels of abstraction. Deep learning algorithms seek to exploit the unknown structure in the input distribution to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features."

In [1] there are important and difficult-to-handle areas where DL has made good progress. Some of these areas are:

- Near-human-level image classification,
- Near-human-level speech recognition,
- Improved self-driving cars,
- Ability to answer natural language questions,
- Superhuman Go, Shogi, and Chess playing.

Near-human-level image classification is the feature of DL that is significant to this research so that the DL approach was chosen to keep in mind that it is necessary to recognize objects from the image, which is a task that DL neural networks solve very well [1, 2, 3].

III. ARCHITECTURE FOR DL

Deep Networks often refer to ANNs with multiple hidden layers, but CNN must have a convolutional layer. In terms of layers of different types, DL networks consist of:

- Convolutional layer,
- Activation function layer (ReLU),
- Pooling layer,
- Fully connected layer,
- Dropout layer.

The convolution layer as the core building block of a CNN uses convolution operation in place of general matrix multiplication. This is a mathematical operation, commutative in nature and performed on two functions written as $(f * g)$:

$$(f * g)(n) = \sum_{m} f(m)g(n-m) = \\ = \sum_{m} f(n-m)g(m) \tag{1}$$

Convolution function parameters are filters or kernels. This layer detects features that are common throughout the dataset and found within local regions of the image. The result of this layer is a feature map obtained for each filter (Figure 1).

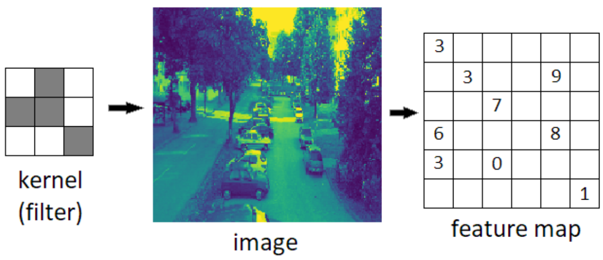


Figure 1. Convolution Layer

The output of the convolution layer is fed to the ReLU function. The Rectified Linear Unit (ReLU) function is given by: $f(x) = max(0, x)$. Currently, this is the most popular function for DL. ANNs with ReLU train much faster than other activation functions like sigmoid and tanh [1].

The pooling layer is an optional layer. This is the down-sampling layer used to reduce the spatial size of the input. Max-pooling is the most common technique and it simply outputs the maximum value in the input region (usually 2×2).

Lastly, a fully connected layer consists of layers in the sense of ANN with multiple layers, usually trained by a sort of Gradient Descent Algorithm (GDA) [4, 5]. This layer outputs class scores. Output neurons count and output activation function depends on the predefined classes.

The dropout layer is also optional and is there in order to overcome the problem of overfitting, thus enabling better generalization.

IV. DRONE PROJECT

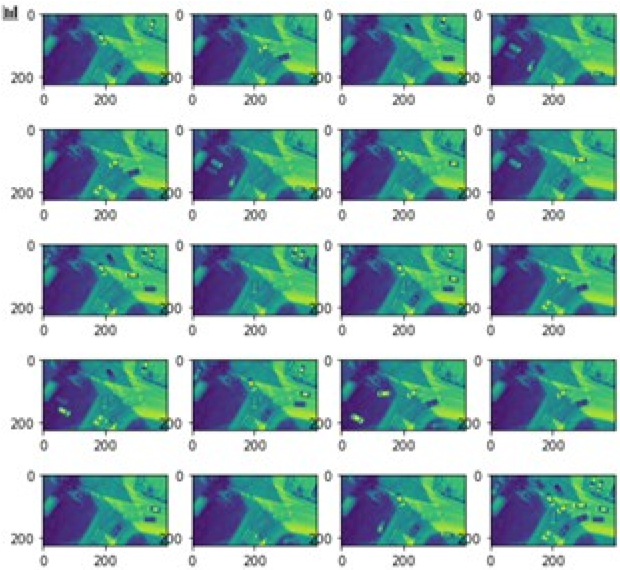
The use of UAVs in traffic as well as for recognizing objects from an image is becoming commonplace. Some of the applications of UAVs of this type are shown in [6, 7, 8].

This particular research involves the use of a drone equipped with a camera that monitors the road and records the situation. DL ANN technique implemented in TensorFlow / Keras environment (Visual Studio Code + Jupyter Notebook) is used for automatic image recognition. ANN training requires a set of images (training data), however, due to technical issues, time consumption, and legal limitations, it is not possible to provide a sufficiently large training sample. At first glance, this may be an insurmountable obstacle, but there are image augmentation techniques that can generate a sufficiently large training set.

V. SMALL DATA SET PROBLEM

In [9, 10] are tackled problems of datasets sizes. Attention is paid to small data sets that in some cases contain only a few images. The question of DL architecture is given which gives acceptable results in cases of small training sets. Frequently accepted data ratios of 80% : 20% (training set : test set) and 75% : 15% : 15% (training set : test set : validation set) are not acceptable in such cases. Data augmentation is a popular technique for generating additional labeled training examples through class-preserving transformations: rotation, reflection, jitter, noise, SNR, scale, shear [10]. In [9] authors have run several experiments and trained different DL architectures on some popular computer vision datasets. It was shown that networks with a low number of trainable parameters are less prone to overfitting and generalize better with small data. It was also shown that static augmentation with transformations that preserve image semantics improve performance and should be considered. For what concerns dropout, it was shown that it consistently improves results and should also be used in low data settings.

In this study, artificial image augmentation was performed by adding vehicles to different background positions. Thus, 40 images of the training set were formed: 20 images showing correct situations and 20 images showing incorrect situations. Having in mind that the ratio of instances by classes is satisfied, no further action has been taken on this issue. To prepare the data set for the training process pictures were scaled down to 800 × 450p, gray-scaled, and then again scaled down to 400 ×



225p. Part of the training set is shown in Figure 2.

Figure 2. Convolution Layer

The next step is to select a suitable ANN architecture for the DL process.

VI. DL ARCHITECTURE

A training set containing 32 images, a validation set containing 4 images, and a test set containing 8 images were formed. Given the size of these sets and the recommendations for selecting the architecture for DL in the case of small data sets, an architecture was selected consisting of: convolutional 2D layer, pooling 2D layer, Dense layer (hidden layer and one neuron output layer). The output layer contains a single neuron to make predictions. It uses the sigmoid activation function to produce a probability output in the range of 0 to 1 that can easily and automatically be converted to crisp class values. The preferred loss function for binary classification problems: Logarithmic loss function ("binary_crossentropy") was used during training. The model also uses the efficient "Adam" optimization algorithm for gradient descent and accuracy metrics. A model of the following architecture was formed:

1. Convolution2D(kernel = "X", kernel_size = (3, 3), activation='relu')
2. MaxPooling2D(pool_size = (2, 2))
3. Dense("Y", activation='relu')
4. Dense(1, activation='sigmoid')

The dropout layer is not used in this research. The model is tuned by choosing the values of X (kernel count) and Y (hidden layer neurons count). The tuning process was performed on 30 images of the training set while the rest of the images made up the validation set, and the values of X and Y were chosen to take into account the size of the training set and the hardware/software limitations. The performance of the model was checked at the training set. The summary results are shown in Table I.

TABLE I. MODEL TUNING

Architecture	Kernel count (X)	Dense (Y)	Loss	Accuracy
1	64	64	0.3535	0.8571
2	164	64	0.3187	0.8809
3	**64**	**164**	**0.2718**	**0.9285**
4	24	256	0.4799	0.7619
5	32	256	0.6931	0.5
6	60	120	0.2519	0.9047

The TensorFlow and Keras APIs were used to build these models. The ANN architecture under number three (Kernel count=64, Dense=164) was chosen to form the DL model since it achieves the highest accuracy. The model was trained on the training set containing 32 images with validation split set to 0.1.

VII. RESULTS AND CONCLUSIONS

The paper describes the initial results of the drone (UAV) application project in traffic control, specifically in the monitoring of particularly hazardous intersections in urban areas with properly or non-properly parked vehicles. The task that the drone should perform is recording with the help of cameras, and the system of which the drone is an integral part, based on the photo, evaluates and concludes: the situation is correct, the situation is not correct. An artificial neural network architecture for Deep Learning has been proposed, and kernel and hidden neuron numbers have been determined. The TensorFlow / Keras platform (Visual Studio Code + Jupyter Notebook) was used. It has been found that increasing the number of kernels significantly affects the length of training, even in training sets with a small number of elements. Achieved accuracy on the training set was 0.90625, while achieved accuracy on the test set was 0.875.

The future of this research is determined by the results achieved so far: the DL ANN architecture was chosen, and its parameters were conveniently selected. What is necessary are additional photos that will enable a larger training set. Also, automatic image augmentation methods will be applied to increase the number of elements of the training set. Drones need to be reused at various times and seasons. Also, a standard route for drone movements must be established, so that this approach would show its full effect.

ACKNOWLEDGMENT

Ministry for scientific and technological development, higher education and information society Republic of Srpska supports this research under the project "Monitoring parking of vehicles using drones", contract number 19.030/3-2-26-2/19.

REFERENCES

[1] M. Arif Wani, Farooq Ahmad Bhat, Saduf Afzal, Asif Iqbal Khan, "Advances in Deep Learning", Series editor Janusz Kacprzyk, Studies in Big Data 57, ISBN 978-981-13-6794-6, 2020.

[2] Pramod Singh, Avinash Manure, "Learn TensorFlow 2.0: Implement Machine Learning and Deep Learning Models with Python", ISBN-13 (electronic): 978-1-4842-5558-2, 2020.

[3] Chris Albon, "Machine Learning with Python Cookbook Practical Solutions from Preprocessing to Deep Learning", O'Reilly, 2018.

[4] Vladimir Brtka, "Machine Learning - Part I", University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, 2019.

[5] Vladimir Brtka, "Soft Computing", University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, 2019.

[6] B. Coifman, Mark McCord, R.G. Mishalani, M. Iswalt, Yuxiong Ji, "Roadway traffic monitoring from an unmanned aerial vehicle", IEE Proceedings - Intelligent Transport Systems 153(1):11- 20, DOI: 10.1049/ip-its:20055014, 2006.

[7] Giuseppe Salvo, Luigi Caruso, , Alessandro Scordo, " Urban traffic analysis through an UAV", Procedia - Social and Behavioral Sciences 111, 1083 – 1091, 2014.

[8] Bipul Neupane, Teerayut Horanont, Nguyen Duy Hung, "Deep learning based banana plant detection and counting using high-resolution red-greenblue (RGB) images collected from nmanned aerial vehicle (UAV)", https://doi.org/10.1371/journal.pone.0223906, accessed 02.09.2020.

[9] Lorenzo Brigato, Luca Iocchi, "A Close Look at Deep Learning with Small Data", https://arxiv.org/pdf/2003.12843.pdf, accessed 01.09.2020.

[10] Michael Majurski, "Small Data Deep Learning: AI Applied to Domain Datasets", NIST, https://www.nist.gov/system/files/documents/2019/08/27/worksho pslides-smalldataconvnets.pdf, accessed 01.09.2020.

Critical Success Factors for Implementing Software Process Improvement Project in a Local Micro Software Company

Z. Stojanov
University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
zeljko.stojanov@uns.ac.rs

Abstract - Software process improvement (SPI) is essential for software organizations that strive to be more productive and competitive at the market. In order to have the best possible effect on the company business performance, SPI initiatives should be implemented concurrently with other activities in the company. This put several constraints and barriers in implementing SPI initiatives, which are treated in literature as critical success factors (CSFs). This paper presents implementation of lightweight SPI project in a local micro software company, followed with the discussion of identified CSFs. Discussion of research validity and benefits for the company, small and micro software companies and researchers are also presented.

I. Introduction

Software companies operate in a dynamic market. To be competitive at the market, software companies tend to improve their products and services, which is mostly related to improvement of processes [1]. Improvement is based on understanding practice, and including human, organizational and technological aspects. In general, there are three paths for improving organization performance [2]: improvement of workers, improvement of processes, and improvement of technology. Each organization selects the most appropriate path that can be implemented with minimal costs and with available personnel. In many cases, in terms of feasibility, process improvement is the most cost effective way for improving organizations [3]. Software process improvement (SPI) assumes changes in software engineering practice in software organizations. Dynamics of process improvement projects and their relationships with everyday practice assume longitudinal research for long period, which is aligned to organizational aspects of industrial practice [4]. Based on an exploratory study, Staples et al. pointed out that software companies adopt SPI if they can anticipate clear benefits for their business [5].

Software process assessment (SPA) is critically important for SPI projects because it is focused on inquiring the current state of existing processes and identifying potential improvements. Process assessment enables measurement of process capabilities, which indirectly reflects maturity of organization whose processes are assessed [6]. Medium and large software organizations usually implement heavyweight approaches (top-down approaches) such as ISO/IEC 15504 - Software Process Improvement and Capability Determination (SPICE) [7] or Capability Maturity Model Integration (CMMI) [8], which are based on best practice guidelines and integrate experience from industry. However, these top-down approaches are too complicated and expensive for small software companies. Small companies rather adopt bottom-up approaches that start from the real state, use available resources, and based on that, through inductive analysis help in identifying potential improvements [9]. Nevertheless of the company size, people full involvement in SPI and SPA primarily relies on their notion of processes, which increases chances to get the more reliable assessment and identification of improvements for existing processes [10].

Small software companies nowadays represent majority in software industry. Majority of very small software companies (up to 25 persons) do not have resources to establish standardized life cycle processes [11]. In addition, since implementation of SPI approaches require resources such as humans, time, budget, and many details about existing processes, small software companies do not readily adopt SPI approaches, especially those approaches based on standards or guidelines [1]. Due to financial constraints, small software companies adopt low risk approaches that relatively quickly show results for any investment [12]. Successful implementation of SPI approaches in small software companies depends on many factors. However, there are also inhibitors that put constraints on the success of SPI approaches, such as the lack of management commitment and resources [13]. The influence of human factor on SPI success has been recognized in literature (company management, senior engineers, employees, customers), and investigated in empirical studies [1].

Based on the above discussion, we propose a research objective for this study: "Which are critical success factors for software process improvement project implementation in a local micro software company?" The paper is structured as follows. The next section briefly outlines work on CSFs in software process improvement. The third section presents a study in a micro software company, focused on CSFs for implementing a software process assessment and improvement. The fourth section presents discussion of benefits of the research and discussions of validity issues. The last section contains concluding remarks and future research directions.

II. Critical Success Factors in SPI

The role of critical success factors (CSFs) in different segments of software engineering practice has been widely investigated in scholarly literature. CSFs are referred to in the literature as the patterns of success and failure, as well as the best indicators of the lessons learned in software projects [14]. Analysis of CSFs and lessons learned are highly valuable for managers at all levels, who can use them for prioritizing the improvement opportunities in their organizations. SPI initiatives generally fail because top management does not pay enough attention to the project in terms of allocating the necessary resources, and often ongoing projects have much higher priority [15]. In addition, during SPI planning and implementation, special attention should be paid to selection of the appropriate measures that consider various contextual factors.

Virtanen et al. [15] investigated SPI success factors in a multinational software company with over 3200 employees. The study included in analysis data collected from eight geographically distributed sites, with low cohesion between distributed teams. Due to that organization of work, each team implements its own development approach which is mainly based on developing software from scratch. SPI project tried to introduce component based development to optimize time and budget costs. The following CSFs were identified in SPI project: top management support, training of programmers, resource allocation, personnel experience, staff involvement, and psychological factors of individuals.

Kouzari et al. [16] researched CSFs and barriers for implementing lightweight SPI initiatives in small and medium software companies. Through literature review, the following factors were identified: senior management commitment, staff involvement, staff training, allocation of resources, SPI team, staff experience, SPI guidance, review and feedback, SPI implementation methodology, monitoring, communication, return on investment (ROI), and awareness of SPI. The authors discussed ROI in SPI project because it is relevant for management that invest money and effort in SPI project, and helps in carrying out the entire SPI initiative in long term.

Sulayman et al. [17] proposed a theoretical framework of SPI success factors for small and medium web companies based on initial framework emerged in grounded theory empirical study and systematic review of SPI in small and medium web companies. Developed framework contains 18 categories of SPI success factors, 148 properties of these categories, and 25 corresponding relationships among categories. Framework was derived by using comparative analysis and theoretical integration as research methods. Framework can be used as a check list for conducting or assessing SPI project.

Niazi [18] conducted a literature review of SPI success factors by using: informal snowball literature review, and systematic literature review. After literature review, 34 interviews with SPI practitioners were organized in order to confirm findings. Findings revealed that practitioners have slightly different view comparing to literature review and stress the importance of training and mentoring. The most important factor, as it was identified in all three study phases, is management commitment. The findings can be useful for planning further SPI projects.

Khan et al. [19] presented a study with a systematic literature review of success factors for SPI in global software development. As study findings, 15 success factors were identified and classified into the six main categories. Among these factors, as the critical success factors for SPI were identified: management commitment, staff involvement, roles and responsibilities, communication, and resources allocation. The factors were discussed in perspective of software vendors and clients.

Sharma and Sangal [20] investigated inhibitors that considerably impact implementation of SPI initiatives in small and medium software organizations in global software development. Totally, 16 inhibitors with percentage greater than 25% were identified, while 5 inhibitors were identified as critical with percentage greater than 50%. Critical inhibitors are: lack of management commitment, lack of communication and information sharing, cultural differences, lack of implementation tools and standards, and lack of resources.

III. STUDY IN A MICRO SOFTWARE COMPANY

Software process improvement project has been implemented in a local micro software company. The project is a part of the project ”The development of software tools for business process analysis and improvement”, funded by the Ministry of Education, Science and Technological Development, Republic of Serbia. The project was prepared and conducted through joint work of the company employees and the researches from Technical Faculty ”Mihajlo Pupin” Zrenjanin.

The company has seven employees, and according to European Commission can be classified as a micro software company [21]. The company is oriented towards local clients, which significantly influences everyday work organization in the company. The majority of daily tasks in the company relates to maintenance of software business applications [22], which clearly points out the importance of software maintenance process assessment and improvement.

Process assessment was performed as a part of process improvement project, aimed at identifying segments of maintenance processes that need improvements.

A. Software Process Improvement Project

Process improvement project implemented in the selected company is presented in Figure 1. The first phase is planning of the whole SPI project, which resulted with plans for process assessment and plans for the whole improvement project. Based on the proposed plans, assessment activities are performed on the selected processes. Assessment plans include: (1) selection of processes to be assessed and potentially improved, (2) determination of the roles and assignments to people involved in assessment activities, (3) proposal of assessment time line, which is mainly constrained with everyday work activities in the company, (4) proposal of field research methods to be used in process assessment [23], and (5) determination of data sources that supports the triangulation of data and increase findings validity [24]. Improvement plans include details on the implementation phase of potential improvements, a general guideline how to track practice in the company after improvement implementation.

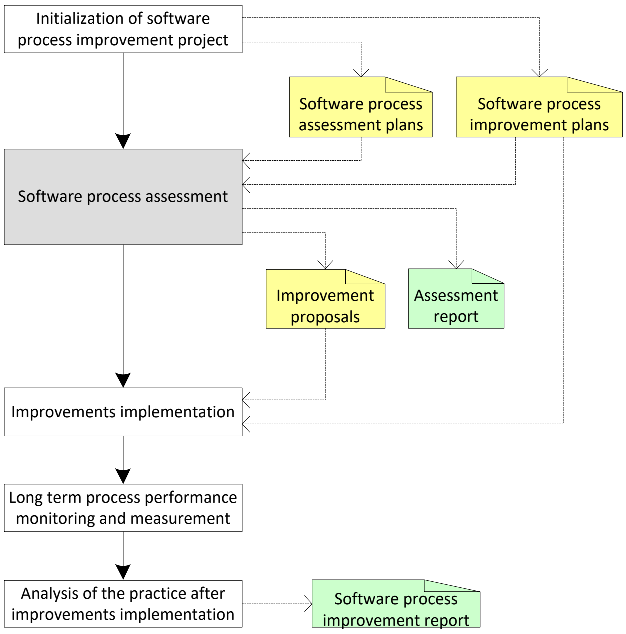


Figure 1. General overview of the software process improvement project

During the project planning phase, the leading researcher (the author of this paper) got to know the organization of the company, the roles, and duties of all employees in everyday work. In addition, access to technical infrastructure and documentation helped in getting more detailed insight into processes.

Implemented improvements are subject to long term monitoring and measuring to detect effects on everyday maintenance tasks and overall company business performance. Based on the analysis of implemented improvements, new cycles of assessment and improvement can be initiated.

B. *Process assessment and potential improvement identification*

Process assessment phase of SPI project is based on inductive approach with frequent feedback to the company employees and management (Lightweight Inductive Method for Process Assessment based on Frequent Feedback, LIMPAF[2]) [9]. The assessment method has been developed for small and micro companies and takes into account their specificities and constraints. The main objectives in designing the assessment method were:

- Achievement of quick and cheap process assessment regarding the time and required resources.
- Easy diagnosis of processes and determination of improvement proposals.
- Everyday working activities of company employees should not be disturbed by assessment activities.
- Frequent feedback to company management and employees, aimed at achieving the optimal results of assessment activities.

Developed assessment method is iterative and includes several loops with data collecting, data analyzing and reporting activities. Step-by-step tracking of process assessment was ensured by organizing working meetings (feedback sessions) with the company employees and informing them about the current state and further activities. This approach enabled organizational learning in the company [25], which improves work efficiency and business performance of the whole company [26].

Process assessment resulted with the assessment report and improvement proposals identified through inductive thematic analysis [27] of records from feedback sessions. Identified process improvement proposals are:

- Optimization of maintenance requests processing timeline with the focus on recording and triaging activities.
- Optimization of maintenance tasks scheduling based on evidence of working hours per maintenance requests.
- Development of web based application for reporting maintenance requests. The application will be available to clients and integrated with the internal software application for tracking tasks.
- Development of a software solution for reporting on maintenance request processing based on statistical data analysis.

Formal ranking of determined improvement proposals, based on the most relevant criteria for the company, was performed by using fuzzy screening method [28]. The first improvement proposal related to maintenance request timeline optimization was selected for the implementation in the company. The implementation was done as a technical solution including modification of software for tracking maintenance requests.

C. *Critical Success Factors*

Critical success factors (CSFs) for the process improvement were examined during the planning phase of the SPI project, and later refined during the implementation of project activities. CSFs are identified and analyzed through the discussions of the company manager (programmer with over 20 years of experience in software industry) and leading researcher (the author of this paper), which ensures that they reflect organizational issues in the company and specificities of SPI project. Analysis of CSFs includes they determination, and how they can be managed during project activities. The following CSFs were identified:

- *Company management support*. SPI project has been tailored to the needs of the company. The company manager participated in the design of SPI approach, which includes proposal of project objectives, identification of all required resources, and selection of research methods that can be applied in a way that do not disturb everyday tasks of company employees. In addition, the company manager and leading programmers participated in feedback sessions and approved all activities and decisions during the project implementation.

- *Availability of company employees on the project needs basis*. SPI approach was designed to involve company employees in all activities, including data collecting activities, data analysis and decision making. Company employees participated in activities when they were needed. For example, when a programmer was interviewed, he participated in feedback session in which interview was analyzed.
- *Access to resources in the company*. The most comprehensive insight into the processes assumes use of different resources in the company as sources of data. These resources include software applications, repository of tasks, company documents, and access to clients where it was needed. Different resources provide different views on the processes.
- *Identification of processes in the company*. The company does not have processes implemented according to any standard regulating software processes [29][7]. Instead, processes are either implemented ad-hoc, or are defined with a minimum specification or requirements. Such a situation makes it difficult to identify processes and their potential improvements. Due to that constraint involvement of all employees and use of various sources of data increases chances for project success.
- *Inclusion of additional researchers*. The SPI project was mainly implemented by the author of this paper and the company employees. However, some situations and complexity of collected data required inclusion of additional researchers that are experts in using specific research methods, especially quantitative methods such as statistical methods [30] or multi-criteria decision making methods [31].

Analysis of CSFs and their handling within SPI project was done by the company manager and leading researcher during the assessment implementation, during post-mortem analysis of assessment activities and outputs, and through long term tracking of the improvement effects on the company practice. The analysis was performed in the company, and it is included in the assessment and improvement reports.

The first critical success factor *Company management support* was identified as the most critical because it directly influences other CSFs. Support of the company management was ensured through involvement of the company manager and leading programmers in all important decisions and activities during SPI project. In addition, company manager validated all research reports and technical solutions produced during the project implementation.

IV. Discussion

A. Benefits

Identification of potential CSFs in the planning stage of the SPI project helped the company management to identify possible obstacles in the project implementation. Deeper and more comprehensive understanding of potential problems, constraints and barriers in the project realization increases chances for success in the project implementation.

Knowledge and experience gained during the project planning and implementation can be used in preparing further projects. Better project planning, more reliable and structured implementation, and more accurate validation of the project outcomes are possible benefits for the company.

Other micro and small software companies can use these identified CSFs as a starting point, or adapt them to their specific needs, when considering the implementation of SPI projects. In addition, the proposed approach for organizing and implementing lightweight SPI project and for considering CSFs specific to internal organization and needs is another benefit for them.

An finally, researchers from academia can find lessons how to consider and identify CSFs when preparing SPI projects in micro and small software companies.

B. Validity

Software engineering empirical research needs to fulfill validity criteria to increase the reliability and usability of reported findings [32]. Two validity types are of importance for this study: internal and external validity. Internal validity, or soundness of research process and accuracy of data and research methods, is ensured through development of a detailed research plan by the author of this paper and the company management, and strict monitoring of research process through frequently organized feedback sessions in the company. In addition, triangulation of data sources and adequate data analysis methods [33][34] increases internal validity of study and reliability of findings.

External validity, or generalizability, can be observed regarding the research methods arranged in the specific research approach presented in several published papers. Although, the presented research findings coincide significantly with the results published in the scientific literature (see section II of this paper), the main issue when discussing generalizability relates to application of presented SPI approach in similar micro and small software companies.

V. Conclusion

This paper presents a study on CSFs identified in a SPI project implemented in a local micro software company. The main contribution of the study is identification and discussion of critical success factors for project implementation. The authors are aware that these CSFs are specific for the selected micro company (context specific), but experiences gained in this project may be useful for the next SPI projects in this and other small companies. This will ensures better planning, more structured and reliable implementation, and more accurate validation of project outcomes and benefits for assessed companies.

Further work will be directed in the following directions. The first direction relates to more detailed analysis of identified CSFs and their comparison with CSFs identified in literature, followed with the development of taxonomy of CSTs for SPI projects in micro software companies. Due to the dynamics of IT industry, controlled evolution of proposed taxonomy is also challenging research direction. The next research direction is the identification of the metrics for measuring the impacts of proposed CSFs, which will lead to the refinement of these factors and more reliable measuring of their importance. Implementation of proposed SPI approach in other micro IT companies will help in getting empirical evidence of its usefulness, as well as in developing and refining taxonomy of CSFs.

ACKNOWLEDGMENT

Ministry of Education, Science and Technological Development, Republic of Serbia, supports this research under the project "The development of software tools for business process analysis and improvement", project number TR32044.

REFERENCES

[1] M. A. Almomani, S. Basri, and A. R. Gilal, " Empirical study of software process improvement in Malaysian small and medium enterprises: The human aspects ", Journal of Software: Evolution and Process, vol. 30, no. 10, pp. e1953. 2018.

[2] M. West, Return on process (ROP): Getting real performance results from process improvement (1st ed.). CRC Press. Boca Raton, FL, USA. 2013

[3] F. J. Vasconcellos, G. B. Landre, J. A. O. Cunha, J. L. Oliveira, R. A. Ferreira, and A. M. Vincenzi, "Approaches to strategic alignment of software process improvement: A systematic literature review" Journal of Systems and Software, vol. 123, pp. 45-63.

[4] I. Allison, and Y. Merali, "Software process improvement as emergent change: A structurational analysis", Information and Software Technology, vol. 49, no. 6, pp. 668-681. 2007.

[5] M. Staples, M. Niazi, R. Jeffery, A. Abrahams, P. Byatt, and R. Murphy, "An exploratory study of why organizations do not adopt CMMI", Journal of Systems and Software, vol. 80, no. 6, pp. 883-895.2007.

[6] T. McBride, "Organisational theory perspective on process capability measurement scales", Journal of Software Maintenance and Evolution: Research and Practice, vol. 22, no. 4, pp. 243–254. 2010.

[7] ISO/IEC 15504 Information technology - Process assessment (Software Process Improvement and Capability Determination (SPICE) (Standard). International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC). Geneva, Switzerland. 2008

[8] Capability Maturity Model Integration (CMMI) for Development, Version 1.2 (Tech. Rep. No. CMU/SEI-2006-TR-008). Carnegie Mellon University. Pittsburgh, PA, USA. 2006

[9] Z. Stojanov, J. Stojanov, and D. Dobrilovic, "A lightweight inductive method for process assessment based on frequent feedback: A study in a micro software company", Journal of Engineering Management and Competitiveness (JEMC), vol. 9, no. 2, pp. 134-147. 2019.

[10] A. Sampaio, I. B. Sampaio, and E. Gray, "The Need of a Person Oriented Approach to Software Process Assessment", In Proceedings of the 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), pp.145-148. San Francisco, CA, USA. 2013.

[11] V. Ribaud, and P. Saliou, "Process assessment issues of the ISO/IEC 29110 emerging standard", In Proceedings of the 11th International Conference on Product Focused Software (PROFES '10), pp. 24-27. Limerick Ireland. 2010.

[12] H. M. Abushama, "PAM-SMEs: process assessment method for small to medium enterprises ", Journal of Software: Evolution and Process, vol. 28, no. 8, pp. 689-711. 2016.

[13] P. Sharma, and A. L. Sangal, "Framework for empirical examination and modeling structural dependencies among inhibitors that impact SPI implementation initiatives in software SMEs", Journal of Software: Evolution and Process, vol. 30, no. 12, pp. e1993. 2018.

[14] V. Garousi, A. Tarhan, D. Pfahl, A. Coşkunçay, and O. Demirörs, "Correlation of critical success factors with success of software projects: an empirical investigation", Software Quality Journal, vol. 27, pp. 429–493. 2019.

[15] P. Virtanen, S. Pekkola, and T. Päivärinta, " Why SPI Initiative Failed: Contextual Factors and Changing Software Development Environment", In Proceedings of the 46th Hawaii International Conference on System Sciences, pp. 4606-4615. Wailea, Maui, HI, USA. 2013.

[16] E. Kouzari, V. C. Gerogiannis, I. Stamelos, and G. Kakarontzas, " Critical success factors and barriers for lightweight software process improvement in agile development: A literature review", In Proceedings of the 10th International Joint Conference on Software Technologies (ICSOFT), pp. 151-159. Colmar, France. 2015.

[17] M. Sulayman, E. Mendes, C. Urquhart, M. Riaz, and E. Tempero, "Towards a theoretical framework of SPI success factors for small and medium web companies", Information and Software Technology ", vol. 56, no. 7, pp. 807-820. 2014.

[18] M. Niazi, "A comparative study of software process improvement implementation success factors", Journal of Software: Evolution and Process, vol. 27, no. 9, pp. 700-722. 2015.

[19] A. A. Khan, J. Keung, S. Hussain, M. Niazi, and S. Kieffer, " Systematic literature study for dimensional classification of success factors affecting process improvement in global software development: client–vendor perspective", IET Software, vol. 12, no. 4, pp. 333-344. 2018.

[20] P. Sharma, and A. L. Sangal, "Investigating the factors which impact SPI implementation initiatives in software SMEs - A systematic map and review", Journal of Software: Evolution and Process, vol. 31. no. 7, pp. e2183. 2019.

[21] European Commission. User guide to the SME definition. Luxembourg: Publications Office of the European Union. 2015.

[22] Z. Stojanov, J. Stojanov, D. Dobrilovic, and N. Petrov, "Trends in software maintenance tasks distribution among programmers: A study in a micro software company", In Proceedings of the IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY2017), pp. 23-27. Subotica, Serbia. 2017.

[23] T. C. Lethbridge, S. Elliott Sim, and J. Singer, "Studying Software Engineers: Data Collection Techniques for Software Field Studies", Empirical Software Engineering, vol. 10, no. 3, pp. 311-341. 2005

[24] L. Bratthall, and M. Jørgensen, "Can you Trust a Single Data Source Exploratory Software Engineering Case Study?", Empirical Software Engineering, vol. 7, no. 1, pp. 9-26. 2002.

[25] Z. Stojanov, J. Stojanov, and D. Dobrilovic, "Knowledge Discovery and Systematization through Thematic Analysis in Software Process Assessment Project", In the proceedings of IEEE 13th International Symposium on Intelligent Systems and Informatics (SISY 2015). pp. 25-30. Subotica, Serbia. 2015.

[26] T. Dyba, T. Dingsoyr, and N. B. Moe, Process improvement in practice - a handbook for IT companies (Vol. 9). Kluwer Academic Publishers. Norwell, MA, USA. 2004.

[27] V. Braun, and V. Clarke, "Using thematic analysis in psychology", Qualitative Research in Psychology, vol. 3, no. 2, pp. 77-101. 2006.

[28] R. R. Yager, Fuzzy screening systems. In R. Lowen & M. Roubens (Eds.), Fuzzy logic: State of the art (Vol. 12, pp. 251-261). Springer Netherlands. Dordrecht, Netherlands. 1993.

[29] Standard CMMI Appraisal Method for Process Improvement (SCAMPI) Version 1.3a: Method Definition Document for SCAMPI A, B, and C. Handbook, CMMI Institute-2013-HB-001. CMMI Institute. Carnegie Mellon University. Pittsburgh, PA, USA. October 2013.

[30] Z. Stojanov, J. Stojanov, and D. Dobrilovic, "Domain Complexity in Corrective Maintenance Tasks' Complexity: An Empirical Study in a Micro Software Company", Theory and Applications of Mathematics & Computer Science, vol. 8, no. 1, pp. 24-38. 2018.

[31] Z. Stojanov, V. Brtka, and D. Dobrilovic, "Ranking Software Maintenance Processes in a Small Software Company in the Context of Software Process Improvement", Scientific Bulletin of The “Politehnica” University of Timişoara, Transactions on Automatic Control and Computer Science, vol. 59 (73), no. 2, pp. 183-192. 2014.

[32] H. K. Wright, M. Kim, and D. E. Perry, "Validity Concerns in Software Engineering Research", In Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research, pp. 411–414. Santa Fe, New Mexico, USA. 2010.

[33] J. Moran-Ellis, V. D. Alexander, A. Cronin, M. Dickinson, J. Fielding, J. Sleney, and H. Thomas, "Triangulation and integration: processes, claims and implications", Qualitative Research, vol. 6, no. 1, pp. 45-59. 2006.

[34] J. Miller, "Triangulation as a basis for knowledge discovery in software engineering", Empirical Software Engineering, vol. 13, no. 2, pp. 223-228. 2008.

Rerouting Traffic Based on Noise Values and Number of Vehicles in Urban Areas

Gordana Jotanović*, Vladimir Brtka**, Goran Jauševac*, Zoran Ćurguz*, Miroslav Kostadinović*, Eleonora Brtka**

* University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Herzegovina

** Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia

gjotanovic@yahoo.com, brtkav@gmail.com, gjausevac@gmail.com, curguzzoran@yahoo.com, miroslav.kostadinovic@sf.ues.rs.ba, eleonorabrtka@gmail.com,

Abstract - The paper deals with the estimation of the number of vehicles at a certain location in an urban area, based on the measured values of noise, pollution, and the number of vehicles at several other locations. The research was conducted to determine the initial model that allows, based on measured values, primarily noise, to make changes in the routes of vehicles in urban areas. Real-life data were used, and the method is compatible with Rough Sets Theory and is based on creating an indiscernibility graph. The values of the noise threshold are determined, which indicates possible traffic jams and therefore serve as indicators of the need to change the traffic mode.

I. Introduction

Noise pollution refers to the level of noise in the human environment that is higher than the acceptable level caused by traffic, construction, industrial, and individual recreational activities [1]. Noise, like unwanted sound, causes various interruptions during work and rest, interferes with voice communication and affects a person's general and work behavior, and at strong levels can have pathological effects. Europe-wide activities to reduce environmental noise have a different priority than environmental issues such as air and water pollution, often because such issues were considered to be best addressed at the national or local level. Noise protection measures are regulated by the "Law on Environmental Protection" of the Republic of Srpska [2]. However, in the Republic of Srpska, this problem is not considered, and he therefore, does not have enough attention, although significantly affects the quality of life of the affected population. The reasons for this behavior may be based on the subjective experience of certain external events, the specific character of noise and difficulties in linking causes with effects on human health [3]. Due to all the above, the goal is to develop a system for monitoring noise, and other pollutants but primarily noise that occurs as a result of traffic, i.e. traffic noise. The prototype of the traffic pollution monitoring system involves testing and development of components that enable condition monitoring, analysis of the condition, and eventual recommendations on ways and possibilities to avoid pollution.

A special challenge we face in this research is how to estimate the number of vehicles in near the hospital, depending on the number of vehicles and noise, at several other locations in the urban environment. The assumptions are as follows: The number of vehicles, by itself, may not be a good indicator because, for example, heavy trucks are unlikely to go in the direction of the hospital. On the other hand, noise does not depend only on the number of vehicles because vehicles of different types generate different amounts of noise, for example, heavy trucks are louder than passenger vehicles, although this is not always the case. The point is that the number of vehicles separately, as well as noise as a separate parameter, cannot be used as good estimators of the number of vehicles in another location.

Based on the above, the goal of the research is to estimate the number of vehicles present at a certain location in an urban environment, by measuring the number of vehicles and noise at several other locations. It is desirable that the number of locations where vehicles are counted and noise is measured, be as small as possible due to savings. Such a system would enable the development of an adaptive automated traffic rerouting system, so that it is possible to influence the number of vehicles present in the vicinity of a particular location.

The rest of the paper is structured as follows: Section II gives some insight into road traffic noise. Section III describes the data collection process and methodology. Results are given in Section IV, while conclusion and future work are given in Section V.

II. Road Traffic Noise

One of many sources of noise is traffic, as well as vehicles participating in traffic, and this type of noise is called traffic noise. The great problem is noise in traffic on city roads because such roads pass through urban settlements, near hospitals, schools, kindergartens, cultural monuments, etc. The European Union defined the term "environmental noise" in the 2002 Directive and refers to unwanted and harmful external influences generated by human activities, including noise generated by transport (road, rail and air) and exposure of the population to urban noise [4].

Traffic noise is one of the main environmental and health problems in modern cities. The European Environment Agency has released a report showing that around 125 million people have been affected by road noise levels above 55 dB a day, causing 10,000 premature deaths, nervousness and anxiety in almost 20 million adults, and 8 million a year with sleep disorders [5].

A. *Traffic noise problem*

The problem of extensive traffic noise is solved by many countries in the European Union. One of the projects funded by the European Union that deals with noise reduction in urban areas is SYLVIE (SYstematisch Lärmsanierung von innerstädtischen WohnVIErteln) [6]. Road traffic noise (cars, buses, trucks, and motorcycles) is the most common source of noise in cities and is a major cause of disruption to human activities. Long-term exposure to high noise levels has far-reaching negative impacts on human well-being and health, with measurable economic consequences: treatment costs, the declining work activity, and productivity, increased mortality, declining market value of buildings in excessive noise zones [7]. When considering traffic noise, one should first start analyzing the noise characteristics of isolated vehicles, passenger vehicles, motorcycles, heavy vehicles, and then analyze traffic noise, especially in urban environments, where there are interrupted and uninterrupted traffic flows. However, in order to better understand the problem of the impact of traffic noise on people, it is necessary to consider the basic sources of traffic noise in road and rail air traffic [8].

III. Data Collection and Methodology

Data were collected in 2018. and in 2020. by measurements at several key locations: the main bus station (SAS), the city center (C), the "Japanese Bridge" (JM), and the hospital (B), as shown in Fig. 1.

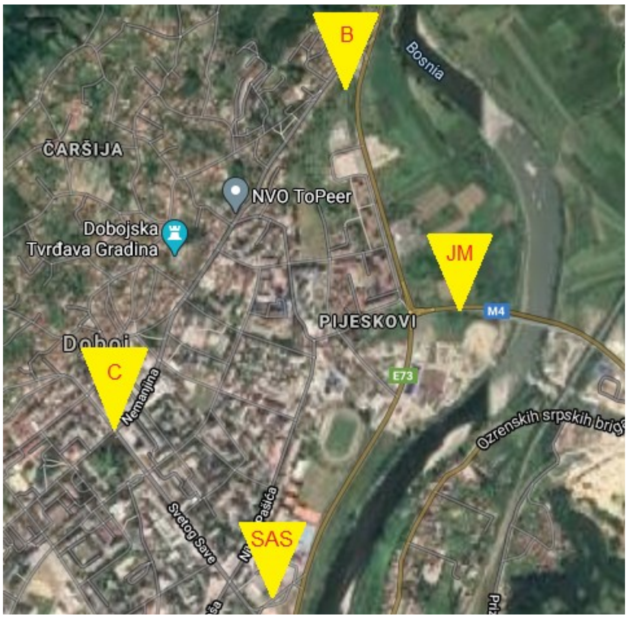


Figure 1. Main bus station (SAS), City center (C), "Japanese Bridge" (JM) and the Hospital (B)

For each location, the following values were measured: vehicle count, noise, count of PM2.5 and PM10 particles, humidity, and temperature. However, the focus of the research is on the impact of vehicle count and noise to the location of choice. In this case, the chosen location is the hospital (B), having in mind its significance in the conditions of a pandemic. The following table gives the correlation between the number of vehicles and the noise, with the number of vehicles that are present near the hospital at that moment.

TABLE I. Correlation to Vehicle Count at Hospital Location

No.	Attribute	Correlation to B vehicle count
1.	SAS noise	0.220704
2.	SAS count	-0.801185
3.	C noise	-0.025784
4.	C count	-0.622172
5.	JM noise	-0.295047
6.	JM count	-0.721873

It is obvious that there is a mostly negative correlation. A smaller number of vehicles at the SAS and JM locations means a larger number of vehicles in the vicinity of the hospital. The same is somewhat true for the location of C. However, based on the noise values measured at the SAS, C, and JM locations, no strong correlation can be established with the number of vehicles in the vicinity of the hospital. This can be justified by the fact that vehicles are not the only noise generator, nor do all vehicles generate the same amount of noise. Thus, the number of vehicles measured at SAS, JM, and C locations can be an indicator of the number of vehicles at location B. The conclusion that if the vehicles are not at SAS, JM and C locations, they must be at location B is logical but certainly not overly useful. Namely, the generality of this conclusion is questionable, and vehicles can certainly be parked elsewhere. The question is how a model based on such a rule would work in each particular case. Having in mind the previous statement, a good approach implies measuring the number of vehicles as well as noise, at several remote locations, so that the combination of these two parameters becomes the basis for estimating the number of vehicles near the hospital.

Based on all the above, it is necessary to perform a qualitative analysis of the data, especially bearing in mind that this is a small data set. Qualitative data analysis was performed by a method analogous to the application of Rough Sets Theory (RST) [9, 10, 11]. Like RST, the method uses the indiscernibility relation (1) and forms an indiscernibility graph. Let $M = \{m_1, m_2, \dots, m_n\}$ be a finite set of measurements while $A = \{a_1, a_2, \dots, a_m\}$ is a finite set of parameters or attributes so that each measurement from M is characterized by attributes from A, while function $f = M \times A \to V$ is called information function, where V is a set of values. All attributes are divided so two sets are formed: set of condition attributes C and set of decision attributes D, so that $C \cap D = \emptyset$. Set D usually consists of a single attribute, in this case, this would be a vehicle count near the hospital. For $P \subseteq C$ indiscernibility relation is defined by:

$$I_P = \{(x, y) \in M \times M : f(x, a) = \\ = f(y, a), \forall a \in P\} \tag{1}$$

Every path that leads to the non-empty leaf, form if-then rule. By applying the indiscernibility relation when forming the graph, only those attributes are included that do not change this relation, i.e. the reduct of the set of attributes in the sense of RST is achieved. Based on indiscernibility graph, it is also possible to synthesize a semi-linguistic summaries in the form of if-then rules, using fuzzy quantifiers: "few", "some", "half", "many" and "most" [12, 13]. The method can be applied to diverse data sets, [14]. Thus, based on the indiscernibility graph, the following sets can be formed: set of if-then rules of the usual form and set of if-then rules that use linguistic terms and enable a summary view of the data: hence they are called semi-linguistic summaries.

The application of the RST based method, as well as the synthesis of the indiscernibility a graph imply that the values of the condition attributes are discrete. The discretization of attribute's values was performed by the Equal Frequency Binning (EFB) algorithm, which is an unsupervised algorithm. All values are divided into three bins, which correspond to low, medium and high. The EFB discretization process was done by the free software system Rosetta. The number of vehicles present in the vicinity of the hospital (decision attribute) is classified into the following categories, respectively: min, avg, submax, and max. The algorithm is implemented in the Java programming language.

IV. Results

A total of 17 if-then rules were generated based on the indiscernibility graph. The following issues were considered: Under what conditions are the number of vehicles in the vicinity of the hospital maximal or minimal? Of the 17 rules, four relate to the situation described. Please refer to Table II for further details on if-then rules.

TABLE II. The Influence of Parameters on the Number of Vehicles Near the Hospital

	IF	**IF**	**IF**	**THEN**
	SAS noise	SAS count	C noise	B count
morning	less than 62dB	NaN	between 52dB and 54dB	Max
evening	NaN	NaN	NaN	Min
noon	greater than 62dB	greater than 275	NaN	Max
noon	greater than 66dB	between 177 and 275	less than 52dB	Min

Condition attributes are SAS noise, SAS count, and C noise, while the decision attribute is B count. Looking at the contents of the previous table, it can be concluded that the conditions for the number of vehicles to be maximal in the morning are: noise at the SAS location must be less than 62dB AND C noise must be between 52dB and 54dB. In the evening, the number of vehicles near the hospital is always minimal. Conclusions on the number of vehicles near the hospital around noon were made in an analogous manner. The NaN value indicates a parameter that does not affect the number of vehicles near the hospital area.

Even more general conclusions can be drawn from the analysis of semi-linguistic summaries. A total of 7 semi-linguistic summaries were generated, and the following three were selected, which refer to the minimal or maximal number of vehicles in the vicinity of the hospital. What follows are three semi-linguistic summaries and accompanying truth values that are an indicator of their truthfulness.

1. IF (morning) THEN SOME are max (0.852)
2. IF (noon) THEN SOME are max (0.852)
3. IF (evening) THEN MOST are min (1.0)

The interpretation of these rules is very simple: In some situations that occur at noon or in the morning, the number of vehicles around the hospital is maximal, while for most situations that occur in the evening the number of vehicles around the hospital is minimal. Such rules provide a very general insight, and their benefits come to the fore in the case of large data sets.

V. Conclusion

The main aim of this research is to provide a strategy of continuous elimination of exceeding levels of traffic noise and air pollution in urban areas produced by cars, trucks, trains, airplanes, etc. This is very important in order to protect human health and cultural monuments.

The described method enables the estimation of the number of vehicles at a certain location, based on the measurement of the amount of noise and the number of vehicles at several locations in the urban environment. In this study, the selected location was the hospital and its surroundings, and noise and vehicle number measurements were performed at three distant locations. The method offers an estimate of the onset of traffic congestion based on the values of the measured quantities. Specifically, for the example mentioned in this paper, during the morning hours, if the noise around the main bus station is less than 62dB, and in the city center zone the noise is between 52dB and 54dB, then the number of vehicles around the hospital is maximal. Also, around noon, the number of vehicles around the hospital will be maximal if the noise around the main bus station is greater than 62dB, and the number of vehicles is greater than 275.

Namely, as soon as the values of noise and the number of vehicles begin to approach these preset threshold values, it means that traffic congestion at the selected location is imminent. This allows for a timely response to traffic rerouting to avoid traffic congestions. Of course, the values of the thresholds vary from case to case, but the method enables their determination and gives an insight into the situations that connect the noise and the number of vehicles. This certainly has a direct impact on the emission of PM2.5 and PM10 particles, although their number is not related exclusively to traffic and motor vehicles.

Future work will include larger datasets and additional experiments with various sensors. Also, extensive testing on large data sets will be conducted, and ways to automate the entire process will be considered.

ACKNOWLEDGMENT

Ministry for scientific and technological development, higher education and information society of the Republic of Srpska, supports this research under the project ”Smart system based on IoT technology designed for monitoring of traffic air pollution”, contract number 19.030/3-2-25-2/1.

REFERENCES

[1] Jotanovic, G., Brtka, V., Curguz, Z., Stojcic, M., & Eremija, M. (2018, October). Mobile Applications for Recording Road Traffic Noise. In Proceedings/8 th International conference on applied internet and information technologies (Vol. 8, No. 1, pp. 94-98). “St Kliment Ohridski” University-Bitola, Faculty of Information and Communication Technologies-Bitola, Republic of Macedonia.

[2] Zakon o zaštiti životne sredine (2007) ,“Službeni glasnik Republike Srpske”, br. 53/02 i 109/05, član 20.

[3] Đurić, B., Trifunović, A., Čičević, S., Josić, A., & Stokić, I. (2018). MERENJE EMISIJE BUKE MOTORNIH VOZILA NA TERITORIJI GRADA KRAGUJEVCA. *Put i saobraćaj*, *64*(1), 61-66.

[4] Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise - Declaration by the Commission in the Conciliation Committee on the Directive relating to the assessment and management of environmental noise. Offi cial Journal L 189, 18.7.2002, p. 12–25.

[5] European Environment Agency. Noise in Europe (2014). EEA Report No. 10/2014. Luxembourg: Publications Office of the European Union, 7. (https://www.eea.europa.eu/publications/noise-in-europe-2014/file).

[6] Kooperatives Lärmsanierungsverfahren, initiirung und begleitungder pilotprojekte, Endbericht, SYLVIE - SYstematisch Lärmsanierung von innerstädtischen WohnVIErteln, Wien, October 2002.

[7] Božić, J., Ilić, P., & Bjelić, L. S. (2018). ECONOMIC ASPECTS OF THE CITY TRAFFIC NOISE: CASE STUDY//EKONOMSKI ASPEKTI BUKE OD GRADSKOG SAOBRAĆAJA: STUDIJA SLUČAJA. *EMC REVIEW-ČASOPIS ZA EKONOMIJU*, *15*(1).

[8] D. Petković, Š. Goletić, (2004), KOMUNALNA BUKA KAO ZNAČAJAN FAKTOR KOD PROJEKTOVANjA I GRADNjE SAOBRAĆAJNE INFRASTRUKTURE, Okolišni menadžment, Fojnica.

[9] Pawlak, Z. (2007). Skowron, A. Rudiments of rough sets. Journal of Information Sciences, vol. 177, pp. 3–27.

[10] Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. European Journal of Operational Research, vol. 99, pp. 48–57.

[11] Pawlak, Z. (1999). Rough Classification. International Journal of Human–Computer Studies, vol. 51, pp. 369–383.

[12] Wu, D., Mendel, J. M., Joo, J. (2010). Linguistic summarization using IF-THEN rules. IEEE International Conference on Fuzzy Systems, (FUZZ) 2010, Barcelona, Spain, 18-23, July.

[13] Boran, F. E., Akay, D., Yager, R. R. (2016). An overview of methods for linguistic summarization with fuzzy sets, Review, Expert Systems With Applications, vol. 61, pp. 356–377.

[14] V. Brtka Vladimir, V. Makitan, E. Brtka, D. Dobrilovic, I. Berkovic , "LP-WAN Performance Analysis by Semi-Linguistic Summaries", AD HOC & SENSOR WIRELESS NETWORKS, (2020), vol. 48 br. 1-2, str. 145-165.

Free and Open Source Software Licenses and GitHub Repository

Z. Kazi
University of Novi Sad/Technical faculty “Mihajlo Pupin”, Zrenjanin, Serbia
zoltan.kazi@gmail.com

Abstract – This paper gives an overview of most often used licenses in free and open source software community with explanation of basic terminology in this area. GitHub repository basic characteristics and features are described. This repository offers organizing the source code of software and software packages, downloading free software, free distribution and installation. Choosing and setting the license on this commonly used repository with example is represented.

I. INTRODUCTION

In the past two decades free software and open source software became the backbone in information and communication technology. It runs on desktop and laptop computers, in mobile phones, on servers, for email receiving and sending, etc. It is embedded in household appliances, automobiles, industry machinery. [1]

The term “Free software” was introduced by Richard Stallman. It means that the software is free for use, installation, launch, execution, distribution, testing and modifications. The term “Open Source Software” means that the source code of the software is available to the user and can therefore be modified. The software may be free but not open source. Conversely, software can be open source but not free. [2]

In short, free and open source software (FOSS) are programs whose licenses give users the freedom to execute the program for any purpose, to study and modify the program, correct any errors and redistribute copies or the modified program without having to pay for services to previous authors. [2]

FOSS became an international phenomenon changing the ICT world based on proprietary software and licenses that gives the “Copyright” to the end users. Two major paradigms in FOSS community are FSF (Free Software Foundation) and OSI (Open Source Initiative). [3]

II. FREE AND OPEN SOURCE LICENSES

Main characteristics of FOSS software [2]:

- It is most often developed in teams of developers over the Internet with intensive use of repositories and control version systems.
- FOSS software support is usually available via the Internet in the form of chats, forums or mailing lists.
- Upgrades and future versions of software are commonly free of charge.
- There is no limit to the number of instances of FOSS software to be installed.
- There are many companies formed to support the organization of businesses that use FOSS.

Open source licenses are divided into two groups:

- Licenses that aim to preserve the freedom and openness of the software itself (“Copyleft” licenses).
- Licenses aimed at giving freedom to software users (“Permissive” licenses).

The most used licenses in the field of free and open source software are [4]:

- General Public License (GPL).
- Lesser General Public License (LGPL).
- Berkley Software Distribution (BSD).
- Mozilla Public License.
- Apache License.
- MIT License.
- Free Document License (FDL).

General Public License (GPL) is the most widespread FOSS license (“Copyleft” type). It has the largest restrictions. It allows users to use and modify software but obliges users that modified versions of software must also be licensed under the GPL. It requires that the original version of the modified GPL software will be available to all users who get only binaries or executable in released versions. The terms of use the GPL license must be available to anyone who receives a copy of the software that contains the GPL license. Any user who adheres to the terms of use has permission to modify the software as well as yes copies and redistributes the software or any derived version. Software under the GPL license is under copyrighted rights, as the user has no right to redistribute it, even in a modified form, unless it is under license terms allowed. [4]

Lesser General Public License (LGPL) - Translated to mean a less valuable GPL license, the other the most common license in the FOSS license. It is intended exclusively for the libraries used by the software. Because FOSS software is written in modules, this license protects the source code modules, i.e. libraries in the same way as well as the GPL license. The difference is that under the LGPL license, "non-free" software can also be used which uses certain modules. The FSF does not encourage the use of this license because the work of the developer is granted to owners of “non-free” software who do not have to give anything to the community from which the software was obtained. [4]

Berkley Software Distribution (BSD) is the third most widely used FOSS license, one of the most open software licenses, has a small number of restrictions on the use of the software. Redistribution and use of the software in its original or binary form, with or without modification, is permitted if redistributions the source code contains a copyright notice, a list of conditions and a disclaimer, if redistributions in binary form reproduce the copyright notice, the list of conditions and the statement on disclaimer (in case of damage) in the documentation and/or other materials supplied with software distribution. The names of organizations and associates cannot be used in derived software without previous permits. [4]

Mozilla Public License is also a “Copyleft” type of license. The users are allowed to copy, distribute, use and modifies the software under the following conditions: the software is provided as is, without warranty of any kind types in connection with the use of the software and the author is not liable for any damage resulting from its use.

Apache License - Redistribution and use of software in source or binary form, with or without modification, is allowed if the following conditions are met: redistribution of the source code or binary form of the sea have a copyright notice, list of conditions, disclaimer; documentation of the final the user involved in the redistribution, if any, must contain a certificate: "This product includes software developed by the Apache Software Foundation (http://www.apache.org/) "This certificate may also appear in the software itself; the names" Apache "and The Apache Software Foundation may not be used to encourage or promote products derived from software without prior written permission; derived software products may not be named "Apache" and "Apache" may not appear in the name without the prior written permission of the software the Apache Foundation. Apache version 2.0 is compatible with GPL 3. This is a “lenient” license because it is not requiring that modified versions be released under the same license, but author information must be kept in individual file

MIT License was created in 1988 at MIT University for X Windows System, under which it is still licensed today. This license was often changed in past. It is almost the same as the BSD license, with the difference that we are allowed with the MIT license use the name of the institution / individual in further modified versions of the software.

FDL (CGNU Free Document License) - a free license for books, texts, manuals, anyone can copy and redistribute the documentation, but must contain the names of the authors, who are not responsible for the modifications. Species “Copy left” license which means that derivative works must be under the same license

The most used licenses in the field of FOSS software are GPL 48%, LGPL 23%, BSD 14%, Apache License 6%, MIT 5%, and all others are represented by 4% in total.

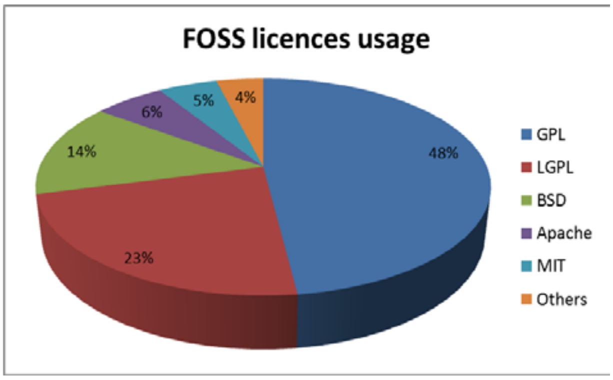


Figure 1. FOSS licenses usage chart

An alternative to FOSS is Proprietary and Closed Software Source Software. Proprietary software that is purchased is protected by copyright (and intellectual property) laws and is subject to various restrictions on the use of copies, most listed in the EULA (End User License Agreement) with the end user. Restrictions on the software are strictly determined by the software owner, reseller, or development manager. The owner can be an individual or an organization. This is the most common type of software, when the installation version of the program is delivered to the user, with the following restrictions: the number of computers on which the purchased software can be installed, especially when it comes to server variants of software, then how many clients the server can connect to software via key/license code, etc.

III. GitHub Repository

GitHub is one of the most famous and most frequently used CVS systems (Control Version System), i.e. repositories with basic aim to store and organize the source code of software and software packages. Software can be downloaded, distributed and installed from these locations. It stores all audits performed on the software to control changes in the software, they provide communication between team members. There are over 10 million repositories created on GitHub last six years, making it for one of the most important sources code on the Internet. The author is Linus Torvalds, creator of the Linux operating system kernel. GitHub is free software and open source software distributed under the terms of the GNU General Public License Version 2 (GPL). [4]

Git's extremely important function is the branching model. It allows and encourages programmers to have more local branches that can be completely independent of each other. Create, merge and delete branches makes

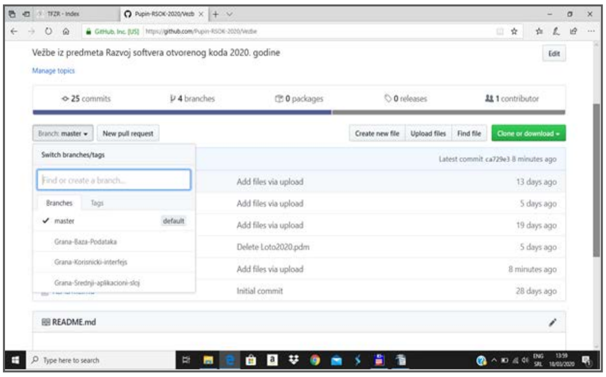


Figure 2. GitHub repository – branches and files

the software development takes extremely fast and easy (figure 2). Most operations in Git only require local files and resources to work - there is usually no information from another computer on the network. Since we have the entire project history on the local disk, most operations look like almost instantaneously. To view the project history, Git does not have to go to the server to get the history and displays it - simply reads it directly from the local database. That means seeing the history of the project almost immediately.

If you want to see changes that have been made between the current version of the file and the file before a month ago, Git could preview the file a month ago and make a local difference calculation instead requires the remote server to do so. You can also download an old version of a file from a remote server to do it locally. Everything in Git is checked before it is saved. That means its impossible change the contents of any file or directory without information about it. This functionality is built-in in Git at the lowest levels and is an integral part of the philosophy of the repository. There can be no loss of information either file corruption occurs without Git being able to detect it. [5]

Although it can be done manually, Git has several advantages that cannot be done manually:

- Possibility to undo changes - if an error is made, you can go back to the previous point in time to recover the work version.
- A complete history of all the changes - if we ever want to see what the project looked like day, week, a month or a year ago, we can check the previous version of the project to see exactly what it is like the state of the files then either.
- Documentation of why the changes were made - it is often difficult to remember why the changes occurred changes. With engraving messages, it's easy to document for future reference why you make changes.
- The certainty that something will change - since it is easy to restore a previous version of a project, it can be had confidentiality of any change you want. If they fail, you can always revert to an earlier version work.
- Multiple history flows - different branches of history can be created to experiment with different one's content changes or to build different functions independently. Then they can reconnect into the history of the main project (master, ie the main branch) or delete them if they do not work in the end.

IV. FOSS Licence and GitHub Repository

The first step in using the GitHub system is to create a user account, with the "Sign up" option. The following is the entry of the username, email address and password. Then it is necessary to do user verification. The type of repository is chosen: "Individual" - free or "Team" - professional, $7/month. If "Individual" is selected, the limit is the number of associates, maximum 3, which makes a total of four members for a private repository.

This restriction does not apply to the public and can be changed later in the settings. User login is via the "Sign in" option and consists of entering a password and username, and the option to send a new password due to a forgotten user password.

On the main page, it is possible to create a new repository by selecting "Your repositories". This is followed by creating a new repository, entering a name, description, type (public or private). Creating a new project is done by entering the project name and a short description of the application/project, and you can also select a predefined type from existing templates. Adding new files is done by finding files or folders on the computer. The technique is to drag a file, multiple files, or an entire folder.

It is possible, but not mandatory, to create new branches in the repository. The initial branch is "Branch: master". All newly created branches contain the complete content of the master branch at the time of creation, so each project participant can work within one, ie. its branches or more. Adding files to a branch is done similarly to the "master" branch.

When adding a new collaborator, programmer or developer, the user receives a message via the email address he entered during registration and must confirm and accept cooperation on the project.

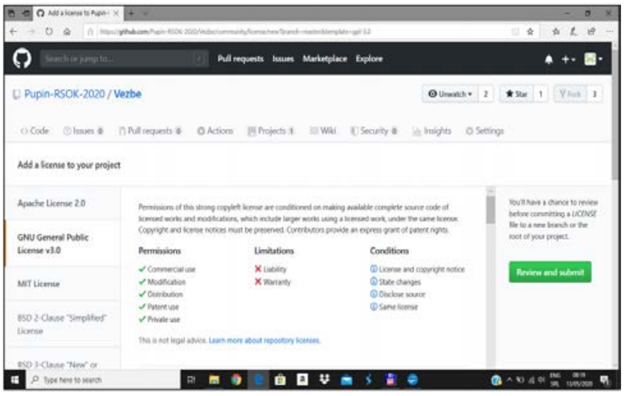


Figure 3. GPL licence characteristics on GitHub repository

The option to transfer files to the "master" branch is "New pull request", and it can also be done by displaying a list of recently updated branches, via "Compare & pull request".

Downloading files from the repository is done with the option "Clone or download", where then follows the choice of download method, the first is in the desktop application GitHub which, of course, must be installed on a local computer, or download the entire content via a packaged archive.

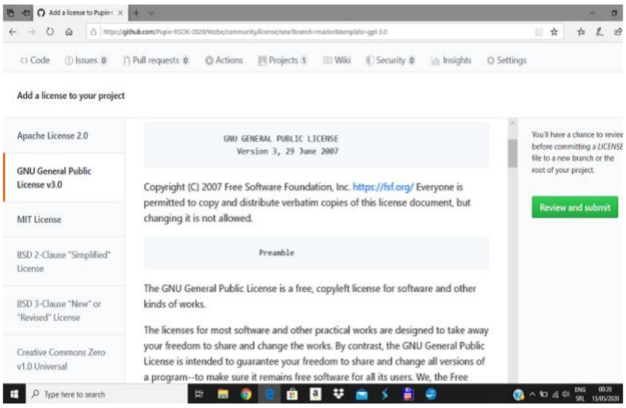


Figure 4. GPL licence text on GitHub repository

The most important decision in setting the FOSS license is choosing between GPL license and a less restrictive license [4]. In this example GPL as the most restrictive license is chosen.

Setting the license is done by first creating a new file with the option "Create new file". Then enter the name of the file in which the license will be located ("License.txt"). The following is a selection of existing licenses in the FOSS world. For each individual type of license, its limitations and characteristics are shown.

GitHub has built-in templates for twelve FOSS licenses [6]:

- Apache License 2.0.
- GNU General Public License v3.0.
- MIT License.
- BSD 2 – Clause “Simplified” License.
- BSD 3 – Clause “New” or “Revised” License.
- Boost Software License 1.0.
- Creative Commons Zero v1.0 Universal.
- Eclipse License 2.0.
- GNU Affero General Public License v3.0.
- GNU General Public License v2.0.
- GNU Lesser General Public License v2.1.
- Mozilla Public License v2.0.

The example on figure 4 shows the contents of the GPL v3.0 license file while the restrictions are shown on figure 3. Confirmation is with the option "Commit new file", and then it is necessary, as with merging application branches, to execute the "Create pull request", then the "Merge pull request" and finally the "Confirm pull request" which completes the process of creating a license.

License file initial location is root folder on the master branch of software project.

V. Conclusion

In the last two decades the free and open source software and programs with licenses that gives user the freedom to execute the program for any purpose, free for use, installation, launch, execution, distribution, testing and modifications became widely accepted in ICT world.

GitHub repository and CVS offers organizing the source code of software and software packages and units, downloading free software, it’s distribution and installation. It stores all changes performed on the software with control changes in the software, provides communication between team members according to chosen repository type. GitHub has attracted many experienced developers thanks to its robust and reliable features as well incredibly fast branching capability.

Choosing existing or creating your own FOSS license at GitHub repository is a simple task technically while choosing the appropriate license for an ICT project can be a difficult and thankless decision. Creating your own license implies knowledge in the law area and intellectual property protection.

References

[1] K. Fogel, “Producing Open Source Software - How to Run a Successful Free Software Project”, Karl Fogel, under the CreativeCommons Attribution-ShareAlike (4.0) license, 2005-2017.

[2] K. Wong, P. Sayo, “Free/Open Source Software – A General Information”, the United Nations Development Programme’s Asia-Pacific Development Information Programme (UNDP-APDIP), Kuala Lumpur, Malaysia, 2004.

[3] V. Van Reijwood, A. De Jager, “Free and Open Software for Development”, Polimetrica, Milan, Italy, 2008.

[4] A.M. St. Laurent, “Understanding Open Source and Free Software Licencing”, O’Reilly, 2004.

[5] S. Chacon, B. Straub, “Pro Git – second edition”, A Press, 2014.

[6] GitHub repository, http://github.com.

IoT Based Predictive Computing

Andreja Samčović
University of Belgrade – Faculty of Transport and Traffic Engineering, Belgrade, Serbia
e-mail: andrej@sf.bg.ac.rs

Abstract – IoT predictive computing and modeling are described in this paper. These techniques have recently gained attention, due to advances in data warehousing, data management, machine learning, statistical modeling, and so on. Today, various applications include those technologies, like e-health, personal and home services, e-transportation, e-logistics, etc. However, IoT systems face a lot of challenges which need to be handled carefully.

I. INTRODUCTION

The IoT (Internet of Things) is a research area that is constantly expanding and evolving thanks to a growing number of applications and various intelligent devices [1]. Connected devices and widely interconnected objects can be recognized by the fact that they are equipped with sensors, computer and communication elements [2]. IoT-based devices enable fast calculations and transfer of sensory data to a connected smart device. Generally speaking, these devices include internet-oriented communication-based parts known as "middleware". The devices are connected via different protocols and have different network topologies (star topology, ring topology, mesh topology, etc.).

The principle of operation of smart devices is as follows - sensor devices forward their collected data to devices that process them, extract the necessary information and further process them as needed. In the IoT scenario, a computer framework allows several intelligent sensors and connected devices to quickly provide services and share knowledge across different platforms. During the information sharing i.e. acquired knowledge, the devices first consider the presence of other devices, objects or things in the environment and that are connected via internet communications (for example, wireless or wired communication networks). Distribution of unique addressing schemes is used to connect and address devices in order to provide patterns for device communication. These computer platforms collaborate with other objects and things in order to achieve common goals by creating new applications and efficient services.

This paper is organized as follows: the second section after introduction describes different IoT applications, with an emphasis on personal and home services, as well as on health care. Several problems and challenges are described in the following section, including technical challenges, device life, efficiency, problems with big data, and incomplete data presentation. Furthermore, some IoT predictive models are described, followed with concluding remarks.

II. IOT APPLICATIONS

The different application domains are discussed in details in this section. These diverse applications are categorized based on the availability of different types of resources. Resources include network availability, coverage, data heterogeneity, repeatability of different processes, customer involvement, and their impact. IoT applications are shown in Fig. 1, while Fig. 2 represents an overview of IoT services [3].

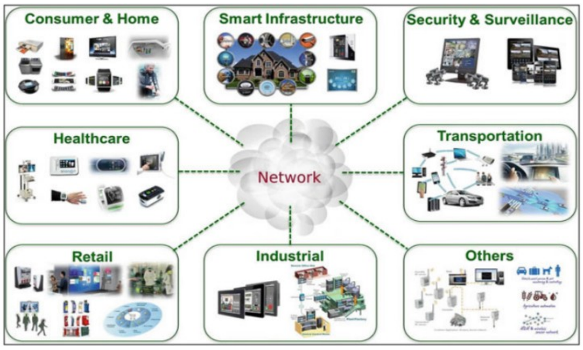


Figure 1. Different IoT applications [3]

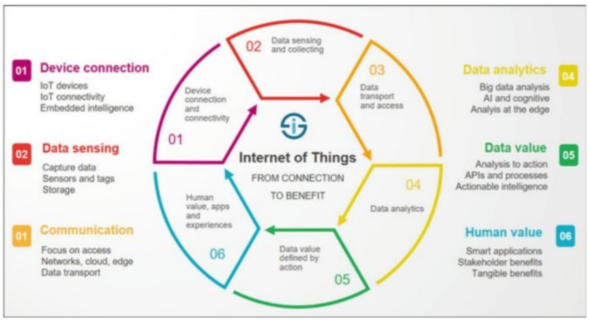


Figure 2. Various IoT services [3]

A. *Personal and Home Services*

Information collected from sensors, such as room temperature, humidity, etc. are used for various applications. They are used by the owners of the network. In general, internet is used as a framework that enables faster transfer of large data (such as video data), as well as faster collection of samples.

In the field of health and health monitoring, sensors have been present for two decades. As for smart devices such as mobile phones, tablets and even some types of intelligent sensors, they can be used to collect sensor data and send it via communication protocols in accordance with the interfaces. Communication protocols can be Bluetooth, WiFi or some other interfaces that connect to devices and allow data to be measured based on certain physiological parameters. With the help of networks, it is possible to monitor elderly people and their health without the need for a personal presence in their homes [4]. Medical staff has access to this information at any time of the day. It can be said that the home surveillance system, in addition to providing the individual with effective supervision, frees from paying for hospitalization services. When it comes to other home smart devices, we can talk about air conditioners, washing machines, automatic ovens, refrigerators, etc. All of them can be connected into a single monitoring and control system to allow easier control and energy save.

B. *IoT in Health Care*

As already mentioned, IoT plays a significant role in the field of health care. Services that use IoT are a system consisting of built-in intelligent sensors and devices that are interconnected and have the ability to exchange data over the internet. These services serve to monitor the condition of patients and possibly timely response of responsible physicians in case of critical conditions of patients [5].

The basic architecture consists of three levels. The first level is body sensors that include devices placed on the body, i.e. have the ability to attach to the body and collect data according to the part of the body to which they are attached. They function as separate data collection units. Blood pressure, heart rate, body temperature, etc. can be collected.

Figure 3. Block-diagram of data processing flow [3]

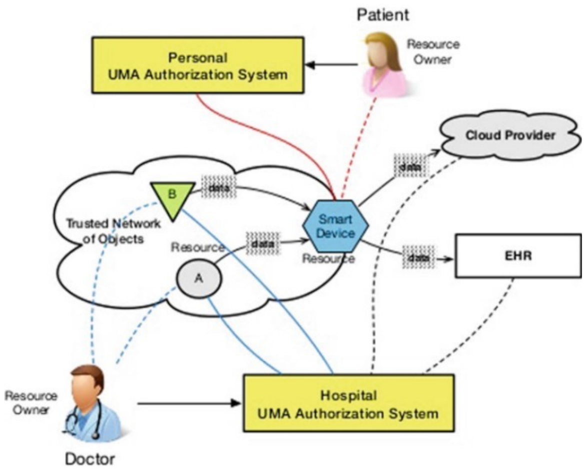


Figure 4. System for monitoring user’s health [5]

The second level consists of the establishing communication and mutual networking. The collected data is forwarded to the next level. The final, third level includes processing of the collected data as well as their analysis with the aim of the most accurate report on the condition of patients. Fig. 3 shows the layout of this architecture, while Fig. 4 shows the system for monitoring an individual’s health. This model uses computer techniques to possibly predict a particular event in a network of connected devices. Various world scientists and engineers are working on developing and perfecting such a model in order to solve major problems that may arise using analytical methods and a computer approach. These techniques provide the basis for creating an intelligent and smart world, that is, systems of connected sensors and other devices. Such connected devices enable the conversion of digital and virtual platforms into an energy-efficient environment and also enable fast communication and transport.

III. Main problems and challenges

A. *Technical Challenges*

Technical challenges in the field of IoT intelligent systems are various and are the subject of many researches [6]. A set of technical characteristics is required for different applications, especially for industrial applications. These applications include capabilities for observation and communication in terms of the type of intelligent sensors and high-speed sampling of received data, fast communication through wireless networks and Internet Protocols (IP), as well as data transfer between smart sensors and receivers. Moreover, device packaging and defined intelligent sensor protocols also play a vital role in the proper use of extracted functions and the transmission of vast amounts of data. These intelligent sensors, physically connected devices and their computing capabilities require a modeling system, based on predictive computing, to predict the results based on the learned system and the visible characteristics of the data, which are necessary for reliable operation.

B. *Device Life and Efficiency Problems*

In the field of IoT, intelligent devices based precisely on IoT are closely related to the lifespan of the main IoT device [7]. This is a big challenge when it comes to life expectancy. The mismatch of intelligent devices over a lifetime requires consideration to be given to the complete design and management of various organizations, configurations, and installations involving IoT devices. Moreover, energy management and full use of energy for intelligent devices are also one of the main challenges. This is especially true for constantly active devices. That is why saving energy is one of the main goals that can be achieved with a good budget.

C. *Problem with Big Data*

The storage and processing of large amounts of data is another significant problem in this area. IoT devices receive various types of data from many devices. They are able to collect a very large amount of raw data in a short period of time. For example, in ordinary industry monitoring, with 1000 sensors, temperature or pressure data are collected with samples of 10,000 per second, while audio or some other type of multimedia is collected with samples of 100,000 per second. Considering that data is collected for only 100 minutes a day, the total amount of data collected on an annual basis exceeds 1000 PB (petabytes). That's a huge amount of data. Storing all this data is very demanding. Predictable computing models are not able to perform any operations on this amount of data. Therefore, they must first be processed and analyzed to be available for basic use. Also, a major problem is the extraction of important information, which is the data needed to perform predictions and calculate the performance of computer devices based on predictive modeling. To solve all these problems, new computer systems based on predictive modeling and efficient algorithms for visualization and presentation of big data are needed.

D. Incomplete Data Presentation

Rare, uncertain and incomplete big data are discriminatory characteristics for data prediction applications. Given that the data are scarce, the number of characteristic points in the given data is too small to analyze and draw completely reliable conclusions. In the field of pattern recognition and machine learning, models and representative systems, a number of algorithms are also unable to present accurate data because sparsely collected data can significantly reduce the reliability of different prediction models and computer systems.

In order to reduce the dimension of huge amounts of data, methods such as data mining, machine learning, deep learning and other intelligent methods are used. After reducing the dimensions, the predictive model extracts characteristics from the reduced data types. Extracted data and functions are stored in a database. After that, algorithms are applied to select specific information of importance. With increasing demands, widespread application and increasing prevalence, predictive modeling and computer-intelligent IoT algorithms have significant roles in several disciplines such as medical sciences, weather forecasting, agriculture, etc.

IoT-based models and frameworks have been proposed to monitor users in real time and to provide better support if needed. These models provide a large number of benefits over classic and standard networks with existing learning capabilities, improved information security during communication, fast access to records and proper energy use [8].

IV. IoT Based Predictive Models

IoT-based predictive methods are discussed in this section. These methods use a coherent set of basic aspects of probability theory and statistics to solve major problems in large amounts of data (Big Data).

The different types of structural design for IoT data pose major challenges in the areas of data warehousing, security, data integrity, warehouse management, transferring vast amounts of data from source to destination server, and distributed real-time data networks. Therefore, IoT can influence the computation and sharing of information among intelligent objects on a different type of computational paradigm. To address these and other major challenges, IoT models based on predictive computing are needed.

A. Predictive Model

Predictive modeling is defined as the process of applying a statistical model or algorithm to excavate data in order to obtain data to predict the outcome of new or future observations or data sets. Moreover, in non-stochastic calculation of predictions for data sets, the goal is to predict the output Y for unknown predictions $x_1, x_2\, x_3$, etc. In the given observations it also includes a time prediction, where each observation is collected according to time T. The inputs are used to predict the results using future values at time $T + t$ where $t > 0$. In computing based on the IoT predictive model, the predictive model makes predictions of future observations. Data prediction involves accurate or prediction within an interval, prediction within a particular area, or prediction analysis based on data distribution. Predictive models choose any method of prediction, regardless of its approach. Moreover, models based on predictive computation extract relevant information using mathematical methods and a set of computational tools from existing data sets to identify patterns. This pattern identification helps to predict future outcomes from given unknown data sets.

B. Descriptive Model

Descriptive modeling is a type of modeling that aims to summarize or present a huge amount of data structure in a compact representative way. Unlike frameworks based on predictive data modeling, in descriptive modeling, reliance is a basic theory of cause that is absent or incorporated in a less formal way. Also, the focus is on a measurable level. The fit of a regression model for a given data can be descriptive if it is used for relationships between dependent and independent variables, rather than for causal inference or prediction.

C. Probalistic Learning Model and Statistical Analysis

For unverified data, the main problem is that each data item is presented as a distribution sample rather than an individual value, so most existing data collection algorithms cannot be applied. Common solutions involve data distributions into account to estimate model parameters. For example, error-based data mining uses mean and variance values for each individual data unit to create the so-called Naive-Bayesian predictive model for classification. Comparative approaches have also been applied to decision making or database queries. Incomplete data is defined as data that lacks values in the fields for a given sample or data. The lack of these values can be caused by various things such as sensor failure or some systematic strategy to intentionally skip some data values (e.g. skipping some data read from the sensor to save the energy needed for transmission). Most modern approaches of data mining have built-in solutions for missing data values (such as ignoring missing data fields), replacing data (replacing missing values to produce improved models).

D. *Predictive Analytics Algorithms for Data Mining*

Predictive analytics and models apply classification techniques to classify data. Classification is a machine learning technique for classifying data into different classes. The primary goal of the classification model is to predict the target variable (unknown variable) using matching techniques that are based on the similarities of the obtained results. The classification model generates results in binary form (for example, a decision to grant a loan to bank's clients) or in a certain form, when a set of input data is given (for example, credit rating, client's name, income, etc.). The data mining process involves forecasting customer requirements, based on data collected in various jobs. The classification model predicts big data using learning techniques to find the general relationship between the unknown variable and all the properties of the input data in the stored data set.

E. *Data Fusion*

Data fusion is a computer technique that connects multiple data received from different sensors or objects. In IoT, various sensors or intelligent devices are equipped with a model or frame of the primary source, and the received data are unstructured and disorganized. The received data must be merged (hence the name fusion). In the envisaged models, the concept of data fusion is applied to business data, telematics or original sensor data. In fusion mechanisms, less data is obtained from external sensors, such as time data or time databases. Such data is received by intelligent devices based on IoT. After that, predictive computers download the collected data and maps, store them in memory, compare and extract similarities between the data. Predictive models and frameworks based on data fusion predict and analyze data using supervised and unsupervised machine learning algorithms.

F. *Fault Prediction Using a Decision Tree Ensemble*

For prediction of failures Big Data records from sensors are also checked using ensemble classifier techniques, based on machine learning. Decision tree ensemble models are applied to teach data collected from sensors to predict future system failures based on past failures. Ensemble classification models use a gradient-based classification approach to represent data using regression trees. A regression tree is used to represent classified data in tree form. In this way, it builds an ensemble of tree classifications one by one. After that, predictions for individual trees are added to make a final estimate of the predictions of the given data. In the final regression tree, the consequent tree attempts to build a model to generate differences between the unknown data units, the tested data and the current value.

The ensemble classification model trains the reinforced backbone of the regression tree. The backbone consists of a series of trees in which fault data characteristics are coded. Based on the values of these characteristics, the model is trained in such a way that each tree can decide which set of records belongs to a particular item. Then, each record is assigned a certain importance, which indicates the evidence for or against the record that belongs to the faulty system. The model collects the proof values of all trees and as a result gives the probability of failure. Thus, the model shows the probability that a given record is an indicator of a faulty system. Fig. 5 shows a decision tree ensemble model created by a regression-based classification algorithm. The classification algorithm based on the regression tree belongs to supervisory learning methods. For each record, it indicates whether it belongs to a regular or a faulty system.

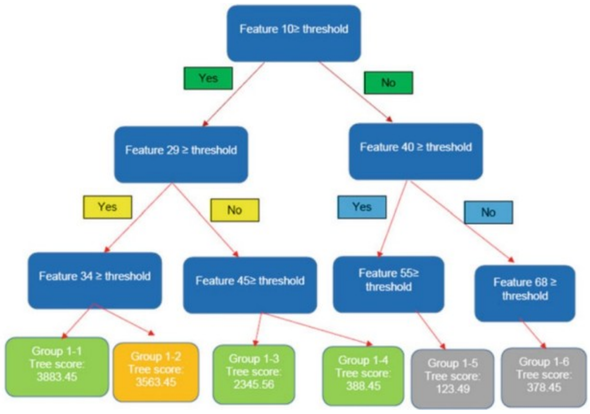


Figure 5. Decision tree ensemble model

V. CONCLUSION

The advances in technology have led to significant changes into existing environment that replace the current framework with new kind of intelligent frameworks. These new techniques are able to predict the information in real time or near real time with any intervention from user of the system. Today, technologies like cloud computing, IoT, pervasive computing, mobile computing, etc. play important role for making prediction. We have discussed IoT predictive computing and the insights of risks associated with the integration of these technologies. Predictive computing makes effective use of machine learning and data mining approaches to process collected data and produce the results in real time for consideration. The role of IoT based predictive computing has been presented to ensure our lives easier and more comfortable.

ACKNOWLEDGMENT

This paper is partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] JA Stanković: "Research directions for the internet of things," IEEE Internet of Things, Vol. 1, No. 1, p. 3–9, 2014.

[2] L. Atzori, A. Iera, G. Morabito: "The internet of things: a survey", Computer networks, Vol. 54, No. 15, pp. 2787–2805, 2010.

[3] PK Gupta, V. Tyagi, SK Singh (eds.): "Predictive Computing and Information Security", Springer, 2017.

[4] Y. Yang, l. Wu, G. Yin, L. Li, H. Zhao: “A survey on security and privacy issues in internet-of-things”, IEEE Internet of Things, 2017.

[5] A. Botta, W. de Donato, V. Persico, A. Pescape: “Integration of cloud computing and internet of things: a survey”, Future Generation Computer Systems, Vol. 56, pp. 684-700, 2016.

[6] Y. Chen, et. al.: “Big data analytics and big data science: a survey”, Journal Management Anal., Vol. 3, No. 1, pp. 1-42, 2016.

[7] A. Gavrovska, A. Samčović: „Intelligent automation using machine and deep learning in cybersecurity of industrial IoT: CCTV security and DDoS attack detection”, Chapter 8 in Cyber security of industrial control systems in the future internet environment, IGI Global, pp 156-174, 2020.

[8] A. Samčović: „Security issues in Internet of Things environment”, 8th International conference on business information security BISEC 2016, Beograd, str. 59-65, 15. oktobar 2016.

Framework for improving privacy behaviour in e-learning environment

S. Mitrović*, S. Čičević ** and A. Samčović***
University of Belgrade, Faculty of Transport and Traffic Engineering, Belgrade, Serbia
*s.mitrovic@sf.bg.ac.rs, **s.cicevic@sf.bg.ac.rs, ***andrej@sf.bg.ac.rs

Abstract - In this paper, the Framework for improving privacy behaviour in e-learning environment is proposed in order to mitigate the risk of academic online services reputation disruption and to enhance the personal e-safety, as well. As the framework strongly relies on the assessment mechanism, this paper also introduces the Self-Assessment Tool as the most convenient questionnaire for privacy attitudes examination among students. It is presented in the study, which results are used to measure students' attitudes towards privacy concern in e-learning environment employing a non-exhaustive list of variables.

keywords – privacy, attitudes, behaviour, Privacy Paradox, framework

I. Introduction

The e-learning environment could be described as a very dynamic area of different learning, teaching and research activities, especially in 2020, when protective measures against the COVID-19 pandemic situation contain different rules, which often oppose traditional forms of academic activities. An extensive set of online activities (that comprise LMS tools, collaborative lab apps, video conferencing, VR/AR apps, etc.) is often used to realize a curriculum in advanced way. However, this involves the use of various devices, not only school workstations and other equipment, but also personal devices such as smartphones/tablets and laptops. Hence, the use of different client devices could impose significant cybersecurity concerns including different kinds of issues based on compromised privacy. Although academic institutions established clear privacy policies in order to meet regulatory compliance, as well as high academic standards, students' security and/or privacy could be at risk due to personal behaviour based on so called Privacy paradox. This kind of behaviour could be simply explained as difference between students' concern about their privacy and willingness to protect their online personal data. On the other hand, there is dilemma if there is any privacy in online environment at the end of this decade anyway. Why is still important to protect the rest of the users' unrevealed private details? There are numerous answers on this question and some of them are related with the measures to protect the reputation of academic online services in form of risk prevention.

Having this in mind, the Framework for improving privacy behaviour in e-learning environment will be proposed in this paper, in order to meet two primary goals: the first is to mitigate the risk of academic online services reputation disruption and the second one is to enhance the personal e-safety [1]. Such a proposal is motivated by the identification of a lack of a comprehensive and integrated theoretical framework that would consolidate various streams of research into one model as a missing link in the literature in the online privacy concern field, and this research is meant to fill that gap. The proposed Framework strongly relies on set of different questionnaires, which have crucial role in testing students' attitudes regarding privacy issues, so the first part of the paper will consider different aspects of these kind of questionnaires.

The paper is organized in the following way. After the introductory remarks, the Literature overview is presented in Section II followed by the methodology in Section III. Numerical results obtained by the questionnaire are presented in Section IV. The appropriate framework is proposed in Section V, while the Section VI presents concluding remarks, as well as future research directions.

II. Literature overview

Regarding the question is there any privacy when online, intrigued is to know how much Internet users are nowadays aware or concerned about privacy intrusion. Online privacy involves the rights of an individual concerning the storing, reusing, provision of personal information to third parties, and displaying of information pertaining to oneself on the Internet [2, 3]. The invasion of privacy on the Internet includes the unauthorized collection, disclosure or other use of personal information [4, 5].

The Concept of Privacy Concerns refers to individuals' beliefs about the risks and potential negative consequences associated with sharing information and it can be used as a privacy management predictor [6-10]. Albeit users declare themselves as has been bothered upon their privacy, yet insufficient effort is made to safeguard their personal information [11]. Multiple theories have explained this paradoxical behaviour from a rational perspective by arguing that users weigh the cost-benefit ratio of online information disclosure both consciously and rationally resulting in a risk-benefit calculation that ultimately chooses benefits over risks [11].

Risk-benefit calculation plays a major role in the context of information privacy, known as the freedom to decide with whom, how and to what extent personal data is shared [12]. Even though social networks users believe they have been taken appropriate actions to be in charge of the flow of their personal data, this does not necessarily lead to significant decline in the disclosure of personal information because of their inability to hang together such huge amount of information [11]. Even supposing better privacy protection strategies/privacy rules, as a consequence of a certain degree of risk perception, this does not represent enough incentive for users to apply them [11, 13]. This results in application of simplified mental models that are often favour the benefits [14]. Some situational cues, as well as, a kind of lack of self-discipline mitigate potential risks in the distant future as users reveal a tendency to favour immediate rewards [11]. This implies that users are lacking in privacy protection knowledge, at both the technological and legal levels; leading to misinterpretation of the likelihood of actual privacy violations and to inaccurate predictions of future hazards [11].

The Concept of a “Privacy Paradox”, initially formulated by Susan Barnes [15] to define the perplexing divide between privacy-concerned adults and self-disclosing digital teenagers, has unfold. Research on online behaviour has revealed discrepancies between user attitude (concerns about privacy) and their actual behaviour in respect to online self-disclosure and privacy management strategies (the so called “Privacy Paradox”) creating the tendency towards privacy-compromising behaviour [9], [16-18].

One of the most frequently acknowledged concepts in explanation the decision to disclose personal information on the Internet is privacy calculus which represent a cost-benefit trade-off analysis that accounts for weighing benefits against costs and potential risks associated with any online action, or transaction [12], [17], [19-21], and which to a great degree counts on individual sensitivity towards what is recognized as privacy invasive [22]

Online self-disclosure despite privacy concerns could also be explained focusing on user trust, a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviours of another, or by their lack of risk-awareness and missing knowledge related to the potential harm associated with online self-disclosure [23-25]. Such a lack of understanding and awareness may be due to users’ digital skills, or lack thereof [17].

Internet users' social privacy (fear of intrusion caused by other people) concerns are much more pronounced than their institutional privacy which deals with companies or public institutions. Furthermore, users more readily adapt their protection behaviour to social than institutional privacy concerns with an emphasis on individual trust, awareness and understanding [17], [26].

More recent, Hoffmann et al. [17], introduce the concept of “privacy cynicism” embodies a cognitive coping mechanism [27] that is not based on a disparity between attitudes and behaviour, encompassing attitudes of uncertainty, powerlessness and mistrust towards the handling of personal data by online services, rendering privacy protection behaviour subjectively futile [17]. It contributes to psychological understanding which underline reshaping of attitudes to perceived circumstances [17].

What is more decisive, the online privacy concern might influence adoption of new technologies, future usage of online services, and other types of protective behaviour i.e. decision-making on how to withhold, fabricate or additionally protect important personal data [28]. It is pivotal to understanding the process of privacy concerns development and perception, as well as the impact they might have on interactions with other individuals, groups, agencies, and vendors [29].

Paramount in an attempt to help users to avoid this contradictory and potentially adverse behaviour amalgamate elevating privacy awareness with measures supporting their privacy decisions [30]. The need arises that some important questions to be answered. Weather and how willingness to disclose certain types of personal information affect students’ intention to limit their privacy exposure, and if they can serve to predict protective behaviour and practices? What are the moderating factors between online privacy concerns and related privacy behaviours [8], [31-32].

Privacy is not yet integrated into the social presentation of e-learning environment and hence, will consequently lead to failed privacy awareness [8]. Thus, it is substantial to uncover the actual barriers to adoption of privacy-protective measures, to throw the light upon the decision-making process that allow for privacy awareness and to develop policies suitable to various students’ needs [8].

Although previous studies have proposed various variables, concepts, and tested different theoretical models of antecedents and consequences of online privacy concern, there is no single widely accepted model of online privacy concern [33-35]. Online privacy literature at the first place deals with the problem of how to measure privacy concern of Internet users. The intensity or range of online privacy concern is hard to measure while it is highly subjective [36].

Thus, the objective of the study is to measure students’ attitudes towards privacy concern in e-learning environment employing a non-exhaustive list of variables in order to offer an integrative framework of online privacy protection behaviour that covers optimized Antecedents - Privacy Concern - Outcomes model.

III. METHODOLOGY

The Self-Assessment Tool used in this study consists from the measurement scales for privacy concern developed in the literature [3], [28-29], [37] and adapted for the internet environment. The questionnaire consists of 15 questions covering 4 dimensions that may have impact on the students’ behaviour in educational context. The structure of the questionnaire comprised the following: Personal Internet Benefits (PIB) (3 items); Attitudes Towards Collection of Personal Information (ATC) (3 items); Control of personal information online (CPI) (4 items); Sharing private information online (SPI) (5 items); questions considering preferences towards using new online services /technologies immediately after they are available (2 items); including also socio-demographic characteristics of respondents. Items used to calculate these variables were measured on a Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree). A small convenience sample of students enrolled in undergraduate Traffic psychology course at the University of Belgrade - Faculty of Transport and Traffic Engineering was selected to participate in an online survey about attitudes towards online privacy concerns.

IV. RESULTS

For the purpose of data analysis, first, descriptive statistics was computed. Data analysis was conducted using Microsoft Office Excel. Subscale scores are calculated by summing the scores on the relevant items and dividing that number by the total number of items within the subscale. Obviously, the dimension Attitudes towards Collection of Personal Information (ATC) received the lowest mean score, while Control of personal information online (CPI) dimension scores show the highest values followed by the Sharing private information online (SPI) mean scores values (Fig. 1).

The participants are hesitative upon the judgments that web sites seeking information should be transparent over data collection and usage, as well as when considering the benefits of gathering information being greater than their concern for privacy. The need for online privacy is moderately manifested, reflected in the opinion that an individual should be able to control personal information when online and that gathering information should not be conducted without the individual’s consent, although receiving the highest mean scores values in comparison to all other items. Similar standpoint is shown in respect of the attitudes towards collection of personal information and a strong consent over ability to control personal information online. The highest animosity was shown for the statement that websites track online activities and sharing private pictures on the Internet (Table 1).

TABLE 1. DESCRIPTIVE STATISTICS OF QUESTIONNAIRE ITEMS

	N		Mean	Std. Error of Mean	Median	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Range	Minimum	Maximum
	Valid	Missing												
PIB1	60	0	3.200	0.140	3.000	1.086	1.180	-0.578	0.309	-0.341	0.608	4.000	1.000	5.000
PIB2	60	0	2.967	0.145	3.000	1.119	1.253	-0.608	0.309	-0.781	0.608	4.000	1.000	5.000
PIB3	60	0	2.883	0.139	3.000	1.075	1.156	-0.353	0.309	-0.915	0.608	4.000	1.000	5.000
PIBavg	60	0	3.017	0.131	3.000	1.015	1.030	-0.672	0.309	-0.569	0.608	3.667	1.000	4.667
ATC1	60	0	2.450	0.143	3.000	1.111	1.235	0.092	0.309	-1.028	0.608	4.000	1.000	5.000
ATC2	60	0	2.650	0.140	3.000	1.087	1.181	-0.154	0.309	-0.936	0.608	4.000	1.000	5.000
ATC3	60	0	3.150	0.140	3.000	1.087	1.181	-0.554	0.309	-0.225	0.608	4.000	1.000	5.000
ATCavg	60	0	2.750	0.106	3.000	0.825	0.680	-0.026	0.309	0.373	0.608	4.000	1.000	5.000
CPI1	60	0	3.233	0.131	3.000	1.015	1.029	-0.894	0.309	0.201	0.608	4.000	1.000	5.000
CPI2	60	0	3.317	0.133	4.000	1.033	1.068	-0.775	0.309	0.138	0.608	4.000	1.000	5.000
CPI3	60	0	3.600	0.143	4.000	1.108	1.227	-0.765	0.309	0.132	0.608	4.000	1.000	5.000
CPI4	60	0	3.600	0.151	4.000	1.167	1.363	-0.812	0.309	0.008	0.608	4.000	1.000	5.000
CPIavg	60	0	3.438	0.125	3.750	0.966	0.934	-1.035	0.309	0.786	0.608	4.000	1.000	5.000
SPI1	60	0	3.433	0.137	4.000	1.064	1.131	-0.782	0.309	0.327	0.608	4.000	1.000	5.000
SPI2	60	0	3.300	0.145	3.500	1.124	1.264	-0.478	0.309	-0.462	0.608	4.000	1.000	5.000
SPI3	60	0	3.350	0.148	4.000	1.147	1.316	-0.663	0.309	-0.310	0.608	4.000	1.000	5.000
SPI4	60	0	3.383	0.149	4.000	1.151	1.325	-0.601	0.309	-0.199	0.608	4.000	1.000	5.000
SPI5	60	0	2.850	0.161	3.000	1.246	1.553	-0.141	0.309	-0.925	0.608	4.000	1.000	5.000
SPIavg	60	0	3.263	0.127	3.400	0.982	0.965	-0.575	0.309	0.416	0.608	4.000	1.000	5.000

Notes:
* List of relevant variables and abbreviations in the questionnaire
* The items were measured on a 5-point Likert scale ranging from 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree. All indexes were calculated as a simple average of their items.

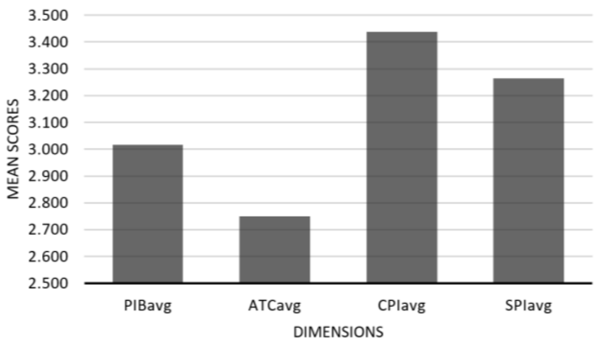


Figure 1. Main dimensions mean scores

There are significant positive correlations among Sharing private information online and Personal Internet Benefits, Control of personal information online as well as, with Attitudes Towards Collection of Personal Information dimensions average scores with the strongest relationship between the first two (Table 2).

TABLE 2. SPEARMAN'S RHO FOR AVERAGE DIMENSION SCORES

	PIBavg	ATCavg
CPIavg	.440**	.334**
SPIavg	.606**	.393**
***p<0.05, **p<0.01 (2-tailed)**		

The strongest relationship was found between the tendency to suppress privacy concerns for sake of the interest to obtain certain information or service from the Internet and posting on the Internet information about the place at the moment, as well as, with putting private information on the Internet and also with no concern to share private pictures on the Internet. Disclosure about the location, putting private information on the Internet and the need to obtain certain information or services were also highly correlated with the attitude that personal interest in the information wanting to obtain from the Internet overrides concerns of possible risk or vulnerability that may have regarding privacy.

On the other hand, the need to obtain certain information or services from the Internet that is greater than concern about privacy shows weak positive correlation with the statement of not seeing problem in sending credit card data when buying online.

Likewise, personal interest in the information that participants want to obtain from the Internet that overrides concerns of possible risk or vulnerability regarding their privacy is weakly related with the notion that when people give personal information for some reason, it should never be used for any other reason.

Worry that websites are collecting too much personal information shows moderate positive correlation with belief that personal information should not be used for any purpose unless it has been authorized by person itself, as well as with the view that online privacy is really a matter of ones right to exercise control and autonomy over decisions about how its information is collected, used and shared.

Concern that websites are collecting too much personal information is only weakly related with the statement of not seeing problem in sending credit card data when buying online.

Judgment that online privacy is a matter of ones right to exercise control and autonomy over decisions about how personal information is collected, used and shared show weak positive correlation with the opinion that they are not worried whether websites track their online activities, nor when websites ask for personal information. Lack of concern that websites track their online activities, or ask for personal information show no significant correlation with the attitudes that personal information should not be used for any purpose unless it has been

Personal Internet Benefits (PIB)

PIB1 In general, my need to obtain certain information or services from the Internet is greater than my concern about privacy.

PIB2 I find that personal interest in the information that I want to obtain from the Internet overrides my concerns of possible risk or vulnerability that I may have regarding my privacy.

PIB3 The greater my interest to obtain a certain information or service from the Internet, the more I tend to suppress my privacy concerns.

Attitudes Towards Collection of Personal Information (ATC)

ATC1 It doesn't bother me when websites track my online activities.

ATC2 It doesn't bother me when websites ask me for personal information.

ATC3 I'm concerned that websites are collecting too much personal information about me.

Control of personal information online (CPI)

CPI1 My online privacy is really a matter of my right to exercise control and autonomy over decisions about how my information is collected, used, and shared.

CPI2 My control of personal information lies at the heart of my privacy.

CPI3 Personal information should not be used for any purpose unless it has been authorized by that person.

CPI4 When people give personal information for some reason, it should never be used for any other reason.

Sharing private information online (SPI)

SPI1 I don't mind sharing private pictures on the Internet.

SPI2 I put private information on the Internet.

SPI3 I don't mind posting on the Internet information about the place I am at the moment.

SPI4 I don't mind posting on the Internet with whom I am at the moment.

SPI5 I see no problem in sending my credit card data when buying online.

authorized by a person itself and with the reasons of using personal information for various purposes, as well as with lack of concern for sharing private pictures on the Internet.

Indifference upon websites tracking their online activities is not related with carelessness of posting information about company, location and credit card. Accordingly, Attitudes towards collection of personal information seem to be not related neither to Control of personal information nor with Sharing private information online (Table 3).

TABLE 3. SIGNIFICANT SPEARMAN'S RHO FOR QUESTIONNAIRE ITEMS

	PIB1	PIB2	PIB3	ATC1	ATC2	ATC3
CPI1	.423**	.429**	.407**	.305*	.391**	.533**
CPI2	.355**	.369**	.420**	.283*	.280*	.464**
CPI3	.379**	.343**	.398**			.549**
CPI4	.318*	.282*	.354**			.497**
SPI1	.494**	.484**	.559**		.406**	.419**
SPI2	.480**	.519**	.600**	.289*	.497**	.353**
SPI3	.516**	.554**	.605**		.365**	.460**
SPI4	.464**	.493**	.514**		.303*	.410**
SPI5	.286*	.348**	.423**			.268*
*p<0.05, **p<0.01 (2-tailed)						

V. PROPOSED FRAMEWORK

In order to meet regulatory compliance, as well as high academic standards, the Privacy Paradox effect should be minimized as much as possible. This is not an easy nor a simple task – it should be rather considered as strategic planning process. Having this in mind, the proposed framework should be considered as the set of activities that should be continuously applied on each generation of students, during their whole four-year undergraduate period. Hence, all activities could be organized in two different layers (Fig. 2):

1. by period of occurrence (lower layer):

- First-year activities, which should be performed during first two semesters after generation enrolment;
- Next-two-years activities, which should be performed during next four semesters;
- Last-year activities, which should performed during last two undergraduate semesters.

2. by frequency of application (upper layer):

- Starting activities, which are repeated at the beginning of each schoolyear;
- Whole-year activities, which are performed during the whole schoolyear.

The whole process starts with the employment of Self-Assessment Tool. Then, responses given by each new student should be analysed in order to discover basic knowledge related to Privacy, as well as personal attitudes

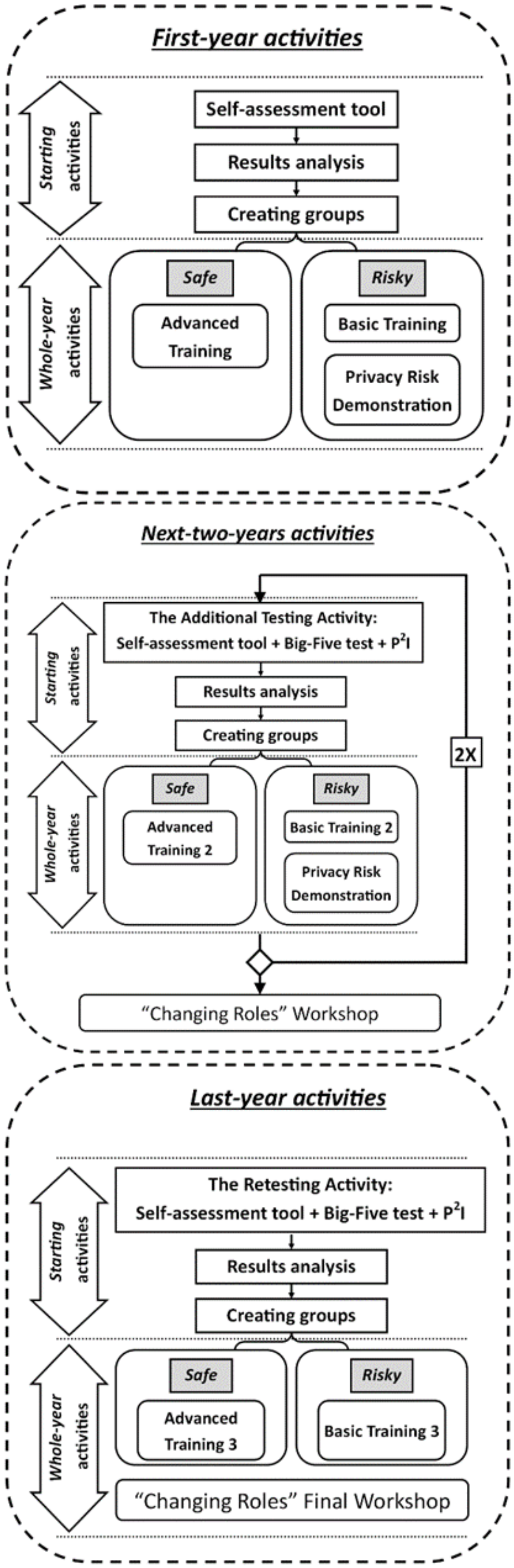


Figure 2. Proposed Framework block diagram

regarding privacy issues. These attitudes should indicate student membership to one of two possible groups of students:

1. The Safe group – the group of students who are aware of privacy issues and who are not susceptible to Privacy Paradox behaviour. These students need to slightly improve their knowledge about privacy issues with advanced training.
2. The Risky group – the group of students whose knowledge related to privacy is not satisfactory or students, which behaviour could be characterized as susceptible to Privacy Paradox. These students need to attend basic training classes in order to significantly improve their knowledge about e-safety concepts and Privacy issues, followed by Privacy issues demonstration exercises in order to familiarize themselves with Privacy Paradox problems.

When promoted to the second and third year, the Additional Testing Activity, that comprise the Self-Assessment Tool, the Big-Five Test [38] and the P2I Test [39] should be employed at the beginning of each academic year in order to discover the progress of each student regarding the general knowledge about Privacy, as well as attitudes toward the Privacy Paradox issues. Then, this activity should be followed by results analysis phase, which goal is to reconsider each membership to one of two abovementioned groups. It is assumed that each student should made certain progress, so both training activities (Simple and Advanced) should be enhanced, regarding the structure of presented topics. At the end of third academic year, it is assumed that student should be aware of Privacy issues enough to participate in a specific activity called the “Changing Roles” Workshop. This event is a kind of hackathon, where two groups of students are trying to discover important credentials to each other using methods to compromise privacy of imaginary private profiles within the virtual social network.

Finally, at the beginning of the fourth academic year, the Retesting Activity, which is also consisted of the Self-Assessment Tool, the Big-Five Test [38], as well as the P2I Test [39] should be employed in order to verify the additional progress in attitudes of each student regarding this topic. Like in previous years, the analysis and selection for the two abovementioned groups, followed by “more sophisticated” Basic and Advanced Training events, as well as the “Changing Roles” Final Workshop. This final event should result with “most uncertain” score, showing in that way minor (or even no) differences between students regarding the knowledge and attitudes toward Privacy issues.

VI. CONCLUDING REMARKS

The Privacy Paradox is one of well-known issues that may have a significant impact on privacy management in academic environment. However, the literature review discovered the lack of widely accepted model of online privacy concern, that could act as response on many identified issues. In order to minimise negative effects of Privacy Paradox on security in academic networks, the Framework for improving privacy behaviour in e-learning environment was proposed in order to minimize the effects of problematic behaviour regarding online privacy. Considering the fact that the framework strongly relies on the assessment mechanism, this paper also introduces the Self-Assessment Tool as the most convenient questionnaire for privacy attitudes examination among students.

The proposed framework is easy to be implemented in existing academic programme, it is not time consuming and it could improve privacy attitudes and behaviour in non-obtrusive manner. However it should be considered just as a starting point for further development in attempt to achieve the most efficient results in building personal privacy integrity.

ACKNOWLEDGMENT

This paper is partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia

REFERENCES

[1] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," *IEEE security & privacy,* vol. 3, p. 26–33, 2005.

[2] A. Ginosar and Y. Ariel, "An analytical framework for online privacy research: What is missing?," *Information & Management,* vol. 54, 2 2017.

[3] N. K. Malhotra, S. S. Kim and J. Agarwal, "Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model," *Information systems research,* vol. 15, no. 4, pp. 336-355, 2004.

[4] S. Bandyopadhyay, "Online privacy concerns of Indian consumers," *International Business \& Economics Research Journal (IBER),* vol. 10, no. 2, 2011.

[5] H. Wang, M. K. O. Lee and C. Wang, "Consumer privacy concerns about Internet marketing," *Communications of the ACM,* vol. 41, no. 3, pp. 63-70, 1998.

[6] H. Cho, J.-S. Lee and S. Chung, "Optimistic bias about online privacy risks: Testing the moderating effects of perceived controllability and prior experience," *Computers in Human Behavior,* vol. 26, no. 5, pp. 987-995, 2010.

[7] T. Zhou and H. Li, "Understanding mobile SNS continuance usage in China from the perspectives of social influence and privacy concern," *Computers in Human Behavior,* vol. 37, pp. 283-289, 2014.

[8] L. Baruh, E. Secinti and Z. Cemalcilar, "Online privacy concerns and privacy management: A meta-analytical review," *Journal of Communication,* vol. 67, no. 1, pp. 26-53, 2017.

[9] T. Dienlin and S. Trepte, "Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors," *European journal of social psychology,* vol. 45, no. 3, pp. 285-297, 2015.

[10] K.-W. Wu, S. Y. Huang, D. C. Yen and I. Popova, "The effect of online privacy policy on consumer privacy concern and trust," *Computers in human behavior,* vol. 28, no. 3, pp. 889-897, 2012.

[11] S. Barth and M. D. T. de Jong, "The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review," *Telematics and Informatics,* vol. 34, no. 7, pp. 1038-1058, 2017.

[12] H. Li, R. Sarathy and H. Xu, "Understanding Situational Online Information Disclosure as a Privacy Calculus," *Journal of Computer Information Systems,* vol. 51, no. 1, pp. 62-71, 2010.

[13] I. Oomen and R. Leenes, "Privacy risk perceptions and privacy protection strategies," in *Policies and research in identity management*, Springer, 2008, pp. 121-138.

[14] S. Potzsch, "Privacy Awareness: A Means to Solve the Privacy Paradox?," *IFIP Advances in Information and Communication Technology 298,* p. 226, 2009.

[15] S. B. Barnes, "A privacy paradox: Social networking in the United States," *First Monday,* 2006.

[16] B. Debatin, J. P. Lovejoy, A.-K. Horn and B. N. Hughes, "Facebook and online privacy: Attitudes, behaviors, and unintended consequences," *Journal of computer-mediated communication,* vol. 15, no. 1, pp. 83-108, 2009.

[17] C. P. Hoffmann, C. Lutz and G. Ranzini, "Privacy cynicism: A new approach to the privacy paradox," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace,* vol. 10, no. 4, 2016.

[18] S. Kokolakis, "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon," *Computers \& security,* vol. 64, pp. 122-134, 2017.

[19] S. Youn and W. Shin, "Adolescents' responses to social media newsfeed advertising: the interplay of persuasion knowledge, benefit-risk assessment, and ad scepticism in explaining information disclosure," *International Journal of Advertising,* vol. 39, no. 2, pp. 213-231, 2020.

[20] H. Krasnova, N. F. Veltri and O. Gunther, "Self-disclosure and privacy calculus on social networking sites: The role of culture," *Business \& Information Systems Engineering,* vol. 4, no. 3, pp. 127-135, 2012.

[21] H. Lee, H. Park and J. Kim, "Why do people share their context information on Social Network Services? A qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk," *International Journal of Human-Computer Studies,* vol. 71, no. 9, pp. 862-877, 2013.

[22] J. B. Hurwitz, "User choice, privacy sensitivity, and acceptance of personal information collection," in *European data protection: Coming of age*, Springer, 2013, pp. 295-312.

[23] M. Tschannen-Moran and W. K. Hoy, "A multidisciplinary analysis of the nature, meaning, and measurement of trust," *Review of educational research,* vol. 70, no. 4, pp. 547-593, 2000.

[24] M. Twyman, N. Harvey and C. Harries, "Trust in motives, trust in competence: Separate factors determining the effectiveness of risk communication," *Judgment and Decision Making,* vol. 3, no. 1, pp. 111-120, 2008.

[25] S. Trepte, L. Reinecke, N. B. Ellison, O. Quiring, M. Z. Yao and M. Ziegele, "A cross-cultural perspective on the privacy calculus," *Social Media+ Society,* vol. 3, no. 1, p. 2056305116688035, 2017.

[26] C. Lutz and P. Strathoff, "Privacy concerns and online behavior--Not so paradoxical after all? Viewing the privacy paradox through different theoretical lenses," *Viewing the Privacy Paradox Through Different Theoretical Lenses (April 15, 2014),* 2014.

[27] C. Lutz, C. P. Hoffmann and G. Ranzini, "Data capitalism and the user: An exploration of privacy cynicism in Germany," *new media \& society,* vol. 22, no. 7, pp. 1168-1187, 2020.

[28] I.-D. Anic, J. Budak, E. Rajh, V. Recher, V. Skare and B. Skrinjaric, "Extended model of online privacy concern: what drives consumers' decisions?," *Online Information Review,* 2019.

[29] T. Dinev and P. Hart, "An extended privacy calculus model for e-commerce transactions," *Information systems research,* vol. 17, no. 1, pp. 61-80, 2006.

[30] A. Deuker, "Addressing the privacy paradox by expanded privacy awareness--the example of context-aware services," in *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*, 2009.

[31] I. Ajzen and M. Fishbein, "The influence of attitudes on behavior.," 2005.

[32] M. Kezer, B. Sevi, Z. Cemalcilar and L. Baruh, "Age differences in privacy attitudes, literacy and privacy management on Facebook," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace,* vol. 10, no. 1, 2016.

[33] A. Gurung and A. Jain, "Antecedents of Online Privacy Protection Behavior: Towards an Integrative Model," in *Online Consumer Protection: Theories of Human Relativism*, IGI Global, 2009, pp. 151-164.

[34] I.-D. Anić, J. Budak, E. Rajh, V. Recher, V. Škare, B. Škrinjarić and M. Žokalj, The extended model of online privacy concern, 2018.

[35] K. Chen and A. Fadlalla, Online Consumer Protection: Theories of Human Relativism: Theories of Human Relativism, IGI Global, 2009.

[36] T. Buchanan, C. Paine, A. N. Joinson and U.-D. Reips, "Development of measures of online privacy concern and protection for use on the Internet," *Journal of the American society for information science and technology,* vol. 58, no. 2, pp. 157-165, 2007.

[37] H. J. Smith, S. J. Milberg and S. J. Burke, "Information privacy: measuring individuals' concerns about organizational practices," *MIS quarterly,* pp. 167-196, 1996.

[38] S. D. Gosling, P. J. Rentfrow and W. B. Swann Jr, "A very brief measure of the Big-Five personality domains," *Journal of Research in personality,* vol. 37, p. 504–528, 2003.

[39] S. Strykowski, B. Borusiak, B. Pierański and R. Romanowski, "Student attitudes towards privacy protection on the Internet," *Handel Wewnętrzny,* p. 352–363, 2018.

[40] E. Hargittai and others, "Facebook privacy settings: Who cares?," *First Monday,* 2010.

[41] A. L. Young and A. Quan-Haase, "Privacy protection strategies on Facebook: The Internet privacy paradox revisited," *Information, Communication \& Society,* vol. 16, no. 4, pp. 479-500, 2013.

The Application of Machine Learning in Business Intelligence

V. Naneva* and K. Stefanova**
* Plovdiv University "Paisii Hilendarski", Plovdiv, Bulgaria
** Plovdiv University "Paisii Hilendarski", Plovdiv, Bulgaria
*veselinarnaneva@gmail.com
**kvstefanova@gmail.com

Abstract: As a part of a rapid development of digitalization, every activity in our live produce large amount of information. This structured and unstructured data need to be extracted, processed and analyzed in order to improve our global existing. Additionally, as consumers we want to have not only answers to our questions, but also predictions for services and products, based on our digital behavior. In a collaboration with Artificial Intelligence stays Machine Learning. By the help of self-learning algorithms, it is possible to create a successful forecasting and to predict user's choices. A parallel direction of the Machine Learning is Business Intelligence. The complexity of collecting, extracting, processing and presenting data in an appropriate way define Business Intelligence as a one of the most powerful decision-making techniques in the digital world. In this paper, we will focus on the application of the Machine Learning in Business Intelligence by defining their specifics, tools and principals for usage. It will be given relevant examples of implementation in order to improve concepts understanding.

I. INTRODUCTION

The improvement of today's digitalization becomes one of the most powerful tool to create a successful life. Of course, we could present our daily activities in a traditional way, but in order to be recognizable, the modern person should know what and how to develop stable digital personality. Every activity generates information that could be analyzed and additionally stored for different purposes. The data comes structured and unstructured. To be clear what is the difference we will give definitions about the two types separately. The structure data is categorized as quantitative data. Examples of structured information include personal names, dates, addresses, stock values, geolocation, and more. This type of data is easily understood by machine languages. As a comparison, the definition of unstructured type is related to qualitative data that cannot be processed and analyzed using conventional tools and methods. More than 75% of all data generated today is considered as unstructured and the possible way to manage it is to have it flow in a data lake, for instance, allowing it to be in its raw, unstructured format. If we consider the ordinary daily activity of digital communication as a chat or emailing with someone, we will find different types of information, generated on separate levels. When we talk digitally with each other, we consume not only the connection between us, but also products, service and devices related to it. Basically, this is a close cycle between the consumer and the producer. Of utmost importance for the producer is to store and analyze daily the behavior of the every user and its product in order to offer more and more appropriate solutions for user interest. There are several aspects that need to be taken into account before answering the question "How could we manage to offer the most appropriate service to the specific user?". Based on the era of digitalization, of course, we could give an abstractive answer – by the help of computer algorithms. But how these algorithms are generated and for what purpose they are applied for? In this paper, we will describe what exactly Machine Learning is and how it affects the computerization and specifics of the Business Intelligence. We will suggest several tools and techniques to illustrate better the related application.

II. MACHINE LEARNING

Machine Learning (ML) could be defined as a computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence [1]. Based on collected data, ML could construct and produce algorithms that can learn from and make predictions on it. These predictions are fully related to strictly static program instructions in order to make data-driven forecasting or decisions. In the body of the algorithms stay mathematical optimizations, which deliver methods, theory and application domains to the field. We could specify three main situations in which machines learning is referred to, based on the nature of the human activities or requirements:

- Supervised Learning – by the help of additional information or labels, the algorithms learn how to communicate between the user inputs and the necessity outputs;
- Unsupervised Learning – the algorithms work with own structures and discover hidden patterns in the data;
- Reinforcement Learning – the algorithms interact with a dynamic environment to learn how to predict the next steps.

According to Tom Mitchell, Machine Learning is the study of computer algorithms that allow computer programs to improve automatically through experience [2]. This idea could be also considered as part of an Artificial Intelligence (AI). Generally, Artificial Intelligence are possibilities to build intelligent programs and machines that can creatively solve problems. As in Kibria's paper is focused on, we could defined ML and AI as two compelling tools that are emerging as solutions for managing large amounts of data, especially for making predictions and providing suggestions based on the data sets [3]. By the help of AI, we could structure our data manipulation in order to improve product performance. Additionally, the development of different applications is closely related to statistically collecting and extracting data by the usage of Machine Learning. We could divide several practices that provide structural definition of building algorithms for AI products. The list of them includes:

- Regression – Linear regression could be found as a type of Supervised Machine Learning. In regression, we predict based on approximating a mapping function from input variables to a continuous output ones. An algorithm that is capable of learning a regression predictive model is called a regression algorithm.
- Classification – Common for the classification models is to predict a continuous value as the probability of a given example belonging to each output class. The output variables are often called labels or categories, which is recognizable also for the Supervised Learning.
- Clustering – In clustering, a set of inputs is divided into groups. Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task. In this practice is more often required to find different structures of data by the help of which the algorithms improve themselves.
- Decision Threes – A decision threes is a predictor that creates forecasting, based on label, which is associated in an instance X by traveling from a root node of a tree to a leaf [4]. The simplified solution that gives this practice is relevant to the binary classification setting, but also the application of it could release complex problems.

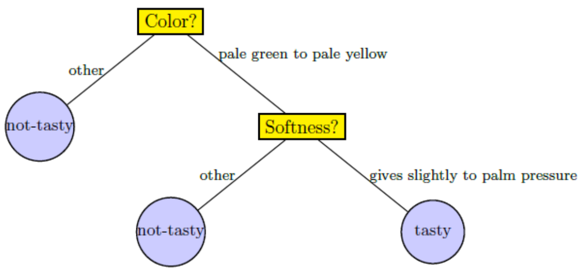


Figure 1. Decision Tree Example

On Fig. 1 we could see the actual presentation of decision trees formed by prediction of papaya's tasty as an example. The nodes define the algorithm and for each of them we have a successor child. Usually, the successor child is chosen on the basis of a splitting of the input space. Then, the splitting is focused on the hierarchy of the tree. We could make a conclusion from the schema of decision tree, that by the help of labels we could make a prediction, which is simple to understand and interpret.

Of course, there are more algorithmic concepts based on Machine Learning. Most of them are concerned with missing values or distance functions and by the help of predefined structures are formed self-learning algorithms. Even there are specified differences in practices or methodologies, all of them are part of the Machine Learning software tools. We will focus on some of the trendy tools related to ML and will compare their application.

The tool that uses highly Automated Machine Learning, classified collection and automated time series analysis is the mathematical environment Wolfram Mathematica. We could consider services related to social media as one of the main application of the Machine Learning. For the aim of optimizing social media usage, ML is applicable to Face Recognition. By the help of scanning the pose and projection in the pictures, software provide the actual suggestion for user, without additional searches. The logic in the back end is a combination of several ML algorithms. Same algorithmic set is found in Traffic Prediction, based on congestion analysis. ML defines areas with more or less traffic and creates periodic prediction to improve user presence on the road. These are several examples of Machine Learning usage in our daily digital life. As might be expected, the implementation of this methodology is applicable in more areas nowadays.

III. BUSINESS INTELLIGENCE

As we have already mentioned in the introduction, we live in a digitalization era and every human activity generates plenty of different information. This data comes in a different format and need to be stored somewhere in a structured and unstructured way. The technique, which is related to extracting, storing, and processing in an appropriate way and finally visualizing a specific information, is Business Intelligence (BI). BI is a part of Decision Support System, by help of which it is possible to be supported the complex decision-making, semi-structured, or non-structured problems [5]. In order to clarify what exactly this concept is, in this paper we will focus on the actual processing of the data in order to provide meaningful analyzation to the end user.

The first part of Business Intelligence is data extracting. For this type of the process, there is a methodology, closely related to Data Mining. It could be described as a logical operation that is used to find a necessary data based on large amount of usually unstructured data. Steps to decide which information is more reliable to others depend on three separately types of data manipulation. As a starting point, we need to use functions and tools for cleaning and transforming the explored data into an appropriate data format.

Additionally in this step is meaningful to determine the nature of the database. Once data is explored and determined, based on the variables in it, it is necessary to choose the patterns which make the best prediction. This is the moment when Machine Learning takes major place. After defining the right patterns, they need to be deployed for the desired outcomes. In the core of Data Mining, there are algorithmic techniques similar to the ML's. In the concept of data extracting stays classification, clustering, and regression, and algorithms used for forecasting are almost similar as in Data Mining. Of course, for the knowledge discovery from databases we need to mention the association rule. In this type of algorithms for finding relevant data among large data sets, it is meaningful to be able to generate directives, based on customer behavior analysis. We could suggest combining steps in a sort of Business Intelligence Architecture (BIA) to have a structural workflow. For this stage of BIA, we could give as an example Operational Systems or External Data. They contain and produce variety of information, and stay as a main segment in the BIA funnel.

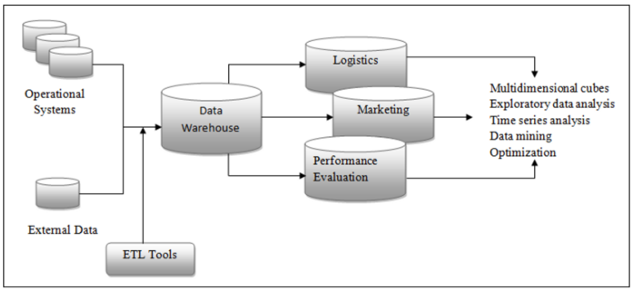


Figure 2. Business Intelligence Architecture

As it could be suggested from Fig. 2, in the process of creating a successful analyzation of the data, after the stage of extracting stays storing. If there have been used the appropriate tools, the data that should be defined in this stage, must be in a structured or a semi-structured type. Nevertheless, there are three different points of data storage architecture. First, we will focus on Data sources. There are plenty of data sources type that different software generated – json files, spreadsheets, external sources, web subscriptions, emails and others. To be possible to store the information correctly, on this stage of storing, sources should be equated partially. After that, as a next level of store architecture stays Data warehouse. This central storage collects and keep data from different sources for improving queries and analyses [6]. Data in a data warehouse is updated regularly, for instance, weekly or sometimes daily, and it is created to support the analytical processing. In this segment, Machine Learning also has impact and there are algorithms for improving data storage performance.

When the data has been stored in an appropriate way, the next level of BI Architecture is data processing. This part of the workflow has several sub stages by the help of which the information will be produced in a correct output. We need to mention a context processing which is related to the ability to process the same data for multiple contexts and then looking for pattern within each result set for further data extracting. Conceptually, here we could apply Machine Learning algorithm to predict which context of use should be applied. The next step after contextualization of data is to standardize data with metadata, master data, and semantic libraries if it needs more manipulation in the Data Warehouses. As we already noted the Data Warehouses could use both structured or semi-structured type. Of an utmost importance is to provide the referential integrity. By the help of this technique, it is possible to use the traditional databases relationships and to manipulate primary keys and foreign keys additionally.

As a last step that could be made in order to provide a successful BI data analyzation is the actual presentation. The distributed stages as an ideology are behind a variety of different analytical application and systems, closely related to reporting. All of these tools provide functionalities such as modelling, prediction, sales analysis and different scenarios. It should be taken into account what the specifics of the data provided are, because it is of utmost importance to decide how a certain information could be presented. Obviously, there are certain rules to follow when presenting a data set. In BI, the conceptual idea for displaying a considered information in order to give a meaningful answer to a specific question is called Visual. There are plenty of visuals, but all of them are divided depending on the application. For example if we want to present data set over a time period, we used linear type of visual. Additionally, if we have structured informational set, based on data hierarchy, we need to use Table visual as matrix table, or regular one. In most BI software products there are features such as visually based predictions that give automotive values and then the tool represents the forecasting by line or a value. The disadvantage is that it is low-manually responsive. Such kind of tools that provide functionalities for representing, manipulating and analyzing information, are closely applicable in the step of data distribution. As an example, we could consider the common BI software products – Power BI and Tableau. Tableau is powerful data visualization tool that helps in simplifying raw data into a very easily understandable format. Its main features include data blending, real time analyses and different type of data collaboration. Data analytics in Tableau could be split into two directions – developer tools and sharing tools. Mostly, ML finds place in the first one. On the other hand, Power BI is a collection of software services, applications, and connectors that work together to turn all unrelated sources of data into coherent, visually immersive, and interactive insights. Power BI provides offline and cloud-based solutions for manipulation and presentation of the data and includes possibilities to use Automotive Machine Learning as a concept, which is implemented in the workflow of the building reports. Despite of that, nowadays there are plenty of data analytical tools, and in all of them Machine Learning takes place not as a helping methodology, but as a main structural unit. In the next section of the paper, we will present some of the common applications of Machine Learning in Business Intelligence.

IV. APPLICATION OF MACHINE LEARNING IN BUSINESS INTELLIGENCE

In the previous sections, we define in details what exactly these two techniques (BI and ML) mean and what their specifics are. If we suggest a result of this combination, it should contains well-processed and meaningfully presented data. For every business life analyzation, it is more than important to know how its activities would change. Analyzation of sales, profit, revenue or distributed changes would be difficult to predict without using the Machine Learning Algorithms. To highlight the advantages of self-learning algorithms in Business Intelligence we will consider several tools in order to define their actual application.

As a starting point, we will focus on the programming language Python. This language is one of the most relevant for scientific computing and provides ecosystems of different libraries. Most of the libraries related to Machine Learning are a collection of algorithms to extract and transform data and additionally to perform data wrangling operation. A good example for this definition are Numpy, Scipy, Pandas, Tensorflow and so on. Machine Learning and scientific computing technologies uses linear algebra operations on multidimensional arrays in order to represent vectors, matrices, and tensors of a higher order [7]. As a detailed example firstly we will describe NumPy. This is a library that provides a multidimensional array object, various derived ones, logical and shaped manipulation. All elements in a NumPy-array should be of the same data type, and thus will be the same size in memory. NumPy fully supports an object-oriented approach, starting, once again, with ndarray. According to Raschka and Patterson's paper, both NumPy and Pandas provide abstractions over a collection of data points with operations that work on the dataset. Additionally, they provide tools for users to execute Machine Learning pipelines end-to-end manually. Tools for automatic Machine Learning (AutoML) are created to correspond with the automation of one or more stages of certain pipelines, making it easier the building ML models while removing a specific repetitive task. This technique is closely related to Business Intelligence and enables to business train, validate and invoke algorithms directly into the software products for BI reports. We will focus on the process of data manipulation with one of the most common tool for data analyzation.

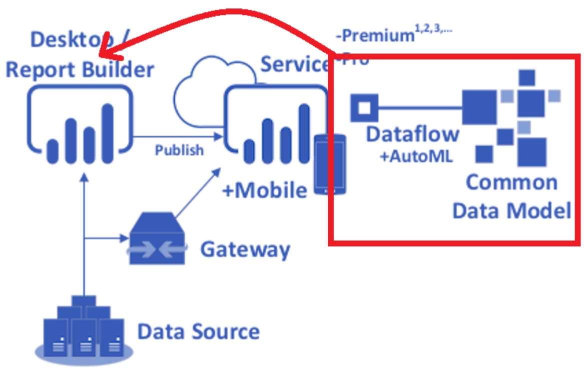


Figure 3. AutoML in Power BI

If we go back to one of the mentioned above BI software, namely Power BI, we could consider the actual application of the AutoML in BI report. As it could be seen on Fig. 3, AutoML interacts with Power BI and provides data manipulation apart from the information included in the specific report. This service automatically extracts the most relevant features and algorithms, and provide ML model validation. There are three main steps:

- Create and Train your model;
- Model Improvement;
- Model Application.

For the first step of the process, it is necessary to define the entity and the outcome field that need to be predicted. Then, there are options for choosing the appropriate ML model/algorithmic type. We need to decide which specific set of self-learning methods we are going to use for the forecasting – binary prediction, general classification or estimate a numeric value based on a regression model. If we want to predict a text field, as we have already mentioned in the definition of regression type, it is not possible for Linear or Logical Regression. As base of Automotive Machine Learning, in Power BI we could make a personal configuration in order to produce accurate values. Additionally to all of the settings that should be made on the specific selected model, it is appropriate to have a ML model training. This is part of dataflow refresh. As a logical process of it, we consider splitting the historical data and testing different datasets. The procedure continues with analyzing fields, reading and manipulating the numeric columns. Then all of the information should be simplified and normalized. If we take the classification model as an example, the AutoML runs the input data through standard sampling and balances the classes to ensure the row counts are equal for all. After the model has been trained, it is necessary to analyze the relationships between the inputs and taking the advantages from the generated report. When the procedure of self-analyzing the performance of the report is finished, the workflow ends with the last step of applying the model. Its implementation creates two new data flows, which contains the predictions and individualized definitions for each row that are stored in the output entity. In addition, there are different type of models, called ensemble ML models, and they are used to obtain better predictive performance. The service of AutoML generates those models if they are found to be optimal for the forecasting.

Additionally to the automotive process of working with Machine Learning models, there is a possibility to generate manually an algorithm suitable to the personal requirements. Such tool is Jupyter Notebook. This technology provides the ability to communicate with wide range of workflows in data science, scientific computing, and machine learning. The programing language of the technology usually is Python. According to Parth Pathak (see [8]), if we want to recreate a Classification model into the Jupyter, we need to include several libraries, such as itertools, numpy, matplotlib.pyplot, pandas and other. The process of manipulating data with this technique starts with pre-processing it, which is related to listing the data sets, converting categorical features to numerical values and normalizing the data. For the accurate results, it is of utmost importance to follow several algorithms as steps. This logical process includes K Nearest Neighbor (KNN), Decision Tree, Support Vector Machine and Logistic Regression. After receiving the required values, the model should be evaluated using test set.

To be possible to integrate Jupyter prediction into the Power BI environment, it is necessary to use an additional workspace. This workspace is a part of a web-based, unified console that provides an alternative to command-line tools. For the service, called Azure, it is possible to converts data processing into analytics and reports that provide real-time insights [9]. As we have already mentioned, the data comes as raw and in unstructured view so by the help of Machine Learning for prediction and manipulation, the information should be stored in an appropriate way. The application of Azure is to provide a dynamically place for ML to communicate with data in order to have refreshed and daily cleaned data. Ordinary, we could make a prediction in Jupyter manually in its interface and recreate in the AutoML modeler, but if we want to have the ability to customize the way the date is predicted, we need to use certain tools for that. As a visually example, we need to present where is the place of Azure in the process of building BI report.

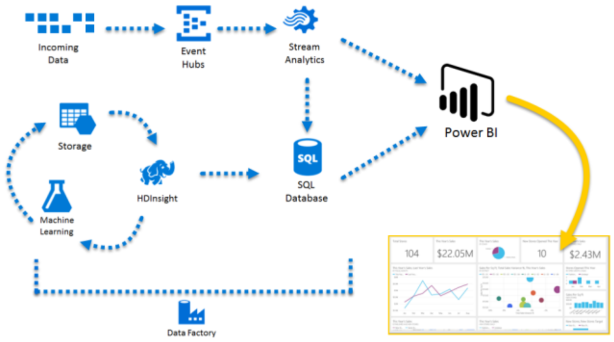


Figure 4. Azure Service in BI

As a conclusion from the Fig. 4, we could suggest that Machine Learning is a non-absent structural unit that is in the composition of the Data Factory in the Azure Environment. Additionally, by the help of Azure Machine Learning studio, it is possible to launch Jyputer and dynamically to include the created prediction. The process of integration first is started with a creation of a Notebook. As a definition, Azure Notebooks are services for developing and running code by using Jupiter. They are based on scenarios to work with and has pre-configured extensions such as JupyterLab, Altair, BQPlot and so on.

Even there are two methods for creating a prediction – automotive and manually, the data that includes forecasting is a part of the information, which is going to be included in the end software product. After creating the relationship in the Power BI, the analyzation report could be made also as a comparison between the data that was before a certain point of period and the values predicted in the future. If we consider an end analyzation report as a presentation of a certain business life and calculate all of the measures related to success of this business, we need definitely to include values in the future as a steps that need to be taken.

V. Conclusion

Nowadays every human activity generates a huge number of information. This data usually is raw and unstructured but is meaningful to be processed in order to to improve our digital life. There are two separate techniques related to data manipulation. On the one hand, by the help of Machine Learning we could create self-learning algorithms to predict user' behavior or choices. On the other hand, to follow data processing rules we need to take into account the specifics of Business Intelligences as a concept. The combination of the two methods for working with different types of information gives the ability to create a structured and architecture-based (BIA) analyzation. In this paper, we suggest the advantages of using not only the data provided for analyzation in reports, but also to take into account Machine Learning features as algorithms and existing tools implementation for an additional information.

Acknowledgment

This paper is partially supported by project MU19-FMI-009 of the Scientific Research Fund of Plovdiv University "Paisii Hilendarski", Bulgaria.

References

[1] Singh, H. and Bikram Pal Singh. "Business Intelligence: Effective machine learning for business administration.", International Journal of IT, Engineering and Applied Sciences Research (IJIEASR) ISSN: 2319-4413, Volume 2, No. 1, January 2013

[2] Thomas M. Mitchell. 1997. Machine Learning (1st. ed.). McGraw-Hill, Inc., USA., ISBN:978-0-07-042807-2

[3] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu and F. Kojima, "Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks," in IEEE Access, vol. 6, pp. 32328-32338, 2018, doi: 10.1109/ACCESS.2018.2837692..

[4] Khan, Rafi & Quadri, Syed. Business Intelligence: An Integrated Approach., Business Intelligence Journal January, 2012 Vol.5, No.1, pp.64-70

[5] In Lih Ong , Pei Hwa Siew and Siew Fan Wong, A Five-Layered Business Intelligence Architecture. IBIMA Publishing,Communications of the IBIM, Vol. 2011 (2011), Article ID 695619, 11 pages, DOI: 10.5171/2011.695619

[6] Raschka, S.; Patterson, J.; Nolet, C. Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. Information 2020, 11, 193.

[7] https://github.com/ParthPathak27/The-best-classifier/blob/master/The%20best%20classifier.ipynb, retrieved on September, 2020.

[8] https://docs.microsoft.com/en-us/power-bi/connect-data/service-azure-and-power-bi, retrieved on September, 2020.

Maritime Single Window and possibility of improving port business

Ana Radulović
Faculty of Maritime Studies, Kotor, University of Montenegro, Montenegro
bukilica@t-com.me

Abstract - This paper briefly points out the importance and application of the port / maritime single window and the possibilities of their improvement in port operations. "Single Window" (SW - Single Window) concept is a formalization of procedures undertaken by the competent services of the European Union through the Center for Trade and Electronic Business in order to achieve efficient exchange of relevant information between trade organizations and government entities. The concept of maritime / port single windows has a basic foundation in the field of formal facilitation of trade and customs clearance in the field of maritime affairs, and is focused on increasing the efficiency of import / export mechanisms and procedures, where detailed information on cargo must be provided in all maritime cross-border activities. The application of single windows was originally focused on an efficient and collaborative form of conducting transactions between commercial and state / administrative entities in cross-border transactions. The paper presents the main advantages of the application of unique windows in the port industry that can be used as a platform for further improvement of the port and maritime economy.

Keywords - SW, MSW, NMSW, VTMIS, FAL Convention, single window

I. Introduction

It is common to assume that the maritime industry is globalized in terms of functional, business and regulatory terms. However, these are also basic elements of the barrier to the development of the maritime industry, given that participants must adhere to a number of complex administrative procedures, including cooperation and work with customs, tax, immigration, security authorities, waste authorities, seafarers' health and the like. [1].

The United Nations Center for Trade Facilitation and E-Business (UN / CEFACT) defines SWS (Single Windows System - Single Window System) as "a service that allows customers involved in trade and transportation to submit standardized data and documents with the aim of fulfilling all formal conditions of import, export and transit of goods and passengers, where information in electronic form is submitted only once in the so - called control point "[2].

The concept of SWS is also explained by the World Custom Organization (WCO) as a trade facilitation tool that allows traders or carriers to meet all the conditions of acceptance of goods in transport in a standardized format only once to the authorities that control it (WCO, 2018). However, the SWS cannot be established in a single step, but in individual steps as defined by UNESCAP / UNECE (United Nations Economic and Social Commission for Asia and the Pacific / UN Economic Commission for Europe - United Nations Economic and Social Commission for Asia and the Pacific) / UN Economic Commission for Europe) during 2012 in the Guidelines for Planning and Implementation through the levels as follows:

- Level 1: Paperless customs clearance;
- Level 2: Regulatory single window;
- Level 3: Unique window through B2B (Business to business) model;
- Level 4: Fully integrated single window;
- Level 5: Cross-border platform for the exchange of single windows (UNESCAP, 2018).

II. MSW (Maritime Single Window)

There are two SW systems, namely customs and maritime / port SW, and they started to be used during 2013. Within the European Union, some countries are already applying the so-called pilot / prototype (first version of the system) NMSW (National Maritime Single Window) in: Bulgaria, Greece, Italy, Malta, Romania and Norway, striving to implement the so-called. national MSW.

MSW (Maritime Single Window) is a concept that aims to digitize and simplify administrative procedures before a ship enters / leaves a port (s), in accordance with the requirements of the European Union Reporting Formalities Directive (RFD) Formalities Directive), or Directive 2010/65 / EU for short, which deals with the standardization of electronic exchange of information and the rationalization of formalities for entering / leaving a ship in / out of port (s). To achieve this, special data entry forms are used, adapted for loading into the MSW, that is, the NMSW system [3].

Thus, the concept of national maritime SWs is an initiative in the field of transport telematics that emerged as a result of several maritime development policies, especially the influence of the European Commission and refers to the application of a national system that will be a single point of electronic delivery and exchange of information. burden between public and private actors of different modes of transport, especially maritime transport and should build on existing telematics applications and systems.

The NMSW model covers the needs of different participants from the local port community, and its architecture integrates and covers communication channels and reduces time and simplifies procedures in port operations (B2P - Business to Port) and operational processes in administration (B2A - Business to Administration) for ship formalities. . Consequently, the business relations of the port affect customers (B2C - Business to Customer), because they benefit from more efficient and simple procedures with a focus on standardizing the exchange of information. NMSW coordination and its functioning differ depending on whether it is observed and analyzed as:

- a system in the center of which is customs clearance, ie. import / export activities;
- a port and ship oriented system with a focus on maritime transport; or
- a system centered on increasing the safety and security of (maritime) transport and traffic [4].

Thus, the system with a single window is a tool that facilitates maritime and other forms of trade by enabling the entities involved in it to transmit all the necessary data to the entities through a single platform, ie. authorities and control bodies that control and monitor them in order to meet all required formalities and requirements [5].

The NMSW platform should be developed in accordance with national rules, European Union legislation and the IMO (International Maritime Organization) in accordance with the Convention on the Facilitation of International Maritime Traffic (FAL). In its original form, the FAL Convention adopted by the IMO deals with the issue of the application of electronic maritime single windows and includes the definition of procedures for all commercial ships navigating international waters.

In this way, Member States can manage different platforms, but all MSWs must meet the following requirements:

- providing electronic reporting formalities by e-transmission through single windows by 1 June 2015;
- develop a one-stop reporting platform that will be available to a number of Member States' authorities and controls;
- receive information in accordance with the reporting formalities provided by EU legislation, which must be available in the national SSN (SafeSeaNet - Safe Maritime Network), and thus to other EU bodies through the SSN system.

Following European and UK regulations on customs, taxes, immigration, security, health and safety, as well as waste management, there are additional, more detailed requirements in the UK than those prescribed by the FAL in the European Union. As NMSW is a new mechanism for providing detailed electronic reports and does not expand the set and content of existing reporting requirements, including an environment for collecting, disseminating and exchanging information, reporting from ships with a structured and commonly defined data structure, rules and access rights management, which are in accordance with the relevant international, national and local legal requirements, the following briefly explains the information flows in the generation of the report, as one of the most important segments, on the example of the British prototype NMSW.

III. Information flows in creating MSW reports

As already pointed out, the purpose of NMSW is, above all, the delivery of electronic instead of paper documents, and users receive automatic confirmation if the sent documents are received correctly, and the same receives multiple entities at the same time. Below is a schematic representation of the information flow within the NMSW system that is applied when generating reports in the UK. Since it is a prototype, detailed instructions for using NMSW for e.g. The UK is likely to undergo changes very soon in the future. Currently, this system requires user registration when accessing by entering company data and user e-mail address, after which it enables and supports the entry of all relevant and requested documents in the form of attachments [6, p.7].

Access to the British NMSW model is protected, and data on the registered user is not passed on to unauthorized parties. The data is not used without the prior consent and consultation of users, eg when a system outage is expected or planned. A schematic representation of sending reports in the direction of NMSW and VTMIS (Vessel Traffic Management Information System), or SSN, is shown in Figure 1.

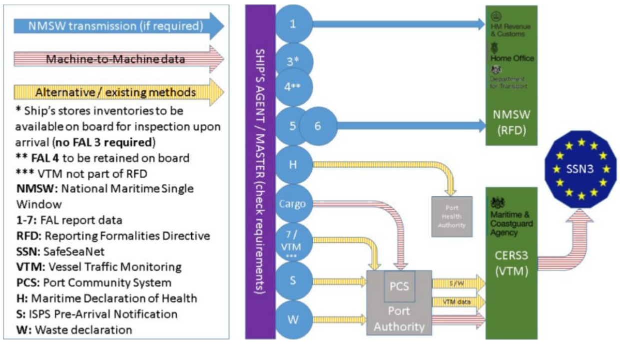


Figure 1. Information flow from the ship to NMSW and CERS (Consolidated European Reporting System) European reporting system) on the example of Great Britain (source: [4, p.11])

VTMIS forms the basis of an electronic platform for the exchange of information at the level of the European Union (as SSN) and includes aspects related to maritime affairs: security, safety, environmental protection, control, ie prevention of marine and coastal pollution, fisheries control, border control, etc. ., in compliance with legal regulations. In addition, VTMIS is largely linked to the physical control of ship traffic. VTMIS aims to increase the safety and efficiency of maritime transport (goods and passengers) and maritime transport (ships) on the basis of regular information, exchange and sharing of information.

The time of receiving the report is contained in the e-mail sent to the user or users after the successful loading of the report into the system. Port authorities require that information be provided on the arrival of a ship in port, in order to provide the necessary services in a timely manner, but this request is currently outside the scope of the NMSW.

In order to make it easier to follow the flow of information, some basic explanations related to NMSW, ie, NSW, VTMIS and SSN [6], are given.

Thus, given that international shipping faces problems in submitting mainly paper documents to national authorities, NMSW is trying to solve the problems of developing and implementing NMSW in creating and exchanging electronic reports and forms. The UK as well as the countries of the European Union have a legal obligation to provide NMSW services in accordance with the RFD.

By providing digital reports and electronic exchange, ships can be sure that their information has been received., The data can be accessed from several different parties, and the same can be used and received by multiple entities. The system is designed to be simple and economical to use, while problems such as the use of inaccurate contact information, faulty or unreliable equipment, unclear handwriting and lack of clarity about reporting are minimized.

The RFD defines the entry of relevant data at least 24 hours before the ship enters UK ports, and if the voyage is shorter than 24 hours, reports should be sent during the ship's departure from the previous port, while if the next port of call changes during the voyage, or is not known in at the time of departure, the report should be sent as soon as the information on the next port becomes certain. FAL reports cover reports related to the elements: FAL 1: General declaration reports; FAL 2: Cargo reports; FAL 3: Ship Trade Reports; FAL 4: Crew Performance Reports; FAL 5: Crew list reports; FAL 6: passenger list reports; FAL 7: Reports on the transport of dangerous goods; and reports related to ship-generated waste as well as reports related to seafarers ’health are included (Figure 1).

With the introduction of national maritime SWs in all EU Member States, based on the flow of information contained in Forms FAL 1-7, PAN (Pre-Arrival Notification) and the Waste Declaration, it is clear that FAL 1 and FAL 5/6 must be entered in the NMSW, while other forms (announcements / declarations) are optional and communicate through the Ports' Community Services (PCS) and port authorities with VTMIS at the national or SSN level at the European Union level. , starting from June 1, 2015.When leaving the port of Great Britain, ships must send an exit check. This report is sent at the time of leaving the port, via the NMSW portal and must contain FAL 1 (Universal Declaration) and FAL 5/6 (Manifesto or crew / punt list) [6, p. 7].If the NMSW is not available for any reason, reports should be sent as soon as the system becomes available. Scheduled outages are posted in advance on the user access page, and are usually scheduled at intervals when traffic is reduced.

In case of prolonged interruption, reports can be sent by fax or e-mail. Reports sent from UK ships to the NMSW are forwarded to: Border Police for Security and Immigration Procedures (FAL 1, 5/6) and - HMRC (Her Majesty's Revenue and Customs) to the Customs Service (FAL 1, 5/6). Confirmation of receipt is sent to the official e-mail address of the user. Users can also provide an alternative e-mail address, to which they will also be sent a confirmation of receipt. All those documents that the commander or agent considers relevant can be loaded into the system as additional.

The forms are simple and contain a header with general information about the ship, they can be copied, ie. used in the same form in each subsequent reporting, thus reducing the volume of administrative work and the number of paper documents on board and on land. NMSW and MCA / CERS (Maritime and Coastguard Agency / Consolidated European Reporting System) are currently separate and serve different but not currently unrelated purposes. NMSW is used for reporting from ships, usually through agents, to the police and customs.

CERS is tied to individual ports and directs VTM (Vessel Traffic Monitoring) reports, such as PAN (Pre-arrival Notification) to the MCA, primarily for security purposes. CERS also has access to PCS. CERS is upgrading and the new version will be able to accept FAL 7, the Waste Declaration and the ISPS PAN (International Ship and Port Facility Security Code) by the ports. MCA works with ports in terms of defining methods for automatic exchange of information [7].

In the future, work will be done on the development of interaction and unification of NMSW and CERS, and any decision on this issue will be forwarded to the users of both systems. For now, in this British prototype there is only an experimental link between NMSW and CERS.

IV. ADVANTAGES OF MSW IN IMPROVING PORT BUSINESS

Notwithstanding a number of implementation problems, a number of key benefits are highlighted below, which MSW provides to the shipping and port industry and onshore administrations involved in maritime transport:

- MSW is a flexible and user-friendly tool for automatically linking relevant information on ships, passengers, crew and / or cargo, related to reporting from the ship to the coast and port and vice versa;
- By exchanging information on the ship, passengers / crew / cargo, among all actors involved in the reporting process, the application of MSW respects the right of access to information;
- The implementation of MSW significantly reduces the time to send reports, which frees the crew from part of the administrative work and provides a better focus on the tasks of direct ship management, increasing the safety of navigation;
- Total reporting costs are reduced and the need to hire intermediaries is eliminated;
- IT complexity is reduced by using simple solutions in the field of maritime transport / traffic, full reporting is provided, ie, monitoring of

ships at the local, regional, (supra) national level (s);

- Compliance with international standards is ensured, such as: ISO 28005, WCO, EDIFACT (Electronic Data Interchange For Administration, Commerce and Transport), including specific EU requirements [8, p.14] etc.
- Maritime and logistical procedures have been simplified;
- Shipping services have been improved in business terms, with a significant reduction in delays, as the most important element in reducing maritime operating costs;
- The need for direct contact is reduced by reducing costs and increasing data transparency;
- Transaction security is increased, and data transfer is adapted to all (included) business systems;
- Paper transactions are eliminated, and the universal single point for entering relevant data avoids duplication of data entry and exchange;
- The system is protected from intrusions, information leaks and viruses;
- The system is scalable with simplified procedures for activating new customer services.

V. Problems in the implementation of MSW

The development and implementation of SW can be analyzed from an international, national, regional or local aspect. All EU members, as well as associate members, are connected (or will soon be) to a central national body that will be the official liaison with the SSN, to which EMSA (European Maritime Safety Agency) is responsible [4].

The most important problems related to the implementation of MSW relate to the issue of ownership, because it is not yet clear whether it represents: public, private or public-private ownership. From the point of view of costs, the question arises whether the services provided by the system are free for users or are paid, that is, who pays for them and how much.

In many cases, the Port SW is identified with the PCS (Port communication System), which is based on integrated procedures, rules, standards and ICT solutions that support automatic data exchange, ie, documents relating to the ship, crew, passengers. and cargo, when entering, staying in port and / or leaving the ship.

However, although the PCS supports the requests of national agencies, and entities interested in cargo, it covers customs and cargo handling requirements, as well as the exchange of information related to services provided in the port to the ship and cargo, emphasizing private and commercial information related to ordering. and charging for port services, rather than just tracking the ship.

On the other hand, EPC (Electronic Port Clearance) is a concept that is more related to ships and their electronic management of formalities related to documentation and procedures at entry / departure, ie, during the ship's stay in port [4].

Thus, there is a need to combine these two (PCS and EPC) very complex and mostly, at least so far incompatible systems that can be combined with a single MSW.

VI. Concluding remarks

Initiatives to establish a European environment with a single window in port and maritime transport solve the problems of the current inefficiency of the environment in reporting in port transport for maritime transport operators, which arose from some limitations of the RFD directive of the European Union. However, the currently valid directives do not provide precisely defined guidelines or define obligations that can guarantee the development of a common and integrated model and data exchange flows in maritime transport. Therefore, the main challenge is to harmonize procedures that may be imposed by national regulations, given that inconsistent reporting at EU level is a lengthy process that simultaneously burdens both shipping, ports, seafarers and maritime and transport companies operating in the international market. However, the most important thing is that the whole port community recognizes the benefits of adopting the MSW model, because only in this way can the entire supply chain through ports and the maritime economy be an efficient and simple model of communication and information exchange tailored to users.

References

[1] Koliousis, I., Koliousis, P., Koliousis, T., Maritime Single Windows: Lessons Learned From The Emar Project, Maritime-Port Technology and Development, Chapter: 4, Publisher: CRC Press, 2015, pp. 27-34.

[2] UN / CEFACT, Single Window Repository, and World Bank's Trading Across Borders, 2011, http://www.unece.org/cefact/single_window/welcome.htm (available: 21.01.2020).

[3] Department for Transport, UK National Maritime Single Window (Pilot) - Moving Britain Ahead. www.gov.uk/government/uploads/system/uploads/attachment_data/file/494258/national-maritime-single-window-guidelines.pdf (available: 26.12.2019).

[4] Fjortoft KE, Hagaseth M., Maritime Single Window: Issues and Prospects, TransNav - International Journal on Marine Navigation and Safety of Sea Transportation, Vol. 5, no. 3, September 2011, pp. 401-406.

[5] Ahn, B.S., Han, M.C., A Comparative Study on the Single Window between Korea and Singapore, 2007.

[6] HM Government, National Maritime Single Window Pilot - Pilot NMSW User Guidance. https://maritime.dft.gov.uk/supportmaritime.olg?opt=guidance (available: 02.12.2019).

[7] IMO, FAL Forms and Certificates http://www.imo.org/en/OurWork/Facilitation/FormsCertificates/Pages/Default.aspx (accessed: 26.04.2019).

[8] Chanda Y., Johnson S., eMar delivering Intelligent Maritime Single Window solutions, INTELCOM, May 2014, pp.1-14.

Technical and User Evaluation of Babbage Cabbage

An Empathetic Biological Media

Adrian David Cheok*, Kasun Karunanayaka+, Chamari Edirisinghe#,
* i-University Tokyo, Japan
+ University of Colombo School of Computing, Sri Lanka
# University of Sri Jayawaradanapura, Sri Lanka
{adrian, kasun, chamari }@imagineeringinstitute.org

Abstract - Babbage Cabbage is a new form of empathetic biological media used to communicate environmental information in the form of a living slow media feedback display. At present, people are generally too busy to monitor various significant environmental issues, such as air pollution, wildlife extinction, and rate of deforestation. By quantifying such information digitally and projecting to the user through living plants, we aim on providing an interface that connects the user to related issues in a manner that traditional digital media cannot. Our system, Babbage cabbage, aims to promote empathy towards the information by coupling important environmental information and the output media by relating these issues to the color changing properties of living red cabbage (Brassica Oleracea). Furthermore, a formal user study was conducted which suggests that the living media promotes more empathy within the user towards the issue than conventional non-living media.

Keywords: ambient media, living media, biological information.

I. INTRODUCTION

In this paper, we describe a system named 'Babbage Cabbage' created using a living red cabbage. We explain how this system provides an ambient feedback medium that promotes empathetic responses within users towards environmental issues. Current environmental concerns encompass a broad range of issues that threaten Earth and its organism. Information related to environmental conditions is readily and freely available in all forms of media. However, due to busy lifestyle and information overload, most people do not process information into a meaningful context and nor do they feel the need to act on information they receive. The Babbage Cabbage system, shown in Figure 1, is intending to provide the user with an empathetic connection to important environmental issues, by using the colour changing the property of living red cabbage.

The novelty of this study can be described in two manners. First, the use of ambient living media to inform and elicit human emotions. Our argument is that, just as a human reacts differently when receiving or appreciating a living flower instead of an artificial one, so too would a living media promote more human interest and empathy about organic based information. Living things evoke feelings related to empathy and recognition of a common struggle in the world different from what can be evoked by artificial or digital imitations. This is based on the fact that it has been shown that humans have true empathy for creatures, which are alive [14,9]. Secondly, the use of a living organism to represent the changes in the environment gives a real meaning to the message. The decay of the living organism symbolizes the environmental destructions, thus evoking an empathetic response. By combining digital media with the living organism, we can develop media that evokes emotions including empathy and care ecological information.

Figure 1. Babbage Cabbage System

In addition, this system can be adapted to other forms of plant-based media. Aside from colour change, we are adapting our communication and control system to involve plants that move as a display. Another possibility is the use of plants as a natural input system such as a leaf or a plant as a substitute for the push button in an elevator. A ubiquitous system using our plant-based media has the ability to raise awareness and motivate users to create and experience plant-based systems. Our system Babbage Cabbage, an empathetic ambient living media system, operates by obtaining information about various environmental issues, for instance, the level of air pollution from various sources (as the input), and coupling them to the colour changing properties of a living red cabbage. Red cabbage changes color according to the pH level of the solution it absorbs (Figure 9, Figure 7). Therefore, the system maps received information to the pH level of the solution delivered to the cabbage. For example, if air pollution level has increased (input), the cabbage would turn red (output) and the colour change is able to convey to the user a message of environmental health state.

Babbage Cabbage system uses a cabbage as the living media due to two reasons. First, due to its controllable and reversible properties (Figure 11, Figure 12) of colour change. This provides the ability to present a range of changing data through the cabbage as an analogue display. Second, the use of a cabbage as a living plant to represent environmental changes creates an 'impedance match between the living media and the environmental issues allowing the user to readily empathize with the system. This impedance match adds semantics to the manifestation since as environmental health flourish and worse, so would the live red cabbage.

Each red cabbage in the system represents one major issue. One cabbage turning red would mean negative changes in air pollution in the area where the data is being collected. Another cabbage might represent the noise pollution data and transform its colour according to the changes in data. A cabbage gradually returning to its natural colour would signify positive changes. This system can be used as a living organism in a smart home to demonstrate consumption data or it can be part of IoT network. We hypothesize that by seeing the cabbage, the impedance match brought forth by this connection between the living media and the environmental information promotes more empathy within the user towards the environmental issues. We have carried out a study to support this claim. Our aims were to identify the most suitable and efficient information to represent through a living media display with slow media feedback property and how empathy is conveyed towards issues. For this, we formulated two questions.

– What kind of information would they prefer to be presented through such a system?

Based on this study, we constructed two similar information display systems, one with the living red cabbage (Babbage Cabbage system) and another with an artificial colour changing replica of a red cabbage (Figure 13). Then we conducted a user study with 22 participants to obtain the answer for the following question.

– Which display promoted more empathy towards the issue?

With this study, it is proved that there is a significant difference between the empathy promoted with living display than the digital artificial representation of the same media towards the represented issue. The only dissimilarity between these two displays was the living factor of the media which caused the significant difference between the promoted empathy.

A poster paper of this system published earlier in the Laval Virtual Conference [6]. There we have discussed the concept and the basic functionality of the system. This paper provides a detailed technical implementation overview of the system together with technical and user evaluations of the Babbage cabbage system.

The next section, related works, describes similar research by other researchers on both living and non-living media and why we based our research on living plants.

Next, the system description follows explaining the system design, and the experiments conducted on the system. The user study details the method and results. The Conclusion of this paper summarizes our finding and the end results of our research.

II. Related Works

There are a few important areas of research that are related to our living media system and have provided important background elements. Our research is concerned with using living ambient media to relate environmental issues to the user. For this, we present. related examples of living organisms empowering interactive media, acting in the broad definition, as an extension of man [12].

A. Ambient Media

Ambient media [8] is fast becoming a familiar keyword in everyday society as researchers are exploring new ways to utilize the spare moments between the shift of focus of various user tasks or utilizing opportunities of transferring information during activities that require only partial attention from the user. Ideally, this media helps people receive, translate and perceive information in the periphery without letting it interfere with their foreground activities. Researchers from different disciplines are developing methods to provide the information to users in more effective ways by appealing to available sensory channels and finding ways for users to process the conveyed information in order to maintain appropriate cognitive loads.

Ishii et al. developed “ambientROOM” [9], a room that had a range of information displays integrated into its architectural structure. The water ripples, active wallpaper, and ambient sound are used in the AmbientRoom as ambient media. This creates an interface between people and virtual worlds, engaging all of the human senses, and blending into the surroundings. Other examples include, LumiTouch [4], a device which is an augmented photo-frame artifact which allows a remote couple to feel each other's presence and abstract feelings for each other, and Pocomz [13], a communicative ambient display which shows the status of online chat pals, displayed in a non-intrusive manner through ambient lighting are a few of several early ambient display technologies.

In above mentioned examples, which allow the user to interact, are based on electromechanical non-living objects. Our prototype uses a non-intrusive biological organism (cabbage) which has a deeper connection to the state of the environment. There exists a connection between the ecosystem and another living object, the red cabbage which is then perceived by the user, who is also alive, and thereby make a connection because of an understanding, perhaps unconscious, that we are on a common struggle in the cycle of life.

B. *Living Media*

Living media encompasses both animals, either DNA altered or non-DNA altered and plants. Plants can act as living media with the help of electronic displays or by depending on their inherent natural features.

Humans and animals have been closely linked throughout history in relationships which have yielded improvements to the quality of life. Animals have served to warn about environmental threats, including water and air pollution for hundreds of years. For instance, in the early 1800s coal miners were accompanied by canaries that were very sensitive to the amount of methane gas and carbon monoxide in the air and therefore acted as an early warning system. At the San Francisco Public Utilities Commission’s water treatment plant in Millbrae, a water security system [18] analyzes the behavior of a number of Bluegill fish kept in a tank to detect water pollution. At the Le Centre International de I’Eau de Nancy, France (known as NANCIE) the biodetector used in the detection of water pollution is apteronotus albifrons, an electric tropical fish [14].

Genetically modified animals, such as the Glowing Green Pigs [17] and Transgenic Zebrafish [19], all of which have been bred to glow under ultraviolet and blue light, have also assisted mankind. The most significant difference in these animals and our system is that their display does not change according to data and nor is it reversible. Our system, in contrast does not depend on genetic modifications. Instead, we rely on the natural, inherent features of the red cabbages to produce an ambient display.

There are several other projects that use living plants as their medium such as Infotropism [7] which combines living plants with sensors and lights in an interactive display by creating a robotic analogue that mimics phototropic behaviour, and Spore [5], a project which uses plants as an actuator so that the plant acts as a self-sustaining ecosystem to visualize flow of stock prices. PLANTS DEMO [1] is a system that allows plants to control their own environments by detecting and analysing plant signals and activating appropriate reactions. Another example of nature-human-computer interaction is the I/O Plant [11], a system that utilizes plants as an input output interface. In these systems the plants’ natural features are not used as a display. Instead, the information is displayed across electronic interfaces and LED displays. Our system, in contrast semantically matches the natural properties of red cabbage to environmental crisis, without the use of any artificial means to convey our message to the user.

III. System Description

In this section we discuss our system in detail with the functionality of mechatronic controller and PC interface. Furthermore, we analyze the results obtained by the experiments such as linearity of the mechatronic controller, color changing results, color reversibility, and color controllability.

The block diagram of the overall system is as depicted in Figure 2. The system extracts information from the assigned environmental issue as the input and produce output as a continuous color change of the cabbages. Living red cabbages contain a type of pigment, called anthocyanins. These pigments, responsible for the colors of many fruits and vegetables [16] are sensitive to pH variations such that their color differs at various pH levels. In certain pH levels these pigments may appear in red, purple, or blue.

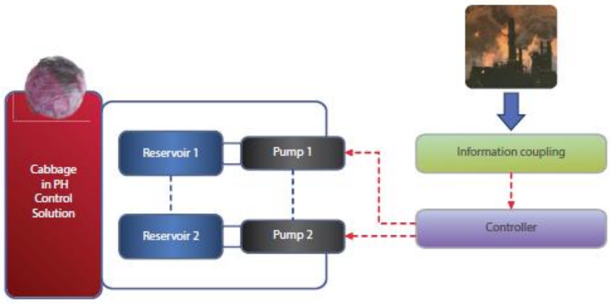


Figure 2. System Architecture

To use this property for living media, Babbage Cabbage employs a mechatronic control system that accurately controls the desired level of pH of the solution that the cabbage absorbs. However, the color change of the cabbage is slow; requiring about half an hour to one hour once the pH is changed.

A computer attached to the system which acts as the interface, allows the user to assign information she wishes to represent in each individual cabbage of the system. Based on the parameters set by the user, and the information received, the computer transmits various set points to the mechatronic control system which in turn controls the pH of the solutions. Currently, the live cabbage is placed in a special glass container that is capable of delivering the pH altered solutions effectively to the cabbage using a sprinkler. This system is easily expandable to connect more than one cabbage to the system.

Figure 3. Setup of one cabbage pixel of Babbage Cabbage system

A. *The Mechatronic Controller*

The optimized mechatronic control system mixes solutions from the reservoirs of pH 2, 7, 10, and 13 solutions to produce the desired solution. The mechatronic controller utilizes a PI (Proportional Integral) control algorithm implemented on a microcontroller in order to mix the pH solutions. Using PWM (pulse width modulation) control, it actuates four pumps to pump different pH solutions from the containers while receiving the 10-bit pH value from the pH sensor in the mixed solution. The resulting solution is then circulated using two pumps (per cabbage) to the cabbage(s). This mixed solution will sprinkle on to the cabbage until system senses a change in the status of the monitoring process. Should the status change, the mechatronic controller takes appropriate solutions from the containers again and mixes with the current solution until it reaches to new pH value. The resulting system for one cabbage is shown in Figure 3.

B. The Interface

The interface lets the user select and assign the status of a slow changing variable that she feels is important to the cabbages. After assigning each cabbage with individual information, if required, it also allows the user to map the cabbage color to the level of change in the assigned information. Upon these selections, the system monitors these various issues and transmits the appropriate set points to the mechatronic controller, which in turn changes the color of each cabbage.

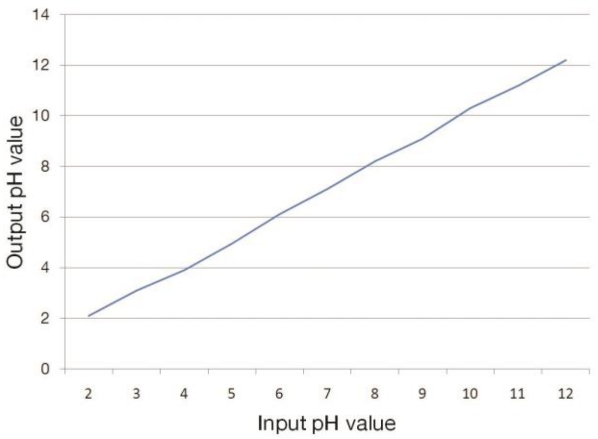


Figure 4. Linearity of the mechatronic controller

IV. RESULTS OF THE SYSTEM

A. Results of the mechatronic controller

Figure 4 indicates the linearity of the system. Inferring from this, the mechatronic control system is able to produce accurate pH solutions with an error of less than 2%. This allows the system to accurately control the color levels of the cabbage increasing the range of information it can represent. In addition, the current system is able to produce the desired pH in 8-10 seconds as depicted in Figure 5. This is acceptable as the cabbage would take approximately one hour to change the color

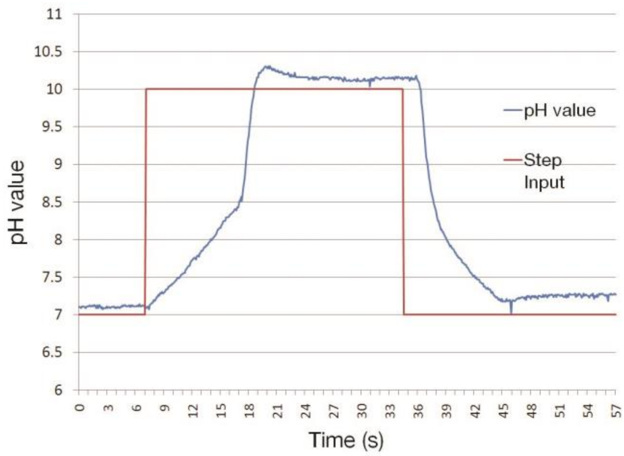


Figure 5. Transient response of the mechatronic controller

B. Color Changing Results

In addition, the system was tested for color changing results of the cabbage at various pH levels. For a quantitative analysis of the color change, we used a KEYENCE(R) CZ-H32 color sensor with the CZ-V21A amplifier. This sensor unit displays the degree of correspondence between the target color calibrated as a reference and the target color currently being detected. The value read as a result of the reflected light intensity affected by the target color. We read the color of five constant points of the cabbage and averaged the value. For this purpose, we calibrated the reference color to be the original color of the red cabbage. For readability, the values are displayed in a normalized form where 0 is the original color of the red cabbage and 1 is the final color of the cabbage. The setup of the system is shown in Figure 6.

The studies were carried out in two steps:

– Color changing results of a cabbage leaf

– Color changing results of the whole cabbage

Color changing results of a cabbage leaf: To see the color change of the red cabbage, first a small-scale qualitative study was carried out in which a leaf of the red cabbage was submerged in different pH solutions to observe the color changes. Color of the cabbage leaves changes its color gradually with the time. It takes about a half an hour to change the color of the whole cabbage leaf. The results are as in Figure 7. In addition, there is a clear color change in the solutions as the pigments of the leaf of the cabbage were dissolved. This study additionally indicated the range of colors that the cabbage can change to.

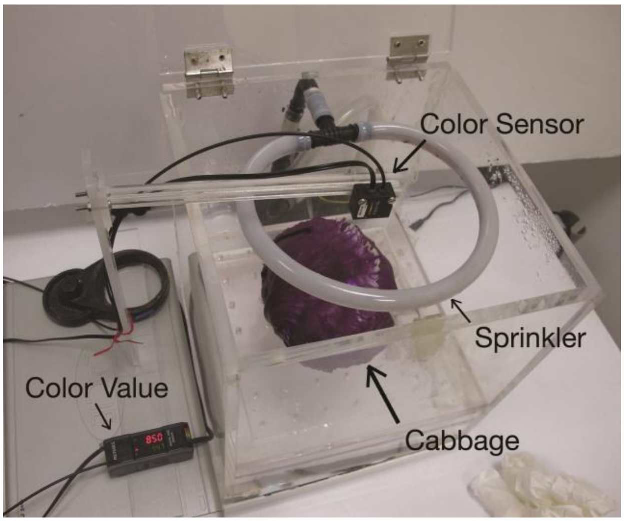


Figure 6. Fig.6: Setup of the system for testing color change results

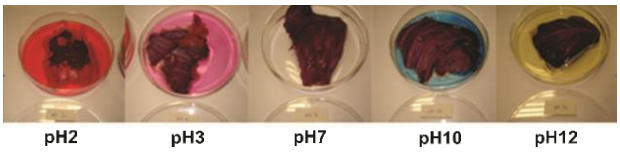


Figure 7. Color changing results for cabbage submerged in different pH solutions The cabbage was initially given pH 2 and the resulting color change was observed through the color sensor.

Color changing results of the whole cabbage: In this study, the full cabbage was examined. Using the mechatronic control system, the full cabbage was delivered different pH level solutions to trigger the color change. The cabbage was initially given pH 2 and the resulting color change was observed through the color sensor. The resulting transient curves of the color is depicted in Figure 8. The rise time of approximately 55 minutes is acceptable for our purpose of the living media as a slow feedback display. The resulting color changes of the cabbage after two hours are indicated in Figure 9. In the current results, even though the cabbage does not change the color as a whole, there is a significant color change in the cabbage.

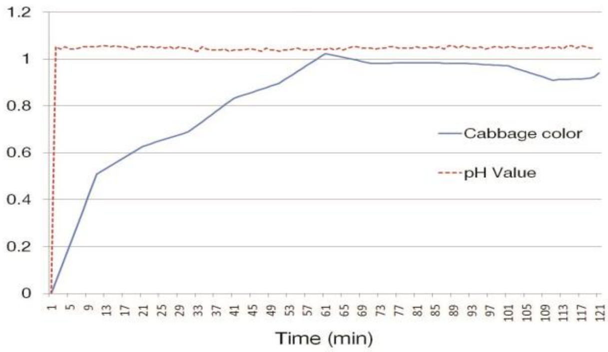


Figure 8. Color transient response for cabbage submerged in different pH solutions

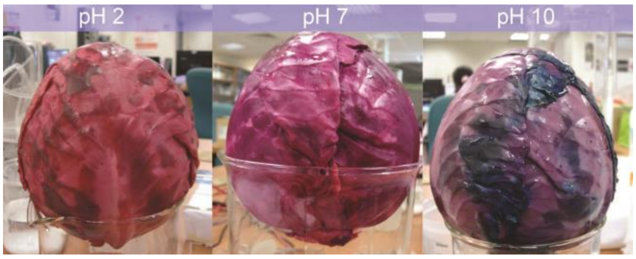


Figure 9. Color changing results of the full cabbage

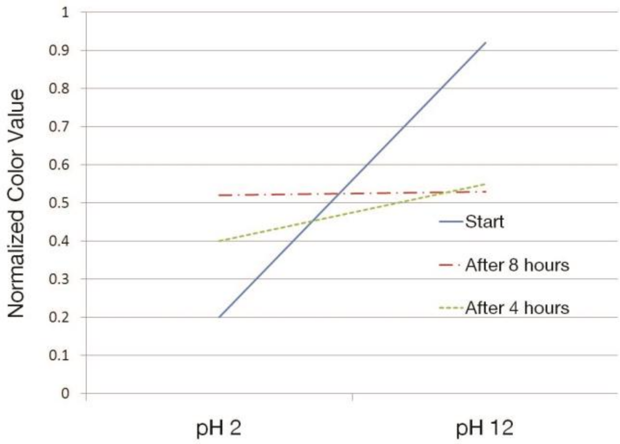


Figure 10. Reversibility test of the cabbage relationship between the colors of cabbage

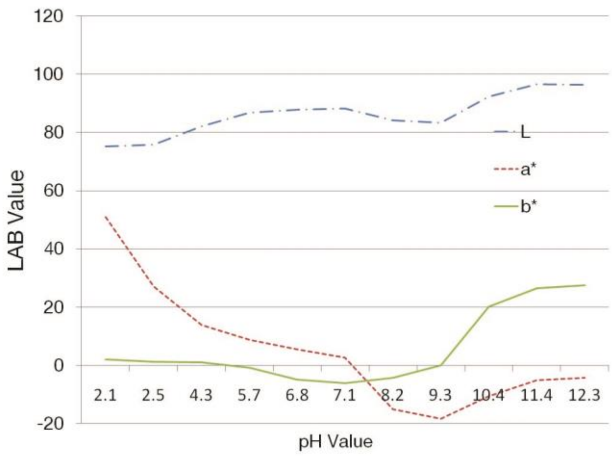


Figure 11. Color vs pH solution mapping

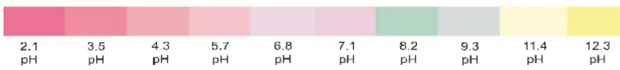


Figure 12. Color spectrum mapped against pH level

Figure 13.

Reversibility study of the cabbage color change: This study was conducted to observe one of the most important properties of the living media, the reversibility. To conduct the study the cabbage was initially exposed to pH 2 solution, then progressed to pH 12 solution with steps of pH 1 every hour. This process was reversed and the sequence of cabbage colors were observed through the color sensor (Figure 10). In this Figure 10 the value 0 indicates the color of the cabbage at pH 2 and 1 indicates the color of the cabbage at pH 12. As observed the hysteresis of the cabbage declined over a period of 8 hours rendering to use a new cabbage afterwards. However, this was due to the relatively fast change of the pH solution continuously. However, in the case of the actual usage of the system, the frequency of data (such as the air quality) is much lower (example: once a day) thus lengthening the life of the cabbage.

Color Controllability: Following this, an analysis was carried out to map the pH level versus the RGB (Red, Green, and Blue) color measurement. This study primarily focused on establishing a relationship between the colors of cabbage and the pH solutions to accurately control the color. To conduct this study, 11 samples of blended red cabbage juice were mixed with different solutions with known pH levels. The resulting solutions were then scanned using SHIMADZU scanning spectrophotometer (uV-3101PC UV VIS-NIR) and the transmittance curve of each of the 11 samples were recorded. For each sample, the tristimulus values of the color is calculated by multiplying the transmittance curve of the sample with the spectral power distribution of the CIE Standard illuminant Source and color matching function D65 of 1931 CIE standard observer [2]. The tristimulus values are then converted to CIELAB values. Graphs in Figure 11 indicate the mapping between the L, a* and b* values against the pH level. The color spectrum derived from these graphs are indicated in Figure 12.

V. User Studies

The main objective of the user study was to explore the different empathetic responses of participants when information was represented through living or non-living media. Hence to explore if there exists a significant difference between the empathy promoted within the user when using a living media as opposed to a conventional digital artificial representation of the same media. We have conducted an initial study to identify the issue to represent through the living media and the main study to address the following hypothesis, both made up of a mix of closed and open-ended questions:

- Living media promotes more empathy within the user towards the issue than conventional non-living media

Participants: For the initial study, 37 participants (undergraduate students, graduate students, and research fellows; age 22-32; M = 25; SD = 2.59) (M = Mean, SD = Standard Deviation) were recruited. There were 23 males and 14 females participated for the study. Similarly, for the main experiment, 22 participants (undergraduate students, graduate students, and research fellows; age 22-35; M = 25.5; SD = 3.35) were recruited. There were 14 males and 8 females participated for this study from same nationalities as the first study. All the participants were from different nationalities, including Singapore, Malaysia, China, Sri Lanka, Japan, Canada, and USA.

Initial Study: Initial Study Prior to the main study, first experiment was conducted to observe the most suitable environmental issue to display through living media. This experiment lasted approximately 15 minutes per person. This was a preliminary study which was carried out two weeks prior to the main study. The main focus of this study was to reveal a possible environmental issue to couple with the information display thus to find an interesting and most suitable issue. Participants were asked to complete a questionnaire containing 20 issues to determine their concerns on different environmental issues, including water pollution, air pollution, and the like. The issues were designed to match the slow feedback nature of the living media.

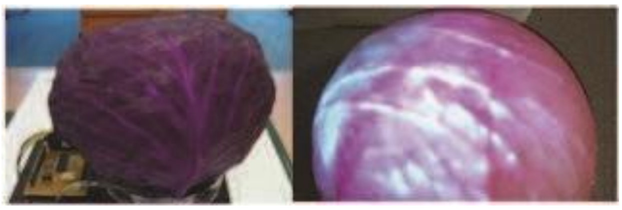


Figure 14. Fig.13: The Babbage Cabbage system with the Artificial Cabbage system setup for the main experiment

Furthermore, the participants were requested to answer these issues based on three factors; importance, accuracy, and frequency of access. This describes how important to know the information related to the issue, how accurately should be the display, and how often the information is needed. All three items were measured with a 5-point Likert scale, ranging from ‘not important at all’ to ‘very important’, ‘not accurate at all’ to ‘very accurate’, and ‘never’ to ‘almost all the time’. In addition, the open-ended questions contained several questions to obtain their general understanding on living media displays. Based on the results from this study the system was setup to display thechosen issue, ‘the level of air pollution’.

Main Study: The main experiment was designed, as illustrated in Figure 13, to evaluate that the living media promotes more empathy within the user towards the issue than conventional non-living media. The artificial color changing replica of the red cabbage (Artificial Cabbage system) was designed to be identical with the Babbage Cabbage (real cabbage) system. Furthermore, the artificial cabbage also changes its color based on the states of the issue as with the Babbage Cabbage system.

This experiment was conducted after two weeks from the first experiment and lasted approximately 20 minutes per session. The questionnaire was designed based on the ‘Index of empathy for children and Adolescents’ [3]. The answers were measured with a 5-point Likert scale, ranging from ‘strongly disagree’ to ‘strongly agree’ while the open-ended questions provided more information on user experience and how will be the final system looks like. For fair and neutral result we have divided the participants into two groups, A and B with 11 participants each. During the first session, group A was introduced to the Babbage Cabbage system while group B was introduced to the Artificial Cabbage system. Similarly, during the second session group A was introduced to the Artificial Cabbage system while group B was introduced to the Babbage Cabbage system.

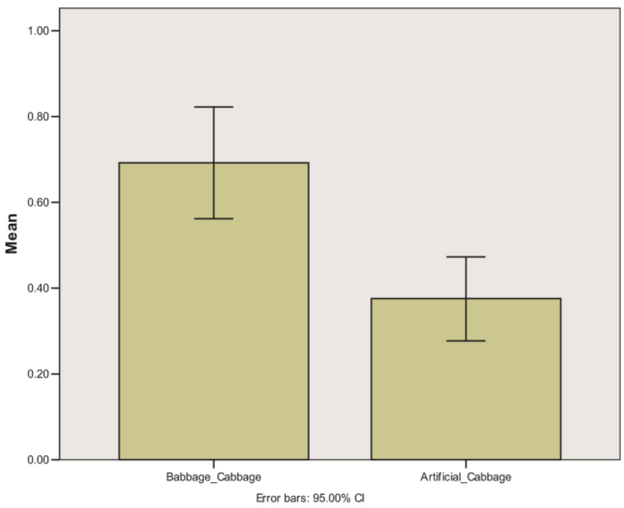


Figure 15. Fig.14: Results of the user study with 95% Confidence Interval (95% CI)

A. *Results*

This section presents a detailed analysis of our results observed through the user studies in order to answer our hypothesis presented. All the values presented in the results section are rounded for two decimal points.

Initial Study: Since we need to observe appropriate information for this living media system (information on high important, less accurate, and less frequency of access), we have revised and standardized the results accordingly. Furthermore, it was assumed that these three parameters are equally important for the final selection.

The mean score for importance, accuracy, and frequency was calculated as a weighted average of the ratings given for the questions, and ranged from 0 to 1 where 1 corresponded to a high important, less accurate, and less frequent information with regards to the issue. Hence the most relevant issue was selected as ‘the level of air pollution’ (Means after ranged and standardized the results for importance, accuracy, and frequency are respectively: 0.82, 0.66, and 0.7 and the standard deviations are respectively 0.24, 0.24, and 0.16)

Main Study: To answer our hypothesis, the main user study was conducted and the results were illustrated by Figure 14 for living media (Babbage cabbage) and non-living media (Artificial cabbage) by analyzing means (calculated as a weighted average of the ratings given for the questions) and standard deviations ranged from 0 to 1 where 1 corresponded to an strongly agree and 0 corresponded to an strongly disagree.

We have analyzed the collected responses in order to prove the hypothesis presented at the beginning of the paper as depicted in Figure 14. Each participant’s score was considered individually. The data was analyzed using paired t-test method. Paired t-test proved that there is a significant difference between empathy generated within the user towards the issue when they are using the Babbage Cabbage system (M = 0.69, SD = 0.18) than Artificial Cabbage system (M = 0.37, SD = 0.13), t(9) = 5.154, p = 0.001. In addition, those results rejected the null hypothesis of no association between the two variables.

From the results, it proves that there is a significant difference on the empathy generated towards the issue when using the two systems, Babbage Cabbage and Artificial Cabbage. Most of the users reported that they feel upset more when they observe the information that indicates worsening of the issue through Babbage Cabbage system. Furthermore, in terms of sadness generated within them when the systems indicate the worsening of the issue also has a significant high mean relate to the Babbage cabbage system (Babbage cabbage: M = 0.67, SD = 0.21 and Artificial Cabbage: M = 0.48, SD = 0.22). There is no significant difference of the statement made with regards to “more than the condition of the system, the condition of the issue that it represents is more important” (Babbage cabbage: M = 0.69, SD = 0.25 and Artificial Cabbage: M = 0.64, SD = 0.22). It is revealed from the study results that the participants were felt they should do something about the issue when they interact with the living cabbage (Babbage cabbage) than the non-living cabbage (artificial cabbage) (Babbage cabbage: M = 0.70, SD = 0.19 and Artificial Cabbage: M = 0.52, SD = 0.23). In addition to those plus points, very few participants believe that using this kind of media to display environmental information is a ‘silly effort’ (Babbage Cabbage: M = 0.18, SD = 0.20 and Artificial Cabbage: M = 0.21, SD = 0.24).

In the experiments, the hypothesis “Living media promotes more empathy within the user towards the issue than conventional non-living media” was tested and evaluated. Results from the first questionnaire helped to obtain a relevant information source (air pollution) to represent through a living media. Based on the main study, we concluded that there exists a significant difference between using living cabbage than an artificial cabbage for the environmental information representation to promote more empathy within the user towards the issue.

VI. Future Work

Our research presented a novel medium of information representation through ambient living media and we demonstrated how such media would promote more empathy towards the information represented. Without limiting to environmental issues, this concept is applicable to a wide range of scenarios including representing personal information such as personal health, time spent with family, etc. In addition, the unique combination of ambient media with natural living media opens up limitless possibilities for such systems to be ubiquitously implemented opening up personal and private information channels in common public spaces in a more calming and subtle manner.

It is expected that this work of living media would trigger novel areas of research specializing in interactive living media. The idea of converting naturally occurring phenomena of living beings into a controllable form allows researchers to push the complex technologies to the background and present more usable, adoptable forms of media in the future. Additionally, research into such closely impedance matched information media, could be furthered into the possibility of promoting action on empathy. However, it should be noted that such research may raise many ethical concerns, and thus should be conducted under proper ethical guidelines and in collaboration and consultation with experts, philosophers, psychologists, and the general public throughout the studies. We hope that this exploration into novel use of living organisms as living media would challenge technological boundaries and bring forth new waves of innovation.

VII. Conclusion

Babbage Cabbage provides the user with information on important environmental issues, based on a reversible ambient display made up of living red cabbages. In the Babbage Cabbage system, an ‘impedance match’ is made between the information about major environmental concerns such as air pollution and the color changing property of living red cabbage. Control of this display was achieved with a system that monitored the environmental information and modified the flow of solution to the cabbage accordingly.

The Babbage Cabbage system can be set in any arrangement that will merge with other houseplants or a floral garden with the minimum of fuss. The use of living plants to convey such environmental information maps effectively the semantics of the subject matter. It helps form an empathic relationship with the news conveyed and helps the users come to a better understanding of their actions.

Based on the formal user study with 22 participants, it was found that living media promotes more empathy within the user towards the issue than conventional non-living media. In addition, many participants reported that the Babbage Cabbage system (living media display) was very effective and they felt that they were attached to the issue through the cabbage display.

References

[1] Barton, J., OFlynn, B., Aherne, K., Morrissey, A., OSullivan, J., Cassells, A., Drossos, N., Goumopoulos, C., Tooke, F., Whitbread-Abrutat, P.: PLANTS DEMO-Enabling Mixed Societies of Communicating Plants and Artefacts. na (2006)

[2] Berns, R.S., et al.: Billmeyer and Saltzman’s principles of color technology. Wiley New York (2000)

[3] Bryant, B.K.: An index of empathy for children and adolescents. Child development pp.413–425 (1982)

[4] Chang, A., Resner, B., Koerner, B., Wang, X., Ishii, H.: Lumitouch: an emotional communication device. In: CHI’01 extended abstracts on Human factors in computing systems. pp.313–314. ACM (2001)

[5] Easterly, D.: Bio-fi: inverse biotelemetry projects. In: Proceedings of the 12th annual ACM international conference on Multimedia. pp. 182–183. ACM (2004)

[6] Fernando, O.N., Cheok, A.D., Merritt, T., Peiris, R.L., Fernando, C.L., Ranasinghe, N. and Karunanayaka, K., 2009. Babbage cabbage: Biological empathetic media. VRIC Laval Virtual Proceedings, pp.363-366.

[7] Holstius, D., Kembel, J., Hurst, A., Wan, P.H., Forlizzi, J.: Infotropism: living and robotic plants as interactive displays. In: Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques. pp. 215–221. ACM (2004)

[8] Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of the ACM SIGCHI Conference on Human factors in computing systems. pp. 234–241. ACM (1997)

[9] Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B., Yarin, P.: ambientroom: integrating ambient media with architectural space. In: CHI 98 conference summary on Human factors in computing systems. pp. 173–174. ACM (1998)

[10] Killen, M., Smetana, J.G.: Handbook of moral development. Psychology Press (2013)

[11] Kuribayashi, S., Wakita, A.: Growable media design: Integrating plants and digital media for information visualization. In: Ubicomp. vol. 5, p. 79 (2005)

[12] McLuhan, M., MCLUHAN, M.A., Lapham, L.H.: Understanding media: The extensions of man. MIT press (1994)

[13] Sakamoto, K., Saito, K., Shigefuji, A., Takahashi, A., Inakage, M.: Pocomz: ambient media device with im application wija. In: ACM SIGGRAPH 2006 Sketches. p. 138. ACM (2006)18 A.D. Cheok et al.

[14] Thomas, M., Flroion, A., Chretien, D.: A new warning biomonitor using a weakly electric fish, apteronotus albifrons (gymnotiformes), and the effect of temperature on the bioelectric responses. Environmental monitoring and assessment 51(3), 605–620 (1998)

[15] Thompson, K.L., Gullone, E.: Promotion of empathy and prosocial behaviour in children through humane education. Australian Psychologist 38(3), 175–182 (2003)

[16] Wang, H., Cao, G., Prior, R.L.: Oxygen radical absorbing capacity of anthocyanins. Journal of agricultural and Food Chemistry 45(2), 304–309 (1997)

[17] Wu, S.C., Lin, S.P., Chen, C.M., Choo, K.B., Cheng, W.T.K.: Production of transgenic pig haboring cdna of human factor ix gene. Journal of Genetics and Molecular Biology 10(1), 36–36 (1999)

[18] Yarnall, A.: Intelligent automation uses bluegills to test drinking water supplies; detection system more complex than canaries in coal mines. San Diego Business Journal 37, 378–385 (2006)

[19] Zeng, Z., Shan, T., Tong, Y., Lam, S.H., Gong, Z.: Development of estrogen-responsive transgenic medaka for environmental monitoring of endocrine disrupters. Environmental

Company's performance prediction using Balanced Scorecard software and neural networks as a tool for strategic management

M. Tufegdžić[*], G. Jovičić [**], M. Trajanović[***] and P. Pravdić[****]
[*]Academy of professional studies Šumadija /Information Technology, Trstenik, Serbia
[**]Technical College of Applied Sciences/Mechanical Engineering, Zrenjanin, Serbia
[***]Faculty of Mechanical Engineering/Department for Production, Information Technology and Management, Niš, Serbia
[****]Academy of professional studies Šumadija /Mechanical and Industrial Engineering, Trstenik, Serbia
milica.tufegdzic@vtmsts.edu.rs, goran.ns@hotmail.com, miroslav.trajanovic@masfak.ni.ac.rs, predrag.pravdic@vtmsts.edu.rs

Abstract – Different sets of quantitative and qualitative Key performance indicators (KPIs), tailored to company's business strategies, require different metrics. The management of the company is very often facing difficulties in the processes of tracking, following and monitoring multiple financial and non-financial indicators, arranged in 4 different perspectives, as a base for accurate prediction of overall performance values using Balanced Scorecard (BSC) software. Combining the capability of the BSC and artificial intelligence it is possible to predict overall performance based on the values of learning and growth, internal processes and customer perspective. The study was conducted on sample of 53 different companies. Appropriate data sets were collected from different sources and converted in a manner which enables processing and calculating the values for financial, customer, internal processes and learning and growth perspectives in BSC software. Overall performance values are also obtained. These values are further used as input data, processed with tools of artificial neural networks in order to generate predictive model for overall performance. The results of our study are presented in the form of tables and graphs. Presented approach is useful for forecasting business performance, time-efficient and applicable to any company, regardless of its size or number of employees.

I. INTRODUCTION

Modern business takes place under conditions of great market competition and constant change, and the vast majority of companies cannot survive without the use of appropriate tools for measuring performance and success. In order to monitor and achieve the goals, to harmonize every day activities with a business strategy, to define priority tasks, services and products, it is necessary to use appropriate tools for strategic management and planning, evaluating and controlling overall business operations, as well as a tool for measuring progress in achieving strategic goals [1].

The Balanced Scorecard principle, as one of the most popular performance management tools, provides a strategic instrument for linking the enterprise strategy management with operations that can be measured. BSC, as a strategic management accounting technique, is a conceptual framework for translating strategic organization's goals into a set of performance indicators that are divided into 4 balanced development perspectives: finance, customers, internal business processes, learning and development [2,3]. Key performances are affected by many factors, which are important for improving business performance [4].

Different approaches for exploring, determination and measuring the most appropriate key performance indicators (KPIs) are used. Prediction approaches are used to provide suitable KPIs and possible value range [5]. Identifying and establishing KPIs relationships are important issues which are elaborated and considered [3]. For that purpose statistical and multi-criteria decision aid techniques are applied. In the cases when enough historic KPI data is available, statistical tools and techniques such as correlation, regression, analysis of variance and multivariate analysis of variance are the most appropriate for predicting KPIs values, although multi-criteria decision aid techniques are less resource consuming [6].

Detecting the right KPIs which are clearly linked with organizational business goals from historical organizational data was conducted using data mining algorithms implemented with Weka. Correlations between KPIs were found using Frequent Pattern growth algorithm [7]. Providing necessary information about the actual business situation from questionnaires was enabled with optimized KPI selection model, implemented in the case of manufacturing company located in three different countries. Applying expert decisions and the outlier's methods in data analysis process leads to reduction of necessary KPIs [8]. KPIs values prediction based on 532 observations values for different KPIs was done using artificial neural network with four inputs and one hidden layer. Finance, customer, operations and learning and growth perspectives were used as inputs, while the outputs were obtained as predicated KPI values.

Association rule Apriori Algorithm was used to discover the relations between KPIs [9]. Scenario from supply chain management domain was the base for developing system dynamics model. Causality relationship between KPIs and dynamic behavior of KPIs, combined with the optimal control, has led to achieving of the objectivities which were previously defined by objective function [10].

Measuring overall performance using BSC software requires data collection of different performance indicators which are arranged in 4 different perspectives. Considering that this procedure requires a lot of time, different methodology is proposed. This methodology involves articifal neural network based approach in nonstochastic prediction of overall performance values.

II. DATA AND METHODOLOGY

The proposed methodology was conducted through the following steps:

- ✓ Acquiring relevant data from different sources,
- ✓ Data analysis and it's conversion to a form that is suitable for processing in appropriate BSC software,
- ✓ Calculating the values for financial, customer, internal processes and learning and growth perspectives in BSC software which are further used as input data for development artificial neural network model.

The sample in our study involves 53 small, middle and big companies from Serbia, in mostly traditional industry branches. Number of employees varies in these companies - from 50 (in small companies) to more than 700 (in big companies). Data has been collected from interviews with responsible persons (managers, employees, department managers, directors, etc.) in companies, while employees have filled out specially prepared questionnaires. Technical and economic documentation from those companies has been studied and used as a valuable source of data.

Tracking financial performance is done using KPIs such as overall assets (in local currency - RSD), overall assets/employee (RSD.), revenues/overall assets (%), profits/employees (RSD), market value (RSD) and return on overall assets (%). Customer perspective KPIs used for the purpose of our study are: number of customers, market share, annual sales per customer (RSD), customers lost (No or %), average time spent on customer relations (hours), customers/employee (No or %), sales closed/sales contacts (%), satisfied-customer index (%), customer-loyalty index (%), number of visits to customers (No), No of customer complaints, marketing expenses (RSD), brand-image index (%), average time from customer contact to sales response (No), average time spent on customer relations (No), service expense/customer (RSD, at year's level). Internal processes perspective KPIs separated for our study are: administrative expense/overall revenues (%), production process time (min.), on-time delivery (%), average lead Time (No), lead time (min.), product development (No), lead time, from order to delivery (No), average decision-making time (No), supplies turn over (No), productivity improvement (%), information technology capacity in the company (No), capacity/employees (No), expenses/administrative expenses (%), the impact of the production process on the environment (No). Learning and growth KPIs important for our case are: investment in training/customers (No), investment in research (RSD), investment in new product support and training (RSD), investment in development of new markets (RSD), direct communications to customers (No, at year's level), patents pending (No), suggested improvements/employee (No), expenses of developing competitiveness/employee (RSD), employee satisfaction index (No), market expense/customer (RSD), employee's view (empowerment index) (No), annual training costs (RSD), communication and support programs per individual.

The values for business goal ranks for all 4 balanced development perspectives represent input values for the BSC Designer Light software, together with their maximum, minimum, target (target-desired), baseline (basic) and actual (current) values. For each perspective goal the weight (so - called goal weight) is determined, ranging from 1 to 10, where the sum of all goals per perspective must not exceed 10. Also, for each perspective a proper value of weight is defined, with the condition that their sum (all 4 perspectives) must not exceed 10. The values of perspectives for all companies are obtained. For each perspective - the mean (M), standard deviation (SD), minimum and maximum were calculated. These values, after eliminating outliers and their robust scaling are used as input data for developing predictive model for overall performance, treated as target data. Part of input and target data is presented in Table 1. For simplicity, learning and growth, internal processes, customer, financial and overall perspective, are marked as p_1, p_2, p_3, p_4 and p_u, respectively. Numerical values of p_1, p_2, p_3, p_4 and p_u are formed into matrices *m x n* within Matlab software. The input data is in the format of *4x53*, while target data is in the format of *1x53*, which are later used for training of artificial neural network model.

TABLE I. PART OF INPUT AND TARGET DATA

p_1	p_2	p_3	p_4	p_u
- 0.29213	- 0.38273	- 3.11386	- 0.36685	- 0.2812
0.418539	0.544454	- 0.92169	1.873693	0.335901
- 0.05267	-0.38103	0.31506	1.246179	0.281972
0.940309	0.859441	0.793976	1.993564	1.212635
- 0.96138	-1.99577	- 0.5753	0.149638	- 0.89522
0	- 0.38273	- 0.03012	0.147224	0
- 0.29213	- 1.14903	0.077108	- 0.16895	- 0.5886
- 0.13624	0.121084	0.562651	0.200322	0.022342
- 0.62008	0.215072	0.29759	0.550282	- 0.03775
0.86236	0.826418	0.953614	1.188254	1.065485

III. ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANN) are mathematical structures that contain interconnected artificial neurons that mimic the work of biological neural network. These artificial neurons have ability to collect, store and use experimental knowledge, actually they have ability to learn.

In order for artificial neural networks to have ability to learn, it is necessary to be trained with appropriate data sets. This dataset must be brought in connection with the change of output quantities. This, indicates the fact that application of artificial neural networks require a sufficient amount of experimental data to conduct the training, then testing the network in term its ability to generalizes the relationship between input and output variables.

The training data sets divided into three groups: training, validation and test data. Training data is used for the purpose of adjusting the weights and bias. Validation data serves to make a decision on terminating the training process and avoid over fitting, while a test data is used for measurement purposes network performance. These data are essential because they don't participate in training process.

In additional to the training input data, there is an additional data set used for artificial neural network testing. Testing of artificial neural networks is done by presenting networks unknown input data.

Development of neural models, Fig. 1, is essentially a complex task and involves solving many questions such as:

- ✓ Determination of input and output quantities,
- ✓ Selection of system for generating the necessary input/output data for training, validation, and performance testing of future neural models,
- ✓ Generating of a sufficient number of input/output data for predetermined ranges changes of each input quantity,
- ✓ Data preprocessing,
- ✓ Selection of architecture,
- ✓ Selection of learning algorithm,
- ✓ Training (for different initial values and weights),
- ✓ Validation of trained network,
- ✓ Testing of the performance of neural models.

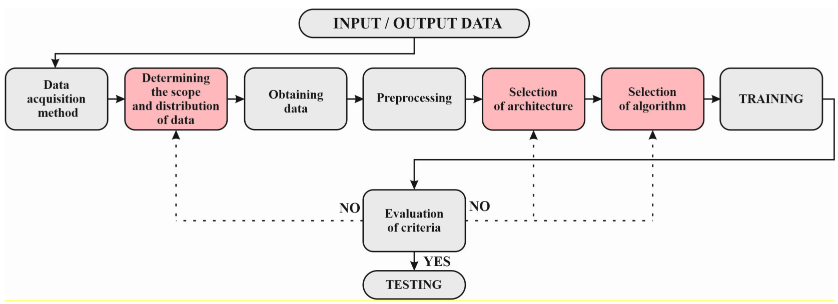


Figure 1. General concept of development of artificial neural network models [11]

IV. ANN MODEL DEVELOPMENT

This section presents our solution for prediction overal perspective based on the feed – forward artificial neural networks with back propagation training algorithm.

The proposed ANN architecture has four input parameters which correspond to balanced development perspectives (p_1 – learning and growth, p_2 – internal processes, p_3 – customer perspective, p_4 – finance perspective), and one output neuron that correspond to the p_u - overall perspective, Fig. 2.

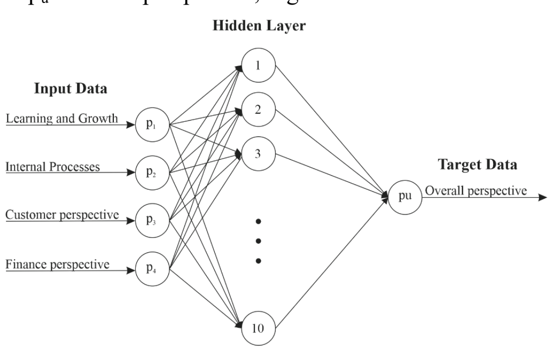


Figure 2. Architecture of the ANN model

The optimal number of hidden neurons can be determined using various rules such as: rules of thumb [12] and rule of pyramid states proposed by Masters [13], but obviously, there is no general procedure to find an optimal ANN architecture and the mentioned rules can serve only as a good guidelines. In this study, optimal ANN architecture will be found based on the experience (trial & error method).

The ANN training is conducted in Matlab 8.1 (R2013) through the ANN Toolbox (*nftool*). Dataset was randomly divided into three parts in proportion 70:15:15. The first part (37 samples) was used as ANN training set, the second part (8 samples) as validation set, and the third part (8 samples) as the testing set. For all three dataset parts *Mean Squared Error - MSE* values were calculated as performance measure.

The teaching, validation and testing sets were obtained using function *dividerand*. As a network training function we have used trainlm (*Levenberg - Marquardt - LM*) that is in most cases the fastest back-propagation algorithm in the Matlab neural network toolbox and highly recommended as a first choice for supervised algorithm, Fig. 3.

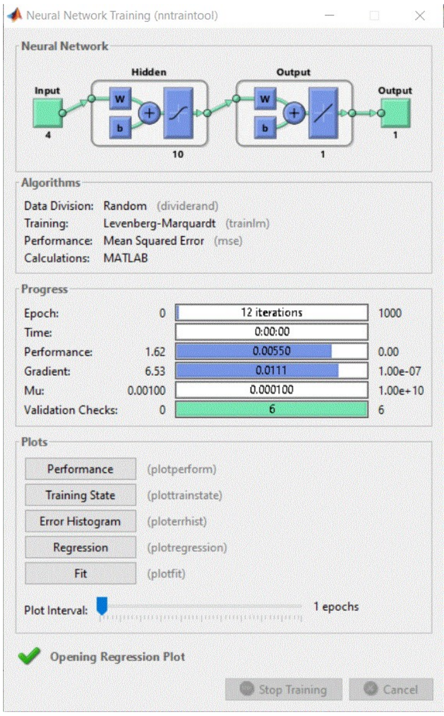


Figure 3. Training of ANN

TABLE II. TRAINING INFORMATION

Name	Training Parameters	Value
Minimum Gradient	min_grad	1e-07
Mu	mu	0.001
Mu Decrease Ratio	mu_dec	0.1
Mu Increase Ratio	mu_inc	10
Maximum Mu	mu_max	10000000000
Maximum Epochs	epochs	1000

In our study, we have tested various architectures, and select an optimal for problem we are dealing with. Various architectures imply different number of hidden neurons, different algorithms, as well as data scope and distribution. Our optimal architecture consists of 10 hidden neurons and LM algorithm with 53 samples.

The process of training, validation and testing ANN with LM algorithm is presented in Fig. 4.

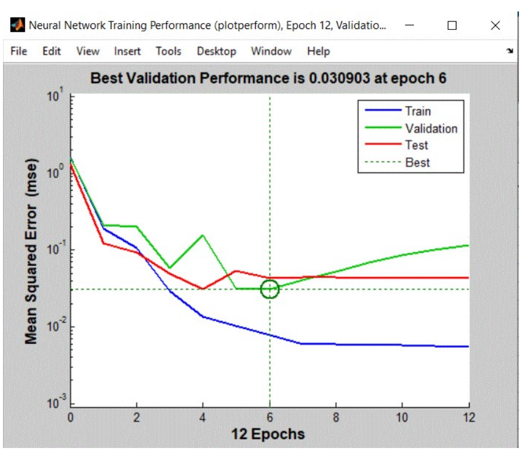


Figure 4. ANN training performace (optimal architecture)

Performance of optimal architectures with correlation factor R and Mean Square Error is shown in Fig. 5 and Fig. 6.

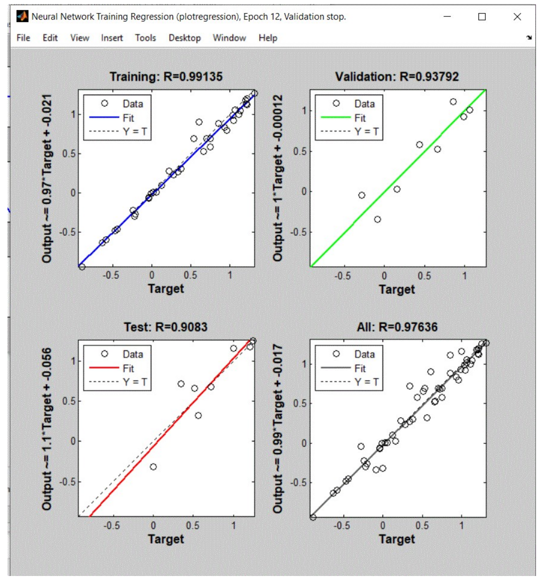


Figure 5. ANN training performace (optimal architecture)

Based on the training performance results, basic conclusion follow. Developed optimal ANN model is acceptable, because correlation coefficient R = 0.97636, which is a measure of how well the variation in the output is explained by the targets. If this number is R=1, then there is a perfect correlation between targets and outputs. In our example, the number is very close to 1, which indicates a good fit.

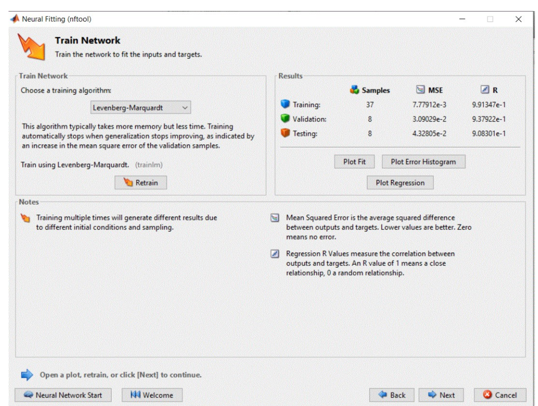


Figure 6. Neural network fitting tool with MSE and R values

V. CONCLUSION

Neural networks offer methods for establishing proper dependencies between learning and growth, internal processes, customer and finance perspectives, with the overall perspective. Connecting qualitative performance indicators with quantitative indicators, enables forecast of the effects of future strategic decisions, being oriented to organization strategic goals and fulfilling the set strategy more efficiently.

Based on obtained results, especially in terms of R values, it is obvious that the developed model is acceptable. However, some improvements of the developed model should be made. In order to find a better solution, it would be necessary to increase a set of input/output data for training, as well as additional data set (which does not participate in the training process) for testing. This is of particular importance due to deterioration of the results by reducing input/output training data and extracting data for additional testing.

REFERENCES

[1] Ishaq Bhatti, M.; Awan, H.M.; Razaq, Z. (2014) The key performance indicators (KPIs) and their impact on overall organizational performance. Quality and Quantity, v. 48, n.6, p. 3127–3143.

[2] Kádárová, J.; Durkáčová, M.; Kalafusová, L. (2014) Balanced Scorecard as an issue taught in th. e field of Industrial Engineering. Procedia - Social and Behavioral Sciences, v. 143, p. 174 – 179.

[3] Köppen V., Graubitz H., Arndt HK., Lenz HJ. (2008) A Procedure to Estimate Relations in a Balanced Scorecard, in: Preisach, C.; Burkhardt, H.; Schmidt-Thieme, L.; Decker, R. (Eds.), Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg, p. 363-371.

[4] Rajnoha, R.; Lesniková, P. (2016) Strategic Performance Management System and Corporate Sustainability Concept - Specific Parametres in Slovak Enterprises. Journal of Competitiveness, v. 8, n. 3, p. 107 - 124.

[5] Badawy, M.; ABD El-Aziz, A.A.; Idress, A. M.; Hefny, H.; Hossam, S. (2016) A survey on exploring key performance indicators. Future Computing and Informatics Journal, v.1, n. 1-2, p. 47-52.

[6] Rodríguez, R. R.; Alfaro Saiz, J., J.; Ortiz Bas, A.; Verdecho Sáez, M. J. (2010) Identifying relationships between key performance indicators. In: 4th International Conference on Industrial Engineering and Industrial Management XIV Congreso de Ingeniería de Organización Donostia - San Sebastián, September 8th -10th, 2010, p. 151-159.

[7] Sultan, N.; Khedr, A.; Idrees, A.; Kholeif, S. (2017) Data mining approach for detecting key performance indicators. Journal of Artificial Intelligence, v. 10, n. 2, p. 59-65.

[8] KaganskI, S.; Majak, J.; Karjust K.; Toompalua S. (2017) Implementation of key performance indicators selection model as part of the Enterprise Analysis Model. Procedia CIRP, v. 63, p. 283 – 288.

[9] Mohamed ABD El-Mongy, A.; Hamouda, A.; Nounou, N.; Wahdan, A., AM. (2013) Design of Prediction System for Key Performance Indicators in Balanced Scorecard. International Journal of Computer Applications, v. 72, n .8, p. 10-14.

[10] Lianjun, A.; Jeng J.J.; Markus, E.; Jen-Yao, C. (2004) A System Dynamics Framework for Sense-And-Respond Systems. In LIN, KJ; LI, T. (Eds.), Proceedings of the IEEE International Conference on E-Commerce Technology for Dynamic E-Business, Beijing, China: CEC-East 2004, p. 6-13.

[11] Miljković, Z., Aleksendrić, D.: Veštačke neuronske mreže – Zbirka rešenih zadataka sa izvodima iz teorije, Univerzitet u Beogradu – Mašinski fakultet, Beograd, 2009.

[12] Panchal, F.S., Panchal, M.: Review of selecting Number of Hidden Nodes in Artificial Neural Networks, International Journal of Computer Science and Mobile Computing, Vol. 3, No. 11, p. 455-464, India, 2014.

[13] Masters, T.: Practical neural network recipies in C++, Academica Press, San Diego, California, 1993.

Structured data chatbot software algorithm prototype

F. Đorđević[*], D. Ivetić [**]
[*] Faculty of Technical Sciences/Department of Power Electronic and Telecommunication Engineering, Novi Sad, Serbia
[**] Faculty of Technical Sciences /Department of Computing and Control Engineering, Novi Sad, Serbia
filip.djordjevic@uns.ac.rs; ivetic@uns.ac.rs

Abstract – In this paper we implement chatbot software and propose an algorithm for the response retrieval. The chatbot itself is part rule-based part conversational and relies on the user's feedback to achieve better performance and connect request messages to the responses from the data. On average, when utilizing the proposed algorithm, the chatbots retrieves responses 19.1x faster, and the biggest gain was 25.6x faster retrieval. The responses are structured so that the chatbot can be used for any domain knowledge, but response knowledge-base has been made so that it can be used for education about the Windows Presentation Foundation (WPF) Framework.

I. INTRODUCTION

A chatbot is software that uses natural language processing (NLP) to understand what its user wants and guides them to their desired outcome, so that the user does as little work as possible.

The term "ChatterBot" was originally introduced by Michael Mauldin (creator of the first "Verbot") in 1994 with the goal to describe these conversational programs [1]. In his famous article "Computing Machinery and Intelligence" from 1950, Alan Turing proposed what is now called the Turing Test which measures the ability of a computer program to impersonate a human in a real-time written conversation with a human judge to the extent that the judge is unable to distinguish conversational content between the program and the human. This is called the "Imitation Game" [2]. The above-mentioned was the first chatbot and its goal was to pass the Turing Test.

Chatbots can be found as a helpful tool in e-commerce, education, entertainment, finance, health, news, and many other fields.

A. Inner workings of a chatbot

Some chatbots just scan for general keywords and generate responses using phrases read from an associated library or a database, while others use artificial intelligence and Natural Language Processors.

Nowadays, many chatbots don't rely upon only typed conversation, but use other functional features such as games and web searching abilities. These bots have found usage on a platform called Discord, which is mostly intended for gamers, where they can play music in voice channels, notify users about tweets, or even display data from internet in chatrooms.

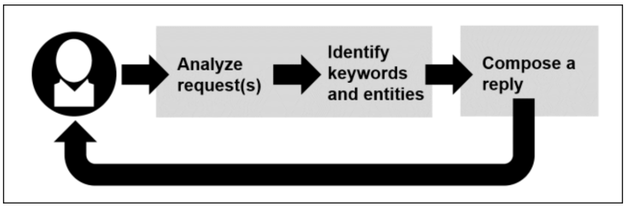


Figure 1. Diagram that represents the way that the chatbot works

Inner workings of a chatbot can be divided into two tasks (as shown in Figure 1):

- User request analysis - **identifying the user intent** and **extract relevant entities**. The ability to identify the user's intent and relevant entities contained in the user's request is the first condition and the most relevant step at the core of a chatbot.
- Returning a response - chatbot must provide the appropriate response for user's request. The answer itself may be:
    - A generic and predefined text
    - A text retrieved from a knowledge base that contains different answers
    - A piece of information based on data the user has provided
    - Data stored in enterprise systems
    - The result of an action that the chatbot performed by interacting with one or more backend applications
    - A question that helps the chatbot to correctly understand the user's request

B. Types of chatbots

Chatbots can be divided into two types:

- Rule-based chatbots follow the predefined paths during conversations. At each step during the conversation, the user will need to pick from presented options that determine the next step in the conversation. That makes it easy to use the chatbot for the simpler scenarios. Interactions with Rule-based chatbots are highly structured and are mostly present in customer support.

- Conversational chatbots are also referred to as virtual assistants or digital assistants. They are much more interactive and personalized than rule-based chatbots. Conversational chatbot's approach to conversing is in the way humans converse and communicate in real-life situations. They can understand the context and intent of complex conversations and try to provide more relevant answers.

C. Related Work

More complex solutions regarding chatbots (in most cases fully conversational chatbots) often seem to utilize machine learning and large training sets, as others go for more structured data and simpler algorithms, as will be shown in these examples.

An interesting chatbot solution which combines both the rules-based and conversational chatbot types of interaction was presented in [3]. In it, Seyner, Savenkov and Vakulenko used pre-generated options in order to engage the user if he/she wants to search or explore the possible data to retrieve and present to the user. They were limited by the structure of the data their chatbot needed to retrieve and present, so their training database contained 250 sample messages. In case of "Search" option, the user can type in his/her request for chatbot to analyze, or when choosing the "Explore" option, the chatbot presents the possible topics to choose from.

In [4] Ilic, Nakano and Hajnal presented a purely conversational chatbot with a constructed dataset of 3000 question-answer pairs that simulate an open-domain chat with generic questions and a mix of humorous, emotional, sarcastic and non-sarcastic responses. Words are mapped into state vectors, which are decoded one token at a time. With the use of a greedy algorithm, next utterance is predicted based on the highest probability at each timestep.

Similarly to the above-mentioned approach, the one described in [5] also features encoders. Vakili and Shakery developed a sequence matching architecture that utilizes the entire training set as a makeshift knowledge-base during making a conclusion. The architecture they introduced is a new one to be used in retrieval-based chatbots, with the model that improves upon the BERT Bi-Encoder baseline without greatly affecting the concluding speed.

Online presence of chatbots has also been explored in [6], where Jo, Yoo, Kim et al. explored the idea of gathering the data from online community posts and using them as pseudo-conversational data. Retrieval of a response is simpler in their solution. Depending on data and the user's query, all the candidates are found by calculating the matching scores, and in the end by ranking the candidates choose the most suitable answers. What they did was to gather data as comments on posts. Words are also translated into vectors and their cosine similarity is calculated to find the best five posts, and the best two ranked comments, measured by likes and dislikes are taken as responses. The only possible "problem" is that the responses will reflect the atmosphere of the community whose posts have been taken as data.

II. Algorithm, Solution Implementation and Results

A. The Algorithm

The Algorithm that is proposed here works as follows: Firstly, it checks if any of the responses in the knowledge-base have the user request message in the associated questions list, and if that particular request message is present, that response will immediately be returned as the response.

If that is not the case, the request message will be split into words and then compared with the keywords of the responses that are associated with the selected topic and the response with the most keywords matched will be the one that the chatbot will return as the response.

The user will then get the option to give feedback to the software if the response retrieved is the one, he/she was looking for. If it is, the question asked will be added to a list of associated questions for that response in order for it to be quicker to access if the user requests the same again.

B. Model and Helper Classes

Since the software is made using the Model View ViewModel (MVVM) Design Pattern and Windows Presentation Foundation (WPF), helper Classes are used to separate the User Interface (UI) logic from Extensible Application Markup Language (XAML) code which is used to define how the software window will look and also Classes that define data models. There are two Model Classes.

Base Class of the Chatbot is the *Response Class*. It is defined with the text of the response, the hyperlink associated with the particular response, path to the image if the response requires an image, two lists with the associated keywords and questions for the particular response and also the topic of the response.

RequestParser Class is used to validate the user's request, choose the response according to request message, and retrieve the answer from the list of answers, and respectively, these functionalities are defined through three methods.

Validation method of the Parser checks if the user is trying to send an empty message, and if it is not empty checks if the request message command-keyword is written correctly.

Choose Response method of the Parser will extract the topic out of the request message and according to the topic call for the retrieval of the response. If the request message was invalid in any, *Choose Reponse method* will return an adequate response.

Retrieve Response method follows the algorithm that this paper proposes. After successfully retrieving the response it will be shown on screen.

Helper Classes in this solution are used to implement the base class for NotifyPropertyChanged functionality, which is achieved with the *BindableBase Class*; *DataIO Class* is used to write and read data from an Extensible

Markup Language (XML) file; and in order to define the interaction between the user and the chatbot software, the *MyICommand Class* contains the base implementation of the Command Design Pattern so that commands can be called from Views and implemented in ViewModels.

C. Software

The rest of the chatbot software is modeled with one View, that being the Main Window and two ViewModels. ViewModel Classes provide data and functionality that will be used in Views. ViewModels define the structure and behaviour of the actual software.

The first ViewModel provides the data and functionality of the message representations inside the chatbot.

MessageViewModel Class defines the color of the "speech bubble" that surrounds the message, and the data defined in the Response Class. Next to that, it is important to define if certain parts of the message design will be shown - almost every response that is retrieved has a link, image, both or none next to the text, but when showing user request messages and responses that a user message is invalid.

These parts do not need to be shown and for that reason it is important to have an option of not showing them. Each message is shown with the keyword: "Request" if typed by the user, "Response" if it was retrieved from the chatbot, and both are shown with a timestamp.

Another part of the algorithm that is implemented here is the option presented to the user to give a feedback to the software itself if the response that he/she got is what he/she was looking for. In the case of user selecting that the response met his/her expectations, the *AddQuestion method* is called to notify the main ViewModel to put the user's message request as the associated question for the chatbot's response.

The *MainWindowViewModel Class* defines the base functionality of the software. It is implemented with the collection of all messages shown, access to the chatbot response data list and also reacts to the event of a user request message being typed in and sent. It then calls the methods explained before to complete the main purpose of the software and return a response to its user. Once it gets the response, it is added to the collection of the messages that are shown.

D. Interaction

The software is designed similarly to other chatting software such as Messenger and WhatsApp. Design and message structured are shown in the Figure 2. On the bottom of the Window is an input field for the user to write their request message. To trigger the sending of a message the user can press the button for sending the message or press enter on their keyboard. After running the algorithm, the software will retrieve the response and show it on the screen, above the user's message input field.

Under the response "chat bubble" the question for the user if the response is what he/she was searching for will be shown with two buttons, where the user can click "Yes" or "No" depending on their answer and the chatbot will associate the user's request message with the response. This is a voluntary action on the user's side, but can be used to "train" the software while it is running.

E. Results

In this part of the paper, results of the measurements of the gains achieved by using the proposed algorithm in the

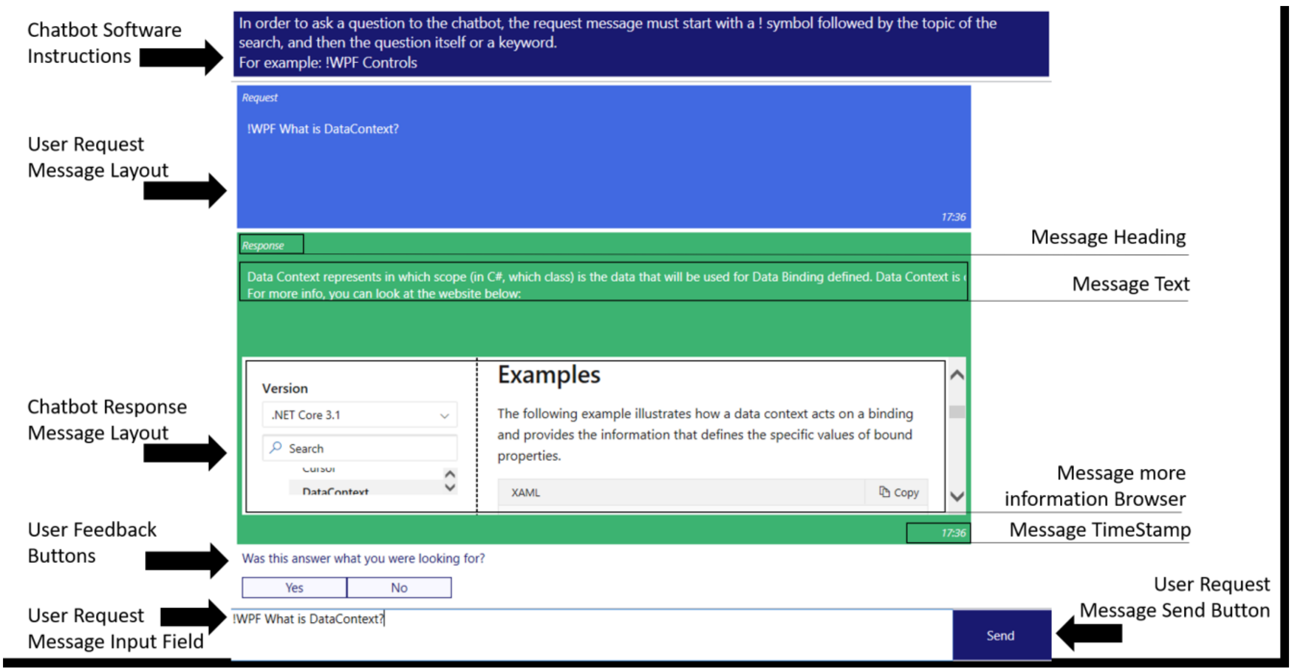


Figure 2. Interface design, interaction example and message structure.

chatbot prototype are presented.

As it was mentioned before, the algorithm is set up with the opportunity for the user to give feedback to the chatbot software and make the algorithm much faster which will be shown in the Table 1.

TABLE I. RETREIVAL ALGORITHM EXECUTION TIMES (MILLISECONDS)

	First time forming the user request message	When the user request message is associated with the response
Only keywords (up to two)	2.872	0.1118 (25.68x faster)
7 word user request message	4.7384	0.2389 (19.83x faster)
12 word user request message	3.236	0.2744 (11.79x faster)

This algorithm is tested on the prepared message list of twenty structured messages that represent the knowledge base for the use in education about the WPF Framework. Test consisted of different ways of forming the user request message, depending on the number of words used and how many keywords are used.

As shown in Table 1, this particular algorithm resulted in, on average, 19.1x faster retrieval of the response if the user request message is associated with an answer. This, since it is voluntary, will not be achieved if the user doesn't give feedback to the chatbot. Algorithm execution time gains are also shown on Figure 3.

III. CONCLUSION

Result of this paper is MVVM implemented software that proposes an algorithm for a chatbot response retrieval. The data used for the prototype is not very impressive, but it still represents the possible gains for this kind of interaction. The solution provided, on average, gets a 19.1x faster execution of the retrieval if the user of the chatbot gives feedback to the software.

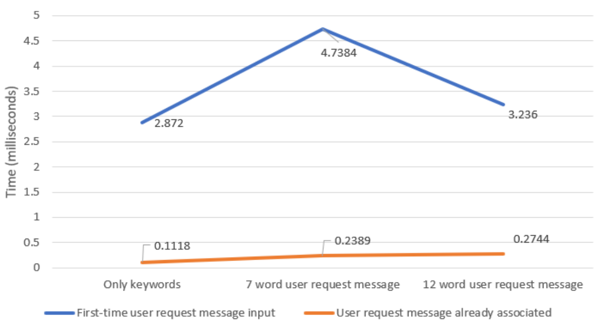


Figure 3. Algorithm execution times.

When compared to other papers presented here, it can be seen that all of them took different ways to test their solutions, and when looking at some concepts that have been proposed have influenced some design decisions here, and also inspired some ideas for future work. The solution with retrieving the answer in [6] has influenced the approach in here, but also the design of the interaction that was presented in [3] has inspired the idea of the feedback part of the algorithm.

One of the greatest challenges when designing chatbot software, as described in [7] is to provide the sample utterances, or templates and message structure with entity placeholders. Next to that, it is quite important to provide a dictionary if entity values, obtain domain-specific knowledge about entities of interest and define all possible structures of user utterance.

Even if this software was made with the intent of it to be used in education, it can be used as a base for basically any domain knowledge, but the future work may be to make it either a local database retriever or a web software with a database where all the users can provide feedback and "train" the software.

Also, with the emergence of Discord as not only the gamer-oriented service, but also a chatting and voice communication software for every possible community, and its bot inclusion possibilities, maybe a chatbot with a similar algorithm and data structure could be implemented on a Discord server.

ACKNOWLEDGMENT

This research has been supported by the Ministry of Education, Science and Technical Development through the project no. 451-03-68/2020-14/200156: "Innovative scientific and artistic research from the FTN (activity) domain".

REFERENCES

[1] M. Mauldin, "ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition", *Proceedings of the Eleventh National Conference on Artificial Intelligence*, AAAI Press, 1994

[2] A. M. Turing, "Computing Machinery and Intelligence", Mind 49, 1950, pp.433–460.

[3] S. Keyner, V. Savenkov, S. Vakulenko, "Open Data Chatbot", June 2019 [The Semantic Web – 16th International Conference, ESWC 2019, Slovenia 2019]

[4] S. Ilic, R. Nakano, I. Hajnal, "Designing dialogue systems: A mean, grumpy, sarcastic chatbot in the browser", University of Innsbruck, MLT Labs, September 2019.

[5] A. Vakili, A. Shakery, "Enriching Conversation Context in Retrieval-based Chatbots", University of Tehran, November 2019.

[6] S. Jo, S. Yoo, S. Im, S. Hee Yang, T. Zuo, H. Kim, S. Han, S. Kim, "A scalable chatbot platform leveraging online community posts: a proof-of-concept study," January 2020 [HCI Conference 2020, Republic of Korea].

[7] W. Shalaby, A. Arantes, T. GonzalezDiaz, C. Gupta, "Building chatbots from large scale domain-specific knowledge bases: challenges and opportunities", December 2019.

Blockchain Technology and its Application in the Finance and Economics

Mimoza Mijoska *, Blagoj Ristevski **
Faculty of Information and Communication Technologies, "St. Kliment Ohridski" University – Bitola
Partizanska bb 7000 Bitola, Republic of Macedonia
* mimozamijoska@yahoo.com, ** blagoj.ristevski@uklo.edu.mk

Abstract - As a new technology, blockchain can be used to analyze and process the data through the effective integration of financial resources. New financial formats or service models are produced to upgrade the financial system and promote the efficiency and quality of financial operations and service from three layers: data, rules and application, based on customers' needs. The blockchain technology can help the financial industry to automatically and accurately identify customer credit conditions, restructure the financial market credit system, and improve the efficiency of cross-border payments. It also posed a challenge for the financial industries' development. In this paper, we systematically analyzed the blockchain technology and its application in the financial and economic domains. With the introduction of blockchain technology in banking operations, the possibility for banks to grant non-performing loans is reduced, thus improving their effectiveness. Using the possibilities offered by the new technology will not allow a situation in which the bank should go bankrupt. Among the other advantages offered by this technology is the great acceleration of the international transfer of funds, which can be done in a few minutes, and in traditional banking, it takes several days in developed economies or even more so in developing countries.

I. INTRODUCTION

Blockchain technology was introduced in 2008 with Nakamoto's white paper "Bitcoin: A Peer-to-Peer Electronic Cash System" [1]. The first blockchain is Bitcoin. Beyond their use in the economic domain, Bitcoin and blockchain technology as articulated by Nakamoto, solve an important computer science problem that had been a barrier to having a functional digital monetary system for years: the double-spending problem. The double-spending problem is that money should only be spent once. The first Bitcoin transactions occurred in January 2009.

Corporate, industrial, and government interest in blockchain technologies is high because applications extend well beyond the domain of cryptocurrencies. There are four main application classes for blockchain technology:

(1) monetary assets (currency, payments, remittance, finance, securities, and financial instruments),

(2) property (land, real estate, and auto title registries),

(3) contracts (business agreements, licensing, registration, wills and trusts, partnership agreements, and IP registration), and

(4) identity credentials (passport, visa, driver's license, and birth registries).

Since 2015, a large number of international financial organizations plan to further develop the blockchain system. In 2014, a consortium called R3 was established, to start research and development of blockchain technology. In March 2017, this group counted about 75 companies to reach 200 in March 2018, including Bank of America, Merrill Lynch, UniCredit Group and many other real estate companies with the goal of better education, law and technology development in blockchain technology [14]. The formation of such a strong corporation with a lot of research and implementation of blockchain technology, especially in the financial sector, tells us that it is a new era in the development of banking before us. In our country, individuals have expressed concern that the conditions for granting credit facilities are often too strict, or that they need to secure a mortgage, or that they do not have access to credit at all. Interest rates are steadily different from the average European Union and significantly higher risks are being borne, which leads to unstable economies, but also high operating costs at which new technologies can be applied can act. Applying new technology, the banking sector would create conditions for a significant increase in operating, reducing exposures, a large number of risks that would result in innovation in lending activities, even lower interest rates on placements.

The rest of the paper is organized as follows. Distributed ledger technology is explained in Section II, while SWOT analysis of using the blockchain technology in finance and economics is depicted in Section III. The subsequent section is devoted to the application of blockchain technology in financial and economic domains and the new challenges of the application of blockchain technology are described in Section V. The last Section gives concluding remarks.

II. DISTRIBUTED LEDGER TECHNOLOGY

Blockchain (distributed ledger technology) is a network software protocol that enables the secure transfer of money, assets, and information via the Internet, without a third-party organization as an intermediary [2]. It can safely store transactions such as digital cryptocurrencies or data/information about debt, copyrights, equity, and digital assets. The stored information cannot be easily forged and tampered because it requires individual approval of all distributed nodes. This significantly reduces the cost of trusting and accounting that commonly exist in non-digital economies and other social activities. Blockchain has four components:

(1) hash, which uses one-way mathematical functions to assign unique indexes;

(2) a digital signature, which is implemented as a public cryptographic key;

(3) peer-to-peer (P2P) network, which serves as a routing structure for nodes to use the distributed hash; and

(4) consensus mechanism, which is a set of digital procedures designed to ensure the accuracy and consistency of the stored information across the participating nodes.

The blockchain data structure is depicted in Fig.1 [3].

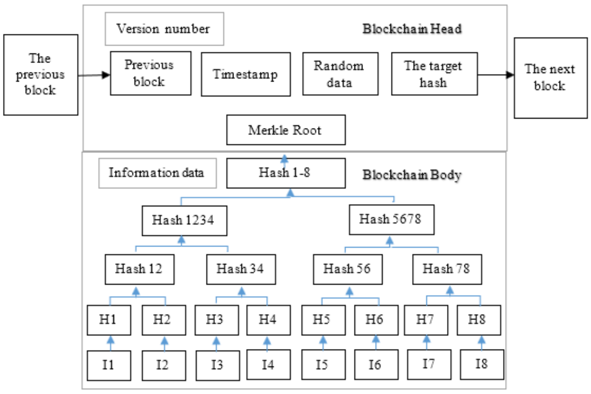


Figure 1. The blockchain data structure.

In the Blockchain Body, the bottom is a part of Merkle Hash Tree which can be either a binary tree or a multitree in the data structure. Specifically, data or information is recorded as the hash value stored in Blockchain Body, and the generated Merkle root through Merkle tree's hash process will be recorded in Blockchain Head.

The blockchain technological platform is gradually shaping up into three directions: (1) underlying infrastructure which includes facilities for mining and manufacturing of specialized computer hardware to perform blockchain-related tasks; (2) middle layer between the blockchain platform and client application services, including smart contracts, a blockchain platform, financial software, and other services; and (3) hotspot distributed applications in various industries, including finance (e.g., cross-border payments, liquidation, financial services, and asset digitization), cybersecurity (e.g., identity protection, data authenticity protection, and critical infrastructure protection), and supply chain management (e.g., logistics tracking and digital works tracking).

This distributed general ledger is replicated to thousands of computer nodes around the world and is publicly available. Despite all its openness, it is also confidential and reliable. This is achieved through a mathematical puzzle and computer power embedded in its "consensus mechanism" - the process in which the nodes agree on how to update the blockchain with each transaction of moving the value from one person to another. Users use public and private keys to digitally sign and make transactions in the system in a secure way. Blockchain users can solve puzzles using cryptographic hashing methods hoping to be rewarded with a fixed amount of cryptocurrency [4].

Blockchain systems seem complex. However, they can be easily understood by examining each technology component individually. At a high level, blockchains utilize well-known computer science mechanisms (linked lists, distributed networking) as well as cryptographic primitives (hashing, digital signatures, public/private keys) mixed with financial concepts (such as ledgers) [5].

A ledger is a collection of transactions. Ledgers are often stored digitally in large databases owned and operated solely by centralized "trusted" third parties however we must trust the third party that the data is backed up, transactions are validated and complete, and the history is not altered. A ledger implemented using a blockchain can mitigate these issues through the use of a distributed consensus method. One of the aspects is that the blockchain ledger will be copied and distributed amongst every node within the system. When new transactions are submitted to a node, the rest of the network is alerted that a new transaction has arrived and at this point, this is a pending transaction. Eventually, one of the nodes will include this new transaction within a block and complete the system's required consensus method. This new block will be distributed across the network and all ledgers will be updated to include the new transaction. When new users join the system, they receive a full copy of the blockchain, making loss or destruction of the ledger difficult. Each transaction that is submitted to the network passes through several steps to be included and published in a block of the blockchain:

A transaction is a record of a transfer of assets (digital currency, units of inventory, etc.) between involved parties. For each input transaction A, an output hash value #A is created using a cryptographic function.

Hashing is a method of calculating a relatively unique fixed-size output for an input of nearly any size (e.g., a file, some text, or an image). Even the smallest change of input will result in a completely different output digest. Hash algorithms are designed to be one-way: it is computationally infeasible to find any input that maps to any pre-specified output. If a particular output is desired, many inputs must be tried by passing them through the hash function until input is found that gives the desired result. Moreover, hash algorithms are designed to be collision resistant: it is computationally infeasible to find two or more inputs that produce the same output. A commonly used hashing algorithm in many blockchain technologies is the Secure Hash Algorithm (SHA) with an output size of 256 bits (SHA-256).

Each block in a blockchain contains multiple transactions, transactions are grouped in sets. Hash values are further combined in a system called a Merkle tree [15].

Merkle tree is a data structure where the data is hashed and combined until there is a singular root hash that represents the entire structure. The root is an efficient mechanism used to sum up the transactions in a block and verify the presence of a transaction within a block. This structure ensures that the data sent in a distributed network is valid since any alteration to the underlying data would be detected and can be discarded. The result of all the hashing then goes into the block's header, and it is combined with the hash of the previous block's header and a timestamp. This combination becomes a part of the cryptographic puzzle. The solution for the puzzle is to find a nonce value.

The nonce value is a number manipulated by the mining node to solve the hash puzzle and with this, it gives them the right to publish the block. After creation, each block is hashed thereby creating a digest that represents the block. The change of even a single bit in the block would completely change the hash value. The block's hash digest is used to help protect the block from change since all nodes will have a copy of the block's hash and can then check to make sure that the block has not been changed. An additional feature of blockchain systems is that they can run so-called smart contracts [7], which is an auto executable code that fires off once certain conditions are met. A smart contract is computer protocol or collection of code and data which runs automatically under defined criteria when deployed on the blockchain. The contract executes the appropriate method with the user-provided data to perform a service. The code, being on the blockchain, is immutable and therefore can be used (among other purposes) as a trusted third party for financial transactions that are more complex than simply sending funds between accounts. A smart contract can perform calculations, store information, and automatically send funds to other accounts. It doesn't necessarily even have to perform a financial function.

III. SWOT ANALYSIS

The advantages and disadvantages of using blockchain technology in financial operations presented through SWOT analysis:

Strength:

1) Operational efficiency; The exchange of information with an electronic system can contribute to financial efficiency;
2) Secure encryption; Registry distribution technology provides secure encryption and capabilities to protect against unauthorized access important data;
3) Lower costs; Registry of distributed data for transactions shows the cost of storage in order with the existing method;

Weakness:

1) Given that there is no regulation in our country, it is difficult to adopt the current service;
2) Lack of experts; Good immaturity and short blockchain history, there are only several experts;
3) Possibility of collisions; In the process of blocking, there is a possibility of a collision that did not exist in the existing methods. (Problems with compatibility that are not present in traditional, well-distributed servers, one of the most widely used and stable technologies, blockchain usage may occur).

Opportunity:

1) Transparency transaction that feeds the blockchain enables gaining credentials publicity;
2) Potential partnership; learning from a mature partner that uses blockchain for financial gain;
3) Advanced financial system; integrated ordinary data can improve financial performance in the future.

Threat:

1) Problems of culture and beliefs; Blockchain application of sensitive data can provoke social antagonism. (It's hard for us laity users to simply think about blockchain. However, it does not help to reason only about an abstract concept. if users do not understand the dangers of security and stability technology).
2) Obstacles caused by excessive regulation; Excessive regulation can limit and hinder the realization of blockchain values.
3) Requires motivation; In order to try to use unknown technologies, it is necessary to motivate the software industry to get involved.

IV. APPLICATION OF BLOCKCHAIN TECHNOLOGY IN FINANCE AND ECONOMICS

There are three main theories related to blockchain finance and economics. First, Illing and Peitz [8] proposed a digital economy theory that promotes the convergence of computer and communications technologies to digitize all assets and then register and transfer notarized digital assets on the blockchain. Ultimately, it will realize the complete digitization of goods and services from manufacturing, sales, and supply chain. Second, free currency studied by Chen [9] states that the dialectical relationship between money and freedom from Marx's economic philosophy perspective. Specifically, the currency should not be issued by the central bank of a government but should be completely decentralized and nonstate owned. All kinds of digital currencies can be exchanged freely in the blockchain economy. Additionally, the digital currency repeats the emergence, elimination, and evolution of the competitive process at an extremely rapid rate. Last, Marcel, Oran, and Otgon adopted information asymmetry theory to examine the trust problem as the different information held by both parties [10]. The theory provides methodological guidance for blockchain finance and economics because blockchain can realize the optimal allocation of resources through the form of digital rewards for mining to establish decentralized credit for universal participation.

A. Cryptocurrency and its trading platforms

Bitcoin is the first decentralized cryptocurrency and a worldwide payment system without a central bank. Transactions take place between users directly over the P2P network without an intermediary. These transactions are verified by network nodes through the use of cryptography and recorded in a public distributed ledger called a blockchain. There are many other types of cryptocurrencies, such as *Ethereum* (it can be used by programmers to pay for transaction costs and services in the Ethereum network), Bitcoin Cash (it is generated after bitcoin's hard fork), *Ripple* (it is a global clearinghouse for other currencies or other value entities such as U.S. dollar, Euro, Pound sterling, bitcoin, airline miles, and commodities), *Litecoin* (it is an early replacement of bitcoin designed to allow ordinary people to mine as well), and *Dash* (it can be paid instantly via a unique, two-tiered network). Therefore, the cryptocurrency has attracted more and more countries' attentions and become an indispensable application of blockchain finance and economics.

1) Comparative analysis of national currency and cryptocurrencies

The differences between the national currency and the cryptocurrency are not so significant and include:

- the absence of an issuer – the cryptocurrency is created by "mining", similar to gold or minerals, with the help of significant computer capacities;

- lack of physical form – if modern physical money (paper banknotes or coins) is provided with several degrees of protection, the cryptocurrencies and money in the banking system, at accounts and cards, are protected by complex information technologies, but banking information security and algorithms for creating and transferring cryptocurrency are fundamentally different;

- the absence of a regulator – cryptocurrencies and transactions on them are not controlled by government bodies or any personified and institutional subjects.

Between national and cryptocurrencies there are a lot of common features:

- they serve as a value meter; both currencies' types are stored in electronic records on computing capacities (usually, huge servers of financial institutions);

- they operate on a competitive basis in the market of currencies (national currencies and cryptocurrencies are exchanged at a floating rate).

Since March 1968, when the US stopped exchanging private dollars for gold at a fixed price ($ 35 per troy ounce), and after, August 1971, when this ban was "temporarily" set for the exchange of the dollar for gold also for national central banks, the trust of people to paper money is gradually reducing.

Cryptocurrencies have several advantages over national currencies:

- cryptocurrencies do not depend on changes in the political situation, their course is not subject to geopolitical risks;

- transactions in cryptocurrencies are controlled by a multitude (hundreds of thousands) of computers, but are not subject to the control of any authorities or persons; they are completely anonymous for people and are characterized by a high degree of confidentiality.

Accordingly, such transactions are nor a subject of the international agreements of competent (tax) authorities on automatic information exchange (Multilateral Competent Authority Agreement, MCAA) or on requesting information (the Convention on Mutual Administrative Assistance in Tax Matters), as well as of the US Foreign Account Tax Compliance Act (FATCA), which obliges all financial and non-financial, insurance and similar institutions to provide the US Internal Revenue Service with information on financial accounts and transactions of US tax residents [6].

B. Cross-border payment

Cross-border payments generally refer to the transnational and transregional transfer of funds between two or more countries or territories through international trade, international investment, and other international claims and debts using certain settlement instruments and payment systems. The traditional cross-border payment is based on the banking system which has such characteristics as time-consuming, high cost, more funds occupied, and low security. However, all these bottlenecks can be effectively overcome by applying blockchain to reconstruct the credit system and expand the payment boundary. Researchers pointed out that applying blockchain technology to the cross-border payment has a high potential effect. Holotiuk, Pisani, and Moormann [11] stated that the blockchain technology will improve the payment system by providing a solid structure for cross-border transactions and removing expensive intermediary costs and gradually weaken or alter the business model of the existing payment industries. R3 has been working with 22 of its member banks to build a real-time, cross-border payments solution on Corda that is the consortium's "blockchain-inspired" distributed ledger [12].

C. Digital asset register and management

Related to digital currencies and money transfer, one of the biggest blockchain applications is digital asset register and management. The blockchain technology can record, transfer, and verify asset ownership (e.g., home, auto, stocks, bonds, mortgages, and insurance) and preserve the integrity and authenticity of sensitive documents or records (e.g., passports, visas, driver's licenses, birth and death certificates, voter registration, contracts, wills, patents, and medical records). An exemplary implementation of a digital asset register for identity services is the State of Illinois's blockchain-based birth register [13]. The blockchain patent has been filed by different companies such as the giant Amazon.com and the start-up Coinbase. Additionally, some companies are filing for blockchain-related patents to protect their digital asset.

Regulatory mechanisms are based on the legitimacy of rules that meet basic values (values' coherence and regulatory efficiency), and on the transparency of their implementation, which can be interpreted as the effectiveness of rules-driven normative regulation.

The development of ICT (information and communication technologies) permits to control if the process is carried out correctly and to assure the total coverage of rules' application, providing the adequate normative (i.e. legal) regulation.

D. The COVID-19 coronavirus versus the Red Cross: better solutions via blockchain and artificial intelligence

Beijing has ordered all public donations for the Wuhan crisis to be transmitted to five government-backed charity organizations. This is a throwback to pre-2016 China, before the Charity Law of China was introduced to enable the establishment of private charities. The Charity Law was intended to develop the charity field and protect the interests of relevant stakeholders. Although all charities in China are required to have in place sound internal governance structures, the charging order implicitly assumes that the five government-backed charities are fit for purpose and better able to manage the current crisis. That assumption may be at odds with historical and more recent evidence suggesting organizations responsible for responding to crises appear to struggle to manage their core responsibilities. And if Beijing's implicit assumption is wrong then the centralizing effect produced by charging merely serves to compound the problem.

In this instance, and not for the first time, the Red Cross in China is in the crosshairs of public anger. 'One of the lessons learned was that emergency response must be better developed at the local level'. This is what the Red Cross said in 2017 on the 10th anniversary of the deadly Wenchuan earthquake in Sichuan province in western China. Billions of dollars had been donated following the Sichuan earthquake but had been 'mishandled'. What has been learned? The public in China has again been angered by the mishandling of donations, and this impacts on the willingness to donate, which retards the objective of addressing a problem.

Blockchain and AI are now in frequent use by global tech companies and represent tools that can be used to better manage crises. A private blockchain network would enable the recording and tracking of anything that is donated, from donation dollars to N95 masks. It also creates clear points at which it is possible to hold a person or organization to account, from the loading of donations for delivery through to its final end-use. Importantly, the blockchain can also be given public visibility, providing transparency to all stakeholders - donors and recipients, as well as public oversight bodies. Anyone could track the progress and use of their donation.

V. THE CHALLENGES OF BLOCKCHAIN IN FINANCE AND ECONOMICS

It usually takes 10 min to add a new block to the blockchain. Thus, each block's capacity is only 1 MB and the online capacity allows only eight transactions per second. The blockchain has a huge gap with the current third-party payment named Alipay, which supports thousands of transactions per second. Blockchain establishes the credit guarantee of the trusted intermediary through a program algorithm. However, its information is irreversible, which makes the system more difficult to recover if the private key or password is lost or leaked, resulting in irreparable loss of customer assets. Although the blockchain has the clear technical logic and is theoretically difficult to be violently cracked, the possibility of a data breach still exists through the hostage of a large number of zombies or the operation mode of a trade unions cluster. Therefore, technical risks such as hackers must be addressed with blockchain development.

A. Problems of financial regulation

Financial regulation is a powerful guarantee for Internet financial information security. However, the emergence of blockchain finance and economics has brought decentralization, which greatly increases the relevance and effectiveness of financial regulation. At present, people have a relatively low understanding and acceptance of the blockchain and it is hard to identify real and effective blockchain financial products. It is difficult for regulators to lock in many anonymous accounts of clients and understand the whereabouts of funds. Additionally, the absence of a central system has facilitated convenience for the money laundering, fraud, and tax evasion as well as increased the difficulty of supervision and management.

B. Complexity in global collaboration

The layout of blockchain finance and economics in the world is difficult due to the great cultural diversities and liberal democracy. Using virtual currency as an equivalent to achieving real-time global liquidation is a challenge for the central bank's legal tender and the right of payment. The legal tender is endorsed by national credit.

However, the virtual currency credit is the mathematical algorithm that is difficult to reflect any single country's financial will. For example, only a few countries such as Japan, Germany, and the United States of America recognize the legal status of bitcoin. Most countries cannot accept Bitcoin's monetary attributes, especially in China. Therefore, the further application and development of blockchain technology in the global financial and economic field need to balance the interests of all countries to reach a consensus.

VI. CONCLUSIONS

In this paper, a system for electronic transactions without relying on trust is proposed. The traditional system is incomplete because there isn't a way to prevent double-spending. To solve this, there is a peer-to-peer network using proof-of-work to record a public history of transactions that quickly becomes computationally impractical for an attacker to change if honest nodes control a majority of CPU power. The network is robust in its unstructured simplicity. Nodes work all at once with little coordination. They do not need to be identified, since messages are not routed to any particular place and only need to be delivered on a best-effort basis. Nodes can leave and rejoin the network at will, accepting the proof-of-work chain as proof of what happened while they were gone. They vote with their CPU power, expressing their acceptance of valid blocks by working on extending them and rejecting invalid blocks by refusing to work on them. Any needed rules and incentives can be enforced with this consensus mechanism.

Blockchains could revolutionize the underlying technology of the payment system and credit information systems in banks, thus upgrading and transforming them. Blockchain applications also promote the formation of "multi-center, weakly intermediated" scenarios, which will enhance the efficiency of the banking industry.

It is worth noting that the problems of regulation, efficiency, and security have always sparked extensive debate in the process of each new financial innovation. History is not stopped by current obstacles, as the technical, regulatory, and other problems of blockchain technology will ultimately be resolved. Hence, the prospect of integrating blockchain technology into the banking industry will most likely occur soon.

REFERENCES

[1] S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, 2008. https://bitcoin.org/bitcoin.pdf (Website accessed 07.09.2020).

[2] Swan, M. Blockchain: Blueprint for a new economy, 2015. Retrieved from https://www.goodreads.com/-work/best_book/44338116-blockchain-blueprint-for-a-new-economy (Website accessed 07/09/2020).

[3] Li Zhang, Yongping Xie, Yang Zheng, Wei Xue, Xianrong Zheng, Xiaobo Xu, The challenges and countermeasures of blockchain in finance and economics, John Wiley & Sons, Ltd 2020

[4] Iansiti, Marco, and Karim R. Lakhani. "The truth about blockchain." Harvard Business Review 95.1, Jan (2017). https://hbr.org/2017/01/thetruth-about-blockchain, (Website accessed 07.09.2020).

[5] https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8202.pdf, (Website accessed 07.09.2020).

[6] N. N. Pokrovskaia, Tax, Financial and Social Regulatory Mechanisms within the Knowledge-Driven Economy. Blockchain Algorithms and Fog Computing for the Efficient Regulation, 2017

[7] Szabo, Nick. "The idea of smart contracts". Nick Szabo's Papers and Concise Tutorials, 1997.

[8] Illing G., & Peitz, M., Understanding the digital economy: Facts and theory introduction., *CESifo Economic Studies*, 2015, *51*(2/3), 187¯188. https://doi.org/10.1093/cesifo/51.2-3.187

[9] Chen, F. , A probe into the dialectical relationship between money and freedom¯From the perspective of Marxist Economic Philosophy. *Nanjing Social Science*, 2017, *8*, 75¯81.

[10] Marcel, B., Oran, T., & Otgon, C., Information asymmetry theory in corporate governance systems. *Annals of the University of Oradea, Economic Science Series*, 2010, *19*(2), 516¯522.

[11] Holotiuk, F., Pisani, F., & Moormann, J.. The impact of blockchain technology on business models in the payments industry. Wirtschaftsinformatik 2017 Proceedings. Retrieved from https://aisel.aisnet.org/wi2017-/track09/paper/6 (Website accessed 07.09.2020).

[12] Crosman, P. *R3 to take on Ripple with cross-border payments blockchain*. American Banker; New York, N.Y., 2017. Retrieved from https://search-proquest-com.ucd.idm.oclc.org/docview/1957965453/ citation/F4C0367615-95447CPQ/1 (Website accessed 07.09.2020).

[13] Evernym.com. State of Illinois Partners with Evernym to Launch Birth Registration Pilot, 2017. Retrieved from http:// globenewswire.com/news-release/2017/08/31/1106400/0/en/Stateof-Illinois-Partners-with-Evernym-to-Launch-Birth-RegistrationPilot.html (Website accessed 07.09.2020)

[14] Dragana Tadić Živković, blockchain technology: opportunity or a threat to the future development of banking, Proceedings of Ekonbiz, 2018

[15] Julija Basheska, Vladimir Trajkovik, Blockchain based Transformation in government: review of case studies, XIV International Conference ETAI 2018, Struga, Macedonia.

NoSQL Databases - Analysis and Directions of Further Development

M. Krstić*, L. Krstić*, V. Stanković*
*Academy of Vocational Studies Southern Serbia, Department of Business Studies Leskovac, Serbia
krstic.marija@vpsle.edu.rs, krstic.lazar@vpsle.edu.rs, stankovic.valentina@vpsle.edu.rs

Abstract - NoSQL systems have emerged from new requirements for greater flexibility and better performance for storing large amounts of data. The reason for that is mainly the popularity of the Internet and information technologies and the growing amount of data that is generated every day. The paper presents research on the topic of NoSQL databases. After introducing NoSQL databases, the research presents all relevant aspects that they include. The main focus is on the analysis of their popularity and further directions of development. The paper aims to encompass the popularity of NoSQL databases by individual types in the observed time period and to consider the directions of their further development.

I. INTRODUCTION

Relational databases have been in use for many years, and the reason is their proven reliability and functionality in many systems. However, for large amounts of data that are generated daily from different sources and in different forms, relational databases are not always the best option, i.e. they are not always suitable for storage and processing. The problem arises when it is necessary to accept and process large amounts of data in the shortest amount of time possible. This is the motivation for the creation of NoSQL ("Not only SQL") databases, in which data are not organized according to the principles of the relational model, but according to a much simpler and more flexible data model. In this regard, if it is not necessary to store data in tables or there are links that cannot be represented by standard SQL relations, and at the same time fast data access is required, the solution is to use NoSQL databases.

II. NOSQL DATABASES

The term NoSQL ("Not only SQL", "non-SQL" or "non-relational") describes a whole class of databases that do not have the characteristics of traditional relational databases and over which the standard SQL query language is generally not used. They are defined as the next generation of databases related to the following properties: non-relational, distributed, open source and horizontally scalable, with features such as non-strict static data structure, simple replication support and simple application programming interface (API) [1].

NoSQL databases are based on the BASE model (Basically Available, Soft State, Eventually Consistent), while on the other hand, relational databases basically apply the ACID model (Atomicity, Consistency, Isolation, Durability). This does not mean, however, that NoSQL databases are unreliable and inconsistent. Eventual consistency means that the system will become consistent once all data is propagated in all nodes in the cluster. Also, systems based on the BASE model, as a rule, are much faster and simpler than those based on ACID.

A. Characteristics of NoSQL databases

NoSQL databases were created in order to meet the requirements for managing large amounts of data, enabling work with data without a previously defined scheme, greater scalability and efficiency. Table 1 provides an overview of the characteristics of NoSQL databases.

TABLE I. OVERVIEW OF NOSQL DATABASE FEATURES [2]

Characteristics of NoSQL databases	Description
Data model	The data model is flexible, and the data is not stored in tables, but in different objects, depending on the type of NoSQL database.
Scalability	NoSQL databases allow the so-called horizontal scalability (if there is an increase in the number of users or the amount of data), it is only necessary to add a new server, and if there is a drop in one of the servers, the data can still be accessed through other servers in the cluster.
Data replication	Data replication takes place according to the master-slave principle (all data entered into the database are stored on the master server (master node) and later replicated to other, slave, servers in the cluster (slave nodes)).
Ability to recover the database	In NoSQL databases, individual parts of the database are stored on several servers that are hierarchically interconnected by a master-slave relationship, and after the server crashes, the database management system reviews the active transaction log and reactivates those transactions that were performed before the server crashed.

B. Advantages and disadvantages of NoSQL databases

Table 2 provides an overview of the advantages and disadvantages of NoSQL databases.

TABLE II. ADVANTAGES AND DISADVANTAGES OF NOSQL DATABASES [3]

NoSQL database	
Advantages	**Disadvantages**
Possibility of use as a primary data source or analytical data source	There are no rules for standardization
Ability to store large amounts of data	Limited query options
Easy replication	It does not offer any of the capabilities of traditional databases, such as consistency when multiple transactions are performed
No need for a separate caching layer	When the amount of data increases, it is very difficult to keep unique values as keys
Allows fast operation and horizontal scalability	It doesn't work well with relational data
Efficient processing of structured, semi-structured and unstructured data with equal effect	Open source options are not so popular with businesses
NoSQL database does not require a dedicated high-performance server	
Supports key development languages and platforms	
Easier to implement compared to RDBMS	
It offers flexible design schemes that can be easily changed without downtime or service interruption	

C. Types of NoSQL databases

There are mainly four types of NoSQL databases. Each of them has its own unique characteristics and limitations. No type can be said to be better than the other, it is necessary to choose the one that suits the specific needs. There are four types of NoSQL databases [4]:

- Key-value databases - the simplest type of NoSQL database that stores data without using a scheme, but uses indexes and data values.
- Document-oriented databases - work according to the concept of key-value models but with additional complexity. Each document in this type of database has its own data and its own unique key that is used to retrieve the data. It is a good choice for storing, retrieving and managing data that is document-oriented but still has some structure.
- Wide column databases - store data tables as columns, not as rows. This way, the data is stored more efficiently, because if there is no data for a given column, they will not be entered, while with relational databases, NULL values are entered.
- Graph databases - are designed for data the relationships of which are represented as a graph consisting of elements interconnected with a finite number of relationships between them.

Table 3 presents the key characteristics of the aforementioned types of NoSQL databases.

TABLE III. KEY FEATURES OF FOUR TYPES OF NOSQL DATABASES [5]

Type	Performance	Scalability	Flexibility	Complexity
Key-value databases	High	High	High	Not complex
Document-oriented databases	High	Variable (High)	High	Low
Wide column databases	High	High	Medium	Low
Graph databases	Variables	Variable	High	High

D. Criteria for selecting NoSQL databases

In continuation of the paper, some of the important criteria for selecting an adequate NoSQL database are given, in accordance with specific needs [6].

- Type of warehouse
  - For example, the get, put, and delete functions are best supported by the key-value system.
  - Aggregation becomes much simpler when wide column systems are used.
  - Data mapping becomes easier with object-oriented software that uses document-based NoSQL databases, such as XML or JSON.
  - The tabular format is changed, and the data is stored in a graphic format.
- Parallel execution control - determines how two users can change the same information at the same time.
  - Lock - prevents more than one active user from modifying an entity such as a document, row, or object.
  - MVCC (Multiversion Concurrency Control) - guarantees a readable overview of the database, but results in the appearance of opposing versions of an entity if multiple users modify it at the same time. MVCC allows seemingly seamless transaction processing by creating multiple different versions of a single object. This means that the consistency of transactions is maintained even if, as a consequence, different users are shown different views at any given time. All changes to the database will be displayed to all users depending on which view they are viewing.
  - Some systems lack atomicity, as a result of which all users who change the database do not have the same view of the database.
  - For reliable transactions, ACID should be selected.
- Replication - ensures that copies of data are always synchronized.

   - Synchronous (slave servers are constantly connected to the master server, so changes take place simultaneously).
   - Asynchronous (slave servers are not permanently connected to the master server, so changes that occurred on the master server are saved in a queue and later take place on the slave server).
- Implementation language - helps determine how fast the database can process. Usually the fastest NoSQL databases are written in low-level programming languages. On the other hand, those written in higher level programming languages are easier to edit.

III. FURTHER DIRECTIONS OF NOSQL DATABASE DEVELOPMENT

Numerous studies on the use of NoSQL databases indicate that organizations that need to store large amounts of data are showing serious interest in said databases. However, this is not the case with smaller organizations. Also, a large number of experts in the field of information systems and technologies are not familiar with NoSQL databases, which indicates that they will have to continue to develop in order to gain the greater popularity they deserve.

Observed from the business aspect, it can be concluded that the main benefit of NoSQL databases is that they are very good for certain purposes. On the other hand, they should not be forced for purposes they were not intended for. One of the reasons for the introduction of NoSQL databases is, among other things, that relational databases take up a lot of resources. In addition, the data structure in NoSQL databases is considered more flexible compared to relational databases.

A. Analysis of use and popularity of NoSQL databases

Relational databases are not always suitable for storing and processing data. Namely, if it is not necessary to store data in tables or there are links that cannot be represented by classical SQL relations, and it is necessary to quickly access the data, in these cases it is more convenient to use NoSQL databases. In order to analyze the representation and predict the further development of NoSQL databases, a cross-section of the current popularity of database management systems by category was made. Figure 1 presents the popularity of individual database categories expressed as a percentage. The results are taken from the DB-Engines initiative website, which collects and presents information on the popularity of database management systems. The popularity of the system is measured using the following parameters [7]:

- The number of system impressions on websites, measured as the number of results in search queries. Google, Bing and Yandex are currently used for this type of measurement. To count only relevant results, the system name is searched for together with the term databases for example "Oracle" and "database".
- General interest in the system, i.e. the measurement for which the frequency of searches in Google Trends is used.
- Frequency of technical discussions about the system. The number of related questions and the number of interested users on the well-known IT Q&A sites Stack Overflow and DBA Stack Exchange are used.
- Number of job offers in which the system is stated. The number of bids in the leading searches for the job Indeed and Simply Hired is used.
- Number of profiles on professional networks in which the system is listed. The internationally most popular professional networks LinkedIn and Upwork are used.
- Relevance on social media. Twitter tweets in which the system is mentioned are counted.

The results are updated once a month. The popularity of the system is calculated using standardization and averaging of individual parameters. These mathematical transformations are made in a way that preserves the distance of individual systems. This means that when system A has twice the value in the DB-Engines ranking than system B, then it is twice as popular when compared to individual evaluation criteria.

In order to eliminate the effects caused by the variable quantities of the data sources themselves, the popularity rating is always a relative value, which should be interpreted in comparison only with other systems. The DB-Engines ranking does not measure the number of system installations, nor their use in IT systems.

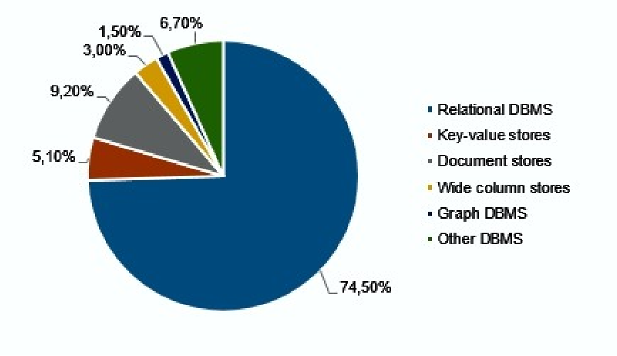


Figure 1. Popularity of database management systems, August 2020. [8]

The results of the current popularity show that the greatest interest is still in relational databases. When it comes to NoSQL databases, the results show that the most popular databases are document-oriented, followed by key-value databases, then wide column databases, while the least popular are graph databases.

Analyzing NoSQL databases from the aspect of business, it can be concluded that they represent a complete solution, when it comes to appropriate purposes. Although the exact market share and percentage of NoSQL databases cannot be determined with complete certainty, their growing financial value can be taken as one of the parameters of development. The best example is the company MongoDB (the leading NoSQL database for many years), whose stock price has been constantly growing since October 2017. It should also be noted that the main cloud service providers offer NoSQL databases, which indicates that these companies see NoSQL databases as an important commercial service. However, if we compare the popularity of NoSQL databases in the period when they expanded in 2015, Google Trends, as an imperfect measure of popularity, but perhaps the only source of data on this topic, shows that interest in them by the end of 2019 fell by more than 40% (Figure 2). What is encouraging about NoSQL databases, viewed in the previous context, is that their popularity is starting to grow again from the beginning of 2020 (Figure 3).

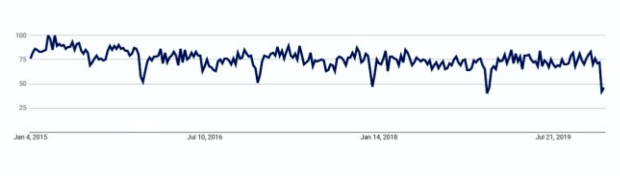


Figure 2. Level of interest for NoSQL databases in the period from 01.01.2015. to 31.12.2019. [9]

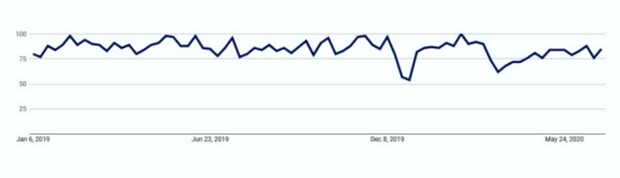


Figure 3. Level of interest in NoSQL databases in the period from 01.01.2019. to 31.06.2020. [9]

The fact that NoSQL databases are no longer such a significant topic in the IT environment does not necessarily mean that they have not become very important. In a multitude of different technologies, when many others are neglected, NoSQL databases remain quite relevant and with a prominent place in enterprise data architecture strategy.

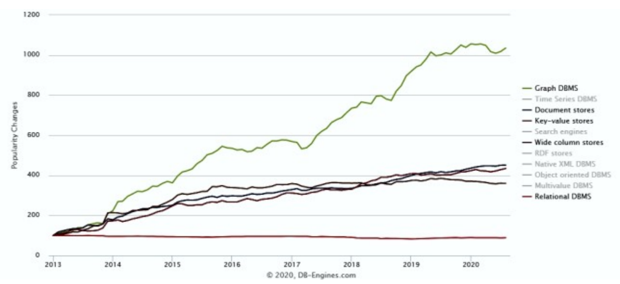


Figure 4. The trend of popularity of NoSQL systems by types and relational DBMS [8]

Based on the presented results (Figure 4), it can be concluded that graph databases have a constant trend of growing popularity, regardless of the fact that the percentage of their popularity is still the lowest compared to other types of databases. As for other types of NoSQL database management systems, with less oscillation, they have all grown in popularity over the years. When it comes to the popularity of relational database management systems, the results show that there is no trend of their growth over the years, it is even noticed that from 2019 it shows a slight decline.

Using the same ranking method, the trend of popularity of NoSQL database management systems by types is presented (Figure 5, 6, 7 and 8) in the period from 2013 to August 2020.

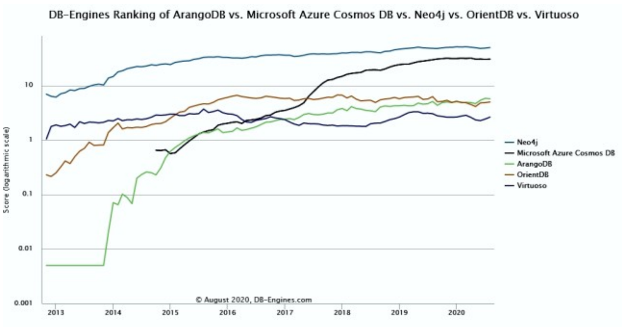


Figure 5. The trend of popularity of graph databases [8]

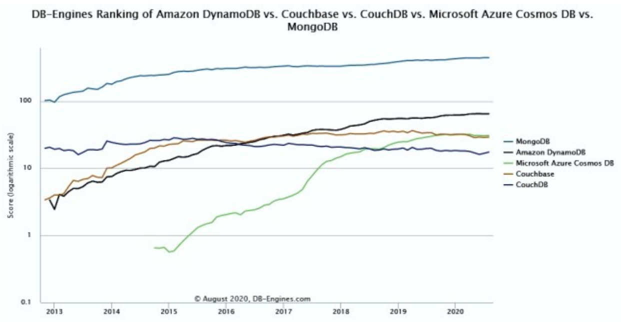


Figure 6. The trend of popularity of document-oriented databases [8]

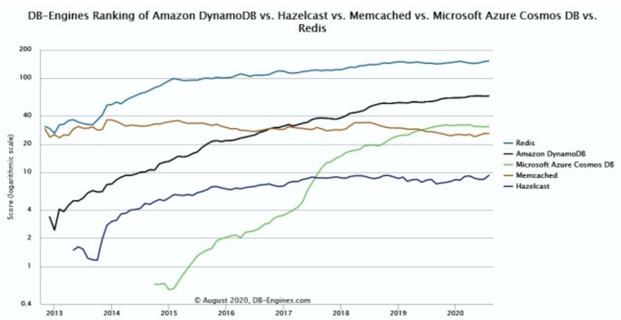


Figure 7. The trend of popularity of key-value databases [8]

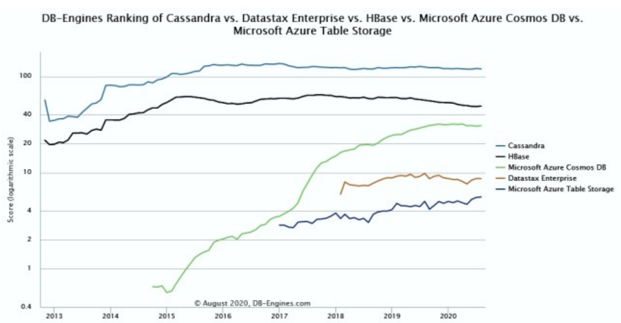


Figure 8. The popularity trend of wide column databases [8]

The five currently most popular NoSQL database management systems by individual types were selected and the trend of their popularity in the observed period was shown. It can be concluded that the characteristics and advantages of NoSQL databases, which are described in the previous parts of the paper, greatly contribute to the growth of their popularity. Also, one of the important factors in the growth of popularity and application is adaptability to modern business conditions.

B. Trends in NoSQL database management systems

A review of available scientific and professional papers dealing with NoSQL databases shows that most of them are focused on analyzing the performance of NoSQL databases or comparing the most popular ones. In the following part of the paper the key trends related to this type of database will be analyzed.

Data and information management, as well as many technologies and practices related to them (from data management to data modeling), have entered a new era. Large amounts of data have changed everything, including the growth of NoSQL platforms, which entails ways in which data will be collected, stored and analyzed in the future. In this regard, the main trends of NoSQL databases are:

- development of cloud microservices leading to many different databases;
- the growth of the Internet of Things (IoT) which pushes many companies to work with large amounts of data;
- the fact that large amounts of data (Big Data) and NoSQL databases have become something that is normal in today's IT environment.

C. Advantages and challenges of NoSQL databases

Relational databases have dominated many organizations for years. With the beginning of expertise, use, growth and need for data, there is more and more interest in NoSQL databases. The reasons for this are [10]:

- Large amounts of data (Big Data) and users of large amounts of data that require a dynamic, flexible and smaller data model. Thus, NoSQL systems meet the above requirement of large amounts of data and their users.
- NoSQL databases have the ability to grow radically to support global large amounts of data and their users.
- NoSQL databases provide improved performance without scalability to meet the expectations of users of large amounts of data.
- Security methodologies must be implemented in practice on the middleware layer by programmers who will solve NoSQL database security problems.

Unlike traditional databases, it is necessary to create strong databases without compromising scalability and performance characteristics. NoSQL has a huge growth in the future as most current applications and software are migrating to the web, and the amount of data needed for storage is growing rapidly. This means that they will face development, improvement and huge growth, and sooner or later they will surely solve their security problems. Table 4 provides an overview of the benefits of NoSQL databases [11].

TABLE IV. CURRENT STATUS: ADVANTAGES OF NOSQL DATABASES

Scaling	Horizontal data distribution between hosts
Scope	Large amounts of data that RDBMS cannot process
Administrators	They are no longer needed due to automatic maintenance
Economy	The use of low-cost servers, reducing the total cost
Flexibility	No data schema, easier design change

NoSQL databases allow developers to store and manage unstructured data and perform complex analytical operations on it. They were created in response to the growing needs of Web 2.0 applications, such as: social networks, Wikipediaes, blogs, folksonomies, online video sharing, web and mobile applications. Also, they typically rely on clusters of low-cost servers to manage rapidly growing amounts of data and transactions, allowing them to store and process larger amounts of data at a lower cost. NoSQL databases can respond successfully to changes. However, in contrast to the many advantages they have, there are challenges that these databases face and which should be seen as available opportunities to improve and further develop NoSQL databases. The current state of the challenges faced by NoSQL databases is presented in Table 5 [12, 13].

TABLE V. CURRENT STATUS: NOSQL DATABASES CHALLENGES

Maturity	Often in the pre-production phase, where some of the main characteristics are mostly missing.
Administration	Sometimes relatively difficult to install and maintain.
Analytics	Missing support for business intelligence and ad hoc queries.
No advanced expertise	There are still a small number of NoSQL experts available on the market. NoSQL databases are still considered new, so most developers are learning how to use them, but over time this situation will be resolved. However, it is easier to find an RDBMS expert than a NoSQL programmer.
Differences in the data model	Companies are struggling with a mental transition from a relational to a NoSQL data model. Projects can be created or discontinued depending on whether the team has properly modeled the data for NoSQL databases to maximize their capabilities, so database professionals need to be trained and familiar with the new NoSQL database model in the database they choose.
Distribution model	Some NoSQL databases use a master-slave architecture that can only somewhat scale read operations compared to a peer-to-peer architecture that can scale both read and write.
Insufficient security	Lack of security function in NoSQL databases.

D. The scope and future of NoSQL databases

As legacy data models, which are mostly relational in nature, are unable to fully cope with today's data needs, a new era in data science begins that leads to the rapid emergence of a wide range of non-relational data models known today as NoSQL and/or Big Data models.

The global NoSQL market is projected to reach $3.4 billion in 2020, representing a complex annual growth rate of 21% for the period 2015-2020 [14]. The expansion of the web infrastructure on a NoSQL basis has proven successful for Facebook, Digg, Twitter.

The databases of the future will be those that can be configured in such a way as to support all or most of the workloads that today require unique and separate database technologies. It can be said that the three main epochs of database technology correspond to the three main eras of computer applications: mainframe, client-server and modern web. It is therefore not surprising that the story of the evolution of modern databases follows the story of the development of the World Wide Web [15]. The dominant drivers for the latest generation of database systems are drivers driven by the demands of Web 2.0, global e-commerce, large amounts of data (Big Data), social networks, cloud computing and - increasingly - the Internet of Things.

For most businesses, Big Data, social networks, mobile devices and cloud computing are key competitive challenges. The ability to use data to create a competitive advantage is key to the survival of a modern organization's business, as is the ability to deploy applications with a global scope and with a mobile and social context. It is hard to imagine a successful modern business that did not have a strategy to use data or to engage users through social networks and mobile channels. For some industries, IoT poses a similar threat and opportunity: Internet-enabled devices are revolutionizing manufacturing, healthcare, transportation, home automation, and many other industries. Today, non-relational systems are an important and growing part of the data architecture of many companies.

IV. CONCLUSION

Although relational databases are designed for reliable transactions and ad hoc queries, they are less suitable for some types of today's applications, and the reason is their rigid scheme. NoSQL databases solve certain limitations by storing and managing data in a way that allows great flexibility and speed of work. Many were developed in companies that were looking for better ways to store and process large amounts of data. This entails that if it is necessary to process unstructured data, speed and availability are important, but not consistency, preference should be given to NoSQL databases.

NoSQL databases are primarily used in intensive read and write systems, in systems where the record structure is not known in advance, but it is flexible and adapts to the structure of records that are just being entered, as well as in systems where horizontal scalability is important. Unlike SQL databases, which are vertically scalable, which means that scaling them requires the purchase of more powerful and expensive servers, many NoSQL databases can be scaled horizontally over hundreds or thousands of cheaper servers.

From all the above, it follows that there is no single database technology that meets all needs. Different tasks require the use of both relational and non-relational databases. The integration of non-relational and relational databases is a challenge for some future research. The solution may lie in the application of NewSQL databases which represent a new approach to relational databases that want to combine the ACID guarantees of good RDBMS and the horizontal scalability of NoSQL.

REFERENCES

[1] NoSQL. Retrieved from http://nosql-database.org/

[2] Šestak, M. (2014). Graf baze podataka. Retrieved from https://bib.irb.hr/datoteka/764460.1-Martina_estak_Graf_baze_podataka.pdf

[3] NoSQL Tutorial: Learn NoSQL Features, Types, What is, Advantages. Retrieved from https://www.guru99.com/nosql-tutorial.html

[4] Krstić, L., & Krstić, M. (2018). Testing the performance of NoSQL databases via the Database Benchmark tool. Military Technical Courier, 66(3), 614-639

[5] SQL vs. NoSQL Databases: What's the Difference. Retrieved from https://www.upwork.com/hiring/data/sql-vs-nosql-databases-whats-the-difference/

[6] Selection of criteria for NoSQL database part III. Retrieved from https://www.3pillarglobal.com/insights/selection-criteria-for-nosql-database

[7] DB-Engines. (2020). Method of calculating the scores of the DB-Engines Ranking. Retrieved from https://db-engines.com/en/ranking_definition

[8] DB-Engines. (2020). DB-Engines Ranking. Retrieved from https://db-engines.com/en/ranking

[9] Google Trends. (2020). Retrieved from https://trends.google.com/trends/explore?q=nosql

[10] Bokhari, M., & Khan, A. (2016). The NoSQL Movement. Retrieved from https://pdfs.semanticscholar.org/a6f0/1c9103d3bafb8ce92641c9f2a4deaccd12f9.pdf

[11] Svoboda, M. (2018). Modern Trends. Retrieved from http://www.ksi.mff.cuni.cz/~svoboda/courses/172-B0B36DBS/lectures/Lecture-12-Modern-Trends.pdf

[12] NoSQL Databases - Overview, Types and Selection Criteria. (2019). Retrieved from https://www.xenonstack.com/blog/nosql-databases/

[13] Mehra, R. (2015). Column Based NoSQL Database, Scope and Future. Retrieved from https://ijrar.com/upload_issue/ijrar_issue_237.pdf

[14] Lodhi, N., & Babu, R. (2015). Column Based NoSQL Database, Scope and Future. Retrieved from https://www.rubrik.com/blog/enterprise-data-management-nosql-trends/https://ijrar.com/upload_issue/ijrar_issue_237.pdf

[15] Harrison, G. (2015). Next Generation Databases NoSQL, NewSQL, and Big Data. New York. Springer

Information Technologies for Supporting Research of Complex Large-Scale Resource-Driven Discrete-Event Systems Under Uncertainty

Evgeny Cherkashin[1], Qiumei Cong[2], Igor Bychkov[1], Nadezhda Nagul[1], Artem Davydov[1], Yue Wang[2], Shi Huiyuan[2]
[1] *Matrosov Institute for System Dynamics and Control Theory of SB RAS*, Irkutsk, Russia
eugeneai@icc.ru
[2] *Liaoning Shihua University, School of Information and Control Engineering,* Fushun, China
cong_0828@163.com

***Abstract*—The paper states basic ideas of an approach dealing with uncertainties in models of manufacturing process of an industrial scale. The models are complexes including discrete-event systems representing the manufacturing schedule and resource-driven game description of its behavior in a competitive multi-agent environment. The model object is a subject of control on various levels. Due to the large scale of the object, the control is synthesized as a result of a logical inference, either a theorem proving or forward-chaining modeling a computer simulation scenarios analysis. The uncertainties appear in the model due to activities of competitors or internal factors having probabilistic nature and represented as internal events or shortage of resources. The synthesized control should ensure that the system corresponds to a set of predefined constraints and/or maximizing/minimizing values of a set of characteristics.**

The aim of the research is to implement computer software platform for visual design of industrial manufacturing processes, defining their environments, internal uncertainties event generator rules, and constraints, realizing control syntheses, as well as model assessment criteria to compare the quality of various scenarios. The software could be applied to the wide range of domain problem solved in, *e.g.*, computer networks; resource allocation in computer cloud systems; automated manufacturing; air traffic control; robotic assembly lines; highly integrated command, control, communication, and information systems; control of formations of underwater unattended vehicles.

I. Introduction

Control of industrial processes is on the top of scientific interest since first automation was introduced to manufacturing goods or supply utilities. The evolution of automation in the last decades was an adoption of various technologies in industry to achieve (1980s) better production quality thanks to improved management, (1990s) better processes organization (reengineering of business processes), (2000s) collaborative manufacturing and management the process between counterparties. With the invention of computers, integrated circuits and communications, information and network communication technology are closely combined with industrial process. Individual manufacturing and network-computational systems now constitute industrial systems, which have improved overall operation efficiency and reduced consumption of raw materials. Due to the large scale of networked systems, it is difficult to realize the traditional centralized control. Therefore, the nowadays control structures of large-scale networked industrial processes is decentralized and distributed. Decentralized control is simpler in structure and more convenient in implementation, but it does not deal with the physical coupling between subsystems. Distributed control deals with coupling relationship through communication between subsystems, the quality of the communication implies the quality of control.

Representing industrial system as discrete-event system (DES), as some other technical and natural objects, consumption, production and concurrent temporal acquiring resources are to be accounted as a significant part of the models. Some examples of the similar kind are computer networks; resource allocation in computer cloud systems; automated manufacturing; air traffic control; robotic assembly lines; highly integrated command, control, communication, and information systems, *e.g.*, groups of underwater unattended vehicles. Uncertainties of the resource volume and the requirements are the parts of the discrete-event system model as its consumption/production cannot be modeled completely accounting all possible factors. In this case the model becomes too complex to evaluate and simulate. Decomposing the sources of uncertainties by the relevance to an industrial system resulted in two kinds: external and internal. External uncertainties are represented with game theory, as events of DES are generated by the environment of the industrial system. This representation will model the behavior of external agents (*e.g.* firms, teams, or individuals) having different competition and collaboration strategies.

The game theory has a direction which considers objects, which agents (players) have above mentioned properties of dealing with resources (consumption, production, locking, *etc.*). At each action, an agent in a general case consumes a resource (money, matter, energy, *etc.*), produces a resource (*e.g.*, commodities, energy, service). Action also can block activities of other agents, preventing a resource consumption/production, stealing a resource, *etc.* There are resources, which are accessible only by limited number of agents, *i.e.* a resource can be locked. These games are referred to as *resource-based games*. A good short introduction has been published recently in [1].

Majority of the research considers the game model to be controlled supervisory ("from top"), *i.e.*, the competing opposites are supposed to be able to negotiate according to the previously synthesized Nash equilibrium. But in practice, it is not possible due to high risk of cheating and complexity arising in accounting all the players' requirements. In the distributed decentralized system, the system have to be considered as being influenced by competitors outside, generating events, and the system reacts on these events. The aim of control is to support systems sustainability (resilience), optimizing its output characteristics. Supervisor control shifts to the level of the agent from the level of the system of competing agents.

The research plan stated in this paper is aimed at investigation of the fundamental problem of development of methods for control synthesis for complex (large-scale) resource-driven DES with uncertainties (RDDESU), which generalize various practical domains, including manufacturing process consisting of planning, scheduling (execution of a planned batch processing) and dynamic control of the schedule implementation. Manufacturing commodities and service quality is a basis of an enterprise existence, including its reputation and financial well-being. From the consumer point of view, the better manufacturing produces better commodities in time and in a sufficient amounts fulfilling demand.

The development of new methods of intelligent control synthesis for resource-driven DES with uncertainties is based on

1) new declarative means of complex DES description in aspects of: event occurrence, state change, resource consumption/production and acquiring, competitor behavior, functioning constraints, as well as computer simulation parameter definition;
2) devising approaches to visual definition of part of the declarative model with standard modelling notation, *e.g.*, UML, SysML, BPMN, CMMN;
3) development of methods for building modular supervisor controllers for RDDESU by means of logical inference, using its descriptive means and generally modified strategies;
4) adaptation of existing logical inference approaches to support resource-driven games and corresponding data structures and operations over them;
5) devising software for DES description and its computer simulation, supporting generation and evaluation of various scenarios.
6) development of the technique for representing near-to-real manufacturing processes as RDDESU;
7) specialization and testing the methods and software under development in the field of manufacturing control.

The main advantage of our approach in comparison to the popular ones is that our method will try to synthesize a control structure (an algorithm or a list of concrete events in the worst case), which will react on occurring events bringing the system in a proper state, if possible, or give the set of additional conditions, fulfilling which the control synthesis could be possible. In the same time the practical requirements imposed development of means for complex DES representation, namely, data structures, visualization, automatic (unattended) intelligent methods and software of control synthesis.

II. A General Idea

RDDESU represents technical, natural and social objects (agents) in its environment of competitors, where all competing agents can consume/produce or acquire/release a resource at each externally or internally generated event changing system state. Such systems are referred to as *sociotechnological systems* (STS). The uncertainties appear due to activities of competitors or internal factors having probabilistic nature and represented as internal events or shortage of resources. The synthesized control should ensure that the system corresponds to a set of predefined constraints and/or maximizing/minimizing values of system characteristics. From control theory point of view, the problem is so called poorly stated, as criteria of optimality cannot be defined (formalized) completely.

Conflicts in STS can be analyzed by construction of a finite chain of sequentially made possible actions (mutual impacts) of opponents [1]. This formulation of the problem is the object of game theory (GT) of modern operational research, and the object is referred to as *resource-based conflict* (RBC), and the resources the agents own are their *resource bases* (RB). According to GT, a sequence of actions (a strategy) has to be determined such that it will lead to the end of the conflict (*i.e.* final state of RB of the players), which meets goals of both or at least one of the opponents, who is declared winner.

Despite widespread and commonly understood essence of the described RBC, GT state of the art is such that available mathematical tools used nowadays by GT developers are not exactly suitable for completely adequate mathematical formalization of conflicts because application of these tools demands explicit representation of sets of possible implementations of the conflicts in a form of trees, matrices or automata. By this dimensionality of practically interest representations of RBC are so large, and time, necessary for their construction, so long, that in most cases it is very hard or even impossible to apply known GT approaches to the adequate modeling of real RBC [1].

Alternative approach, which consideration in the context of RBC may have evident reasons, have raised from the artificial intelligence area. It differs from GT by basic representation of the potential activities of the opposing sides and possible ways of application of their capabilities during conflict. This approach is called "multi-agent", following that every STS is represented as a set of active entities, called agents, operating independently in accordance with their internal logic. Agents form multi-agent system (MAS). In comparison with modern GT models, MAS do not demand explicit representation of action of the opposing sides during conflict. It is sufficient to define implicit logics of action of all agents, comprising MAS, and to construct the initial state of the MAS, *i.e.* set of initial states of its agents. All further behavior of the modeled STS until some final state will be explicated by MAS itself [1].

Analysis of the manufacturing process modeling domain, the primary scientific interest of LNHU (Liaoning Shihua University, Fushun, China), showed that the object instances are represented as STS. The accuracy of the controlled processes represented as RDDESUs ensuring the prescribed parameters properties in the conflicting environment and internal properties causing uncertainties is the subject of LNHU research. Internal uncertainties are mostly the shortage of resources, namely utilities, absence of staff due to illness, equipment fault. These problems are solved with reserving power, employees, rescheduling batch processing to the equipment used on less prioritized processes. The amount of the reservation and a set of other activities aimed at preventing occurring internal uncertainties and overcome them is figured out by a computer simulation.

Thus, an upgrade of well-established approaches for industrial systems manufacturing processes modeling with game theory and DSS representation is being proposed, the unknown parameters of the model are figured out in a computer simulation based on control synthesis algorithms for a corresponding DSS. The simulation processes are organized as theories of first-order logical language of Positively constructed formulas (PCFs) and its theorem proving techniques as the scales of the industrial systems require automating high-level activities of the object analysis process.

III. Related Works

Currently, there are almost no publication on RDDESU due to the problem complexity related to the representation of the object (system) state, resource conversion and lock, uncertainties modeling. Various aspects of the stated problem, however, are in active research.

Modeling the competitive environment is relatively new in the field of research of industrial systems. The environment is described as set of possible events, which can occur, the possible actions are the way of reacting these events [2], [3]. In the paper, the solution is designed as set of rules for reacting to events. Reactions are constructed for both competitors and non-competing counterparties (suppliers, consumers). After that, the amount of additional resources to be spent to implement the reaction is accounted. The similar technique is used to assess possible risks of startups.

Most works do not cover the organization behavior in a perspective more than a couple of possible steps. This is probably due to the complexity of its theoretical representation. In the 2000s, a new direction of application of GT with usage of classical control methods appears, namely, *mean field games* models (see a recent review [4]). The theory started to attract scientific attention after publication of J. M. Lasry and P. L. Lions (2006), and, in the same time, of P. Caines, Minyi Huang and Roland Malhamé (2007). The theory allows one to apply regular approaches of continuous mathematic to figuring out a Nash equilibrium, resulting in less combinatorial and complex computation. But the problem state is based on the subsumption that there exists infinite number of rational opposing agents to the main player. In the environment, an agent activity influence to main player is infinitely small and, hence, the influence of the opposites are accounted aggregated and one can even use iterative techniques efficiently to figure out fixed point estimates. The main application areas are modeling crowd behavior, energy GRID modeling, market price formation.

Digitization of the business processes management supply chains, namely information technology support, resulted in preferable discrete representation of the manufacturing process. Discrete-event systems and their derivatives are the mathematical abstraction widely used in representation of the manufacturing processes of all types (continuous, discrete and batch). Most attention is paid to the local processes (within a company). The influence of the competing organizations in these models are presented as uncertainties and accounted by constructing a *predictive control*. Predictive control is a model-based optimization control approach. The system model is used to predict the future state of the system. One of the approaches that uses system state forecasting is the *receding horizon* optimization strategy improving the control inputs of the system. In the paper [5], the physical coupling among subsystems are considered, and the communication network is utilized to obtain Nash equilibrium solution for the optimization of each subsystem output through repeated iteration of coupling among subsystems, with the constraints being satisfied; in that case the system performance is optimized. Based on the previous paper [5] and paper [6], authors perform the nominal stability and performance analysis, and verify the distributed predictive control algorithm on a benchmark process (heavy oil fractionation process). The above method belongs to a class of non-cooperative distributed predictive control.

When the real process is strongly coupled, the stability and control performance of the system will be greatly affected by the local performance indexes obtained by the non-cooperative distributed predictive control. Therefore, researchers put forward the cooperative distributed predictive control method. Compared with non-cooperative distributed predictive control, the former takes the global performance index of the system as a function of each subsystem's performance indexes, allowing calculation the optimal global performance index and improvement of the convergence speed while considering the subsystem's performance. Liu Yubo *et al.* [7] proposed a distributed predictive control method based on global coordination, the correlation of input and output of all associated subsystems are comprehensively considered and verified it in Shell's heavy oil fractionator and other process. Venkat *et al.* [8] used game theory to study the optimality of a distributed predictive control and concluded that the iterative solution has the global optimality, but the computation is complex.

Aiming at the predictive control method with limited resources, Asadi *et al.* [9] proposed the distributed coordinated predictive control method under the resource limitation and adopted the time-division and multiple-access technology to coordinate the sub-system. Zou *et al.* [10] proposed an event-driven distributed predictive control method, which implements an asynchronous coordination for each agent (subsystem) and reduced the computation and communication burden. Li *et al.* [11] adopted the non-zero-sum game reinforcement learning method to obtain the optimal solution of each agent iteratively and each subsystem is coordinated and controlled.

In the system with structural uncertainty and parameter uncertainty, Li *et al.* [12] proposed a time-varying multi-step control rate algorithm in the distributed predictive control method, in which a "multi-step control set" is designed, and the degree of freedom of the controller is increased. For the system with the bound change of the parameters, the performance of predictive control is further improved by combining the multi-step control method. For dealing with uncertainties of the resource allocation and the demand, all achievements focus on the production planning and scheduling stage and the existing distributed predictive control algorithm has not been involved.

IV. The Proposed Approach

As stated in [13] the uncertainties is the main point of reduction of the model ability to make precise forecast and reveal real properties of the consumer products manufacturing and service provision. In order to minimize impact of the uncertainties model must account as much information as possible, as well as the object should have the ability to adopt itself to perturbation (a property of *resilience*). We divided the sources (see section II) of the uncertainties as external, a competitive multi-agent environment and internal, which arise at industry site. In contrast to the nowadays fashionable approach of complex systems modeling based on machine learning techniques, classical mathematical methods of control theory for discrete models will be used with advantage of contemporary information technologies (IT) supporting automated research based on computer simulation, logical inference (automatic theorem proving), and application to physical and cyber-physical units. The control synthesis will be realized as a result of logical inference: back-propagation theorem proving and forward-chaining deduction. The first one is used to synthesize a modular supervisor control if the logical description formula has a proof, or show the system's properties preventing the synthesis. The second one is a computer simulation controlled by forward-chaining logical inference, where the control is synthesized as sequence of actions shifting system to a admissible state after an undesired event occurrence; the events occurrence are modeled with rules.

Computer simulation and control synthesis are based also on novel application of the authors' automatic theorem proving methods, previously used in construction of the complexes of models, their identification, and driving the computer simulation. The authors' PCF approach is to be extended with new techniques and modifiers of the default logical inference strategy supporting the implemented simulation techniques. So, the new calculi with special, including intuitionistic properties, will be devised. This should enrich existing techniques, which already support parallel processing; as compared to the Resolution method, the approach has less combinatorial inference search process, weak requirements of the formula structure, it is highly configurable and compatible with heuristics.

In comparison with Prolog and its constructive notation, PCFs have intuitionistic semantics in a wider range of rule representation, non-Horn ones. The PCF language expresses logical inference as a question-answering procedure, which looks like application rules in forward-chaining inferences (CLIPS, DROOLS), but supports disjunctive branching in the consequent part, which is naturally expresses variants of system evolution in a virtual future. The resulting inferences can have intuitionistic semantic if some general constraints on the whole formula are hold. These constraints are more general, than in Prolog-like systems. Another useful features of the PCFs inference is the ability to adjust to the properties of the domain. The default complete strategy is modified by means of constraints and guides (heuristics), which are added to the default search strategy, this does not force the user to program new one.

The proposed research is based on the complex representation of a large-scale sociotechnical system as RDDESU, accounting more information for the object representation as compared to the approaches based on GT and DES, resulting in more adequate real technical systems modeling with its environment, usually competitive for resources. The competitive environment generating external uncertainties is represented with RDG, namely, mathematical structures of multisets, operations over them, filter, and the rules of the behavior definitions. All these mathematical structures are compatible with DES representations. Development of data structures for representation of RDG as DES is a new problem in IT, as well as the amounts of structural and behavioral data of modeled object force the research to use intelligent and visual technologies to represent the object and convert its representation into data to be processed with computer simulation and control synthesis software.

The research will deal mainly with the development of techniques for

1) definition the model structure of the RDDESU and their environments (further, just RDDESU),
2) adaptation of the standard visual notations for representation of the RDDESU,
3) construction of transformation software converting the visual representation of a RDDESU and other model aspects into a computer simulation object,
4) implementation of control synthesis techniques for the RDDESU,
5) development of new modifiers of the default inference engine to support computer simulation for RDDESU,
6) adaptation of three-level dynamic model of manufacturing as RDDESU,
7) implementation of scientific software for RDDESU research.

Consider methods and approaches main items of the list in details.

A. Resource-driven games representation of an industrial system

A RD DES with uncertainties are represented as competing agents spending/producing or acquiring/releasing resources. In manufacturing industry, resources are raw materials, commodities, services and utilities (energy, gas). The competitive environment comprises such kind of agent, which activity always affect to other agents' resource production and consumption by eliminating an amount of resource or suppressing the activity altogether that is represented as zero productivity and consumption. The activity of an agent can also produce some amount of resource at competing agent side.

Mathematical representation of the resource-driven games is based on usage of multisets, where each item of a multiset is an amount of a resource representation (<resource-kind, amount>). A multiset represents the set of resources owned by an agent at a time moment. A number of operators (intersection, union, addition, subtraction, transformations), filters and constraints are defined to express the resource possible ways of conversion within an action. The conversion is expressed, *e.g.*, with a grammar, where each of the grammar rule represents a conversion act. All the possible grammar inferences results in multisets of various consequences. Over these sets, similarly, operations, filters and constraints can be also defined. Filtering the resulting multisets produces various model forecasting or property evaluations [1].

The approach that uses simpler multi-agent model was applied in distribution of the load in a heterogeneous distributed computing environment [14]. A computational process is descried as an abstract program implementing within a problem-solving scheme. The scheme represents the applied modules to be run and their relationships. A multi-agent algorithm for distribution of the modules between environment nodes realizing a possible computation instance of the scheme has been developed. The resource reallocation is started as soon as an existing computation process fails. The instance plan is generated as a result of tender of jobs, allocation of a core class to the instance costs a resource. The bidding data are collected in the run time accounting various kind of failures. In comparison to the known algorithms, the proposed one implements an adaptive multi-scenario solving this issue and therefore increases a degree of computational process fault-tolerance. The further direction of the study is related to using this algorithm in complex technical systems with decentralized group control of problem-solving processes in the conditions of conflict and cooperation.

B. System modeling

DES aspect of a RDDESU is represented with usual means, such as, automata, which is systems of state conversions defined as functions. Uncertainties are expressed as possible competing agent actions and probabilistic events occurring internally in the system. Real-world problems (RWP) usually have large scale as compared to theoretical testing examples. The structures of the RWP representation are complex and comprises various aspects. A popular way of overcoming the complexity of object representation is usage of visual models, where various aspects are represented with corresponding visual images. In the IT field, standard ways of the representation are UML, SysML, BPMN and other notations. These notations can be used for RDDESU representation, especially SysML, which domains are description of complex systems, including software, organizations, processes etc. For each of the model notation a visual editors and export/import formats (such as XMI), usage techniques were designed. In our project we will use visual notations and their tools, with conversion of the models into internal representation.

In order to proceed with computer simulation and other object synthesis, the model of RDDESU must be represented as internal objects, which is essentially a PCF formula defining RDDESU. This object is constructed from the set of visual models using SysML diagrams (State diagrams, Parametric diagram, Requirement diagram, *etc.*) by models' transformation. These models in form of XNI-files are loaded in an object-logical environment constructed as network of objects of the Logtalk logical Prolog based language [15].

The standard XMI representation of the visual models is transformed into set of triples `<subject, predicate, object>` as an internal representation of input models [16]. This representation provides uniform access to all model structures of all models simultaneously. Logtalk objects analyze structure of models and other objects producing intermediate structures representing the structure of modules to be generated as source code. The transformation obeys a scenario of software construction in the image and likeness of Model Driven Architecture paradigm. Usage of such kind of transformation approach allows one to develop software system carcasses on the level of abstract models, involve various sources of model data in the transformation, define and structuring conversion knowledge as objects.

The approach was tested in representation of 144 Mothur modules (an applied package for genetic sequencing analysis) as dataflow block of Rapidminer studio, allowing natural scientists to operate the process of genetic investigations themselves. The technique will be used to construct transformation of SysML representation of RDDESU into an internal representation. The usage of the logic object-oriented language Logtalk in the model transformation instead of standard ATL, QVT approaches allows one to use both power of logical inference and the expressiveness of object-oriented programming systems, as well as the integration with existing libraries in within the same programming system. For example, inheritance in a prototype hierarchy allows programmer to change rule sets, encapsulation allows simplification of access to input data, recognizing patterns in input data on-the-fly.

C. Control synthesis

Control for RDDESU is synthesized either as supervisor, preventing occurrence of events, or as a set of trajectories of system's evolution constrained with a time interval (set of ordered time moments), where each time moment corresponds to an admissible state of the system.

For most common and convenient way to represent a DES, finite state automata, results of P. Ramage and W. Wonham [17] are applied. The main approach to decentralized control synthesis on the basis of the theory of supervisory control is the conjunctive architecture of local supervisors when global permitted events are the only ones which were allowed by each of the local supervisors. Most of the results of decentralized control are associated with this approach [17] and a generalized approach, so-called general architecture of a decentralized supervisor [17].

According to the synthesized control, some events of automata-based DES are supposed to be prohibited from occurring to restrict system behavior within constraints given by some specification. The means of such control is the supervisor. The problem of designing a proper supervisor for fully or partially observed DES is the main problem of supervisory control theory (SCT). Controllability of a specification is a crucial property which must be verified to start building a supervisor. It may be compared to the safety property of the controlled system: no uncontrolled event which is not allowed by specification may occur. In the case when the input RDDESU specification is not controllable, a controllable sublanguage (subautomaton) of this specification is to be found as a new admissible specification. In the case of manufacturing modelling, the sublanguage denote a set of events, which appearance can be controlled.

Modular supervisors for DSS have been constructed for control of on a top level and AUV (autonomous underwater vehicles) group with automatic theorem proving on the base of PCFs. An inference result of theorems represented with PCFs with subformulas corresponding to the controlled system and the requirements, which are to be hold, a conclusion is being drawn on the possibility to construct a corresponding supervisor control, or, in case of the impossibility of inference construction, a set of the system parameters are given, which prevent the inference. The obtained data is useful as heuristic for reduction of the number subsystem combinations needed in the modular supervisor synthesis [18], [19].

DES is a wide class of behavior representation of a dynamical systems changing state at a discrete time moments thanks to event occurring. In this direction, last decade we obtained results on decentralized hierarchical control synthesis for groups of autonomous underwater vehicles [20]. The number of the trajectories can be constructed, *e.g.*, movement of the group in a conserved formation along a prescribed trajectory, patrolling a water area with gateways. The control is synthesized accounting the uncertainties of the environment, non-ideal functioning the communication and measurement equipment, discretization of the measurement (sampling) and constraints imposed to control resources.

In Russian Science Foundation project 16-11-00053, a logical programming like language grammar for PCFs was developed to describe team missions of robots [21] and synthesize

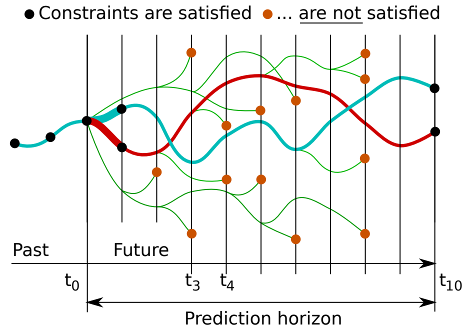


Fig. 1. Predictive horizon based computer simulation

a control to achieve the mission goals. For the purpose, a PCFs language has been developed containing special system predicates and so-called computed function (functions, which are substituted as its calculated value instead its expression), allowing one to control default inference strategy and construct resulting terms. The forward-chaining inference of the PCFs calculus is being adapted to real time computational systems. The system constructs the inference till a timeout is occured. If the inference corresponds to a set of criteria, it is considered as successful. If a set of results are infered, they are analyzed and the best one is chosen. Other original features of PCFs are investigated and adapted for solving other domain IT (real time inderence) and technical problems.

An opposite approach is the compensation of the event appearance consequences, shifting the system in an admissible state with spending a reserved resource. This can be achieved with a set of rules recognizing the unsatisfactory state, analyze its properties, and add new actions to the schedule to move the system to an admissible state. In a general case, the admissible state is defined as a set of constraints. It is very hard to construct a complete set of rules for recognition and a transition synthesis for the domain, which is under investigation. In this case, a set of actions are performed in the same time moment splitting the current scenario to a set of subscenarios (alternative virtual futures) for each possible action. After each splitting and transition set of constraints are checked. The scenarios, which do not comply the constraints are rejected. The trajectory, which reaches the modeling horizon corresponds to the synthesized control driving the system in an intermediate, but admissible state (Fig. 1). When a bunch of alternative admissible futures were synthesized, the choice of the best one is carried on with assessing criteria, *e.g.*, minimizing additional resource consumption. Thus, the control is synthesized by means of logic controlled computer simulation, and respects some ideas presented in section III.

In order to implement the computer simulation on PCFs representation of RDDESU, data structures for multisets and a default strategy modification are to be designed, and operations within construction of logical inferences are to be implemented. The latest version of our automatic theorem proving (ATP) software (provers) have RAM (Random Allocated Memory) allocation engines efficiently implementing indexing over the sets of terms. Filter operations for multisets have the similar specifications. Items of the multisets are similar to term and atom in sense of data representation. Special subroutines will be programmed to account specific properties of multisets, their elements and relations, making multiset manipulations memory and time efficient.

The inference engine is to be adjusted with new modifiers and guides to support time and RAM efficient application of operations, filters and constraints as well as representation of grammars over the multisets. The default strategy can be modified with adding each rule (a question) a priority, arranging the rules in an order of application, restricting application of a rule when a pattern is recognized, backtracking the inference in case if criteria have not met within a time interval or a number of the inference rule application, checking constraints to the indexed subsets on the stage of substitution search.

V. Future Work

Expected results and their practical values are expected to be as follows:

1) A technique for representation of various aspects of RDDESU as a complex of models will be developed, data structures for representation of the complex will be proposed.
2) For a visual representation of the aspects, a corresponding standard visual representation (*e.g.* within UML, SysML, CMMN notations) and its editing software will be adopted, and a usage technique will be proposed.
3) A software for a) transformation of the visual models to the internal representation will be realized as structured logical objects, encapsulating transformational knowledge and algorithms; b) methods of building supervisors for the DES will be developed based on the use of means of automatic proving theorems in PCF-calculus (calculus of positively-constructed formulas).
4) The authors' technique of constructive inferences constructions on the base of PCFs will be improved to support multisets and operations over them as first class objects, being subject of the inference, as well as new modifications of the default strategy of the inference search will be obtained, making the inference be more productive; a parallel versions of new strategy will be devised for cluster computing systems.
5) Methods, developed in the project will be organized as a modular software for carrying on scientific research (scenario generation, computer simulation and criteria assessment) in the field of RDDESU, intelligent user interfaces for process control will be devised, as well as various means of the result representations.
6) For application aspect of the research, techniques of the near-to-real manufacturing process formulation as complex of visual models will be developed, examples of the formulation will be presented and their control synthesized.
7) An extensive testing of the synthesized control on testing and near-to-real examples will be carried out to assess the performance and other characteristics with the competing approaches.

In parallel, the following investigations will be carried on in LNHU. The work is related to the adaptation the RDDESU approach to the real-world classes of manufacturing process.

1) Research of resource-driven multi-objective optimization method based on the resource-driven game model, a new efficient method for obtaining the optimal set point (the optimal solution) and the coordinating multi-objective strategy will be developed.
2) Research of distributed predictive control for RDDESU with regards of industrial scale and complexity.
3) The establishment of visualization model of typical industrial process of ethylene and polypropylene production.
4) Application of the proposed control algorithm of RDDESU to typical petrochemical processes accounting the uncertainties.

The expected results will discover new opportunities for rising level of commodities batch manufacturing control, as well as service. The project expands the theoretical knowledge in DES, GT, ATP, constraint programming, which are the fields of artificial intelligence, as well as manufacturing modeling. The produced software could be used to construct CAE/CAM (Computer-aided engineering/Computer-aided manufacturing) systems for concrete enterprises.

VI. Conclusion

The problem of construction of an automated research software for investigating properties and control synthesis for a complex large-scale industrial dynamical system is considered in the paper. The control must hold the system within prescribed set of constraints describing admissible states, while the system is affected by uncertainties of various kind. A system model general design is proposed, where the uncertainties are divided onto external, caused by competitors and counterparties, and internal occurring at place being result of utility shortage, equipment failure, staff absence. The theories of Discrete-event systems and Resource-based games are used in the representation of the industrial system batch manufacturing processes, which are a general case.

The following results are to be obtained: 1) a resource-driven game based model of a manufacturing process accounting competitive environment of an enterprise will be developed, 2) techniques of control synthesizing minimizing deviations of the process execution due to perturbation, 3) new strategies of inference search for positively constructed formulae, 4) new formalized knowledge for model conversion and interpretation represented as objects, 5) computer simulation and visualization system implementing the above mentioned methods and techniques. The scientific novelty is justified the usage of data close to real, application of the transformational technologies in software that allow one to deal with more complex models and account more data. The theoretical results could be used to construct decision support systems for engineering manufacturing processes, assessment of the robustness of the synthesized control.

The collective has a certain foundation on the control synthesis, modeling and optimization of production scheduling and planning, description and management method for uncertainties occurring in industrial process. For a concrete manufacturing process, techniques for formulation of the production scheduling model and dealing with the internal uncertain factors of manufacturing process based on the chance constraint programming and fuzzy theory [13] have been developed. For the typical concrete manufacturing process with multivariate, strong coupling, large time delay and multi-time scale, a set of systematic and practical optimization and control schemes are proposed by the two-layer way, in which the upper layer adopts the set-point optimization method with the slow sampling time and the down layer adopts the distributed predictive control with the fast sampling time. In view of this, some factors such as the change of the raw material, the lifetime drop of the total load, external environment and the associated coupling among the sub-systems that influence the control quality of the system in industrial control can be overcomed, the control performance can be improved, the product quality and economic benefit can be increased.

The enrichment of the two-level standardized model of manufacturing process with resource-driven representation and parts simulating internal disturbances will add a third level of execution dynamics under a control. The manufacturing process naturally is being under impact of various uncertainties (perturbations) such as delay of resource supply, the equipment breakdown, temporal lack of labor, sudden production need, requiring equipment occupation on the urgent tasks. These uncertainties are modeled either “in overage”, *e.g.*, as probability to finish production on time, or an average resource loss (see section III). We suggest rectify these approaches with synthesis (on the third level) of an optimal control supervisor automata preventing some events to happen, or as a result of computer simulation in discrete time. The control is synthesized as a sequence of actions of a trajectory holding imposed constraints in virtual timescale during a significant time period.

Acknowlegements

The research is being supported by Russian science foundation, grant No. 21-41-00022.

References

[1] I. Sheremet. “Resource-based games,” CEUR-WS proceedings, Vol. 2638, Proceedings of the 2nd International Workshop on Information, Computation, and Control Systems for Distributed Environments, Irkutsk, Russia, July 6-7, 2020, pp. 234-251.

[2] A. A. Osorio-Londoño, J. C. Naranjo-Valencia, G. Calderón-Hernández, “Training and its influence on competitive strategy implementation,” Human Resource Development Quarterly, 31, 2, 2019, p 149-172. DOI:10.1002/hrdq.21381

[3] M. U.Ahmed, M. M. Kristal, M. Pagell, T. F. Gattiker. “Building high performance supply-chain relationships for dynamic environments”, Business Process Management Journal, Vol. 26, No. 1, 2019. p. 80-101. DOI:10.1108/BPMJ-05-2018-0139

[4] D. A. Gomes, J. Saúde, “Mean Field Games Models—A Brief Survey,” Dyn. Games Appl. 4, 2014, p. 110–154. DOI:10.1007/s13235-013-0099-2

[5] X. Du, Y. Xi, S. Li. “Distributed Model Predictive Control for Large-scale Systems,” American Control Conference, 2001, Proceedings of the 2001, IEEE, 2001, 4, p. 3142-3143.

[6] S. Li, Y. Zhang, Q. Zhu. “Nash-optimization Enhanced Distributed Model Predictive Control Applied to the Shell Benchmark Problem,” Information Sciences, 2005, 170(2), p. 329-349.

[7] Y. Liu, X. Luo, F. Xu. “Global Coordination and Stability Analysis for Distributed Model Predictive Control System.” Journal of Chemical Industry, 2013, 64(4), p. 1318-1331.

[8] A. N. Venkat, J. B. Rawlings, S. J. Wright. “Stability and Optimality of Distributed Model Predictive Control,” In: Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference. Seville, Spain: IEEE, 2005, p. 6680-6685.

[9] E. Asadi, A. Richards. “Scalable Distributed Model Predictive Control for Constrained Systems,” Automatica, 2018, 93, p. 407-414.

[10] Y. Y. Zou, X. Su, S. Y. Li, Y. G. Niu, et al. “Event-triggered Distributed Predictive Control for Asynchronous Coordination of Multi-agent Systems,” Automatica, 2019, 99, p. 92-98.

[11] J. N. Li, J. L. Ding, T. Y. Chai, C. Li, L. L. Frank. “Nonzero-sum Game Reinforcement Learning for Performance Optimization in Large-scale Industrial Processes,” IEEE Transactions on Cybernetics, 2019, Early Access Article.

[12] L. Dewei, X. Yugeng, Zh. Pengyuan. “Constrained Robust Feedback Model Predictive Control for Uncertain Systems With Polytopic Description,” International Journal of Control, 2009, 82(7), p. 1267-1274.

[13] Y. Wang, X. Jin, L. Xie, Y.Zhang, S. Lu. “Uncertain Production Scheduling Based on Fuzzy Theory Considering Utility and Production Rate,” Information 2017, 8, 158.

[14] A. Feoktistov, R. Kostromin, I. Sidorov, S. Gorsky, G. Oparin. “Multi-agent Algorithm for Re-allocating Grid-resources and Improving Fault-tolerance of Problem-solving Processes”, Procedia Computer Science, Vol. 150, 2019, pp. 171-178. DOI:10.1016/j.procs.2019.02.034

[15] E. Cherkashin, A. Shigarov, V. Paramonov. “Representation of MDA Transformation with Logical Objects,” 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, Russia, 2019, p. 0913-0918. DOI:10.1109/SIBIRCON48586.2019.8958008

[16] E. Cherkashin, I. Terehin, V. Paramonov, V. Tertychniy, “New transformation approach for Model Driven Architecture,” Proceedings of the 35th International Convention MIPRO, Opatija, 2012, p. 1082-1087.

[17] T.-S. Yoo, S. Lafortune. “A General Architecture for Decentralized Supervisory Control of Discrete-Event Systems,” Discrete Event Dynamic Systems. 2002. Vol. 12, Iss. 3. p. 335-377. DOI:10.1023/A:1015625600613

[18] A. Davydov, A. Larionov, N. Nagul. “Logic Inference Based Construction of a Supervisor for a Discrete Event System,” CEUR-WS proceedings, Vol. 2638, Proceedings of the 2nd International Workshop on Information, Computation, and Control Systems for Distributed Environments, Irkutsk, Russia, July 6-7, 2020, p. 53-67.

[19] A. Davydov, A. Larionov, N. Nagul. “The Construction of Controllable Sublanguage of Specification for DES via PCFs Based Inference,” CEUR-WS proceedings, Vol. 2638, Proceedings of the 2nd International Workshop on Information, Computation, and Control Systems for Distributed Environments, Irkutsk, Russia, July 6-7, 2020, p. 68-78.

[20] I. Bychkov, A. Davydov, M. Kenzin, N. Maksimkin, N. Nagul and S. Ul’yanov, “Hierarchical Control System Design Problems for Multiple Autonomous Underwater Vehicles,” 2019 International Siberian Conference on Control and Communications (SIBCON), Tomsk, Russia, 2019, p. 1-6, DOI:10.1109/SIBCON.2019.8729592

[21] A. Davydov, A. Larionov, N. Nagul, “PCF-based Formalization of the Parallel Composition of Automata,” Proceedings of the 1st International Workshop on Information, Computation, and Control Systems for Distributed Environments, ICCS-DE 2019, Irkutsk, Russia, July 8-9, 2019. CEUR Workshop Proceedings. p. 29-41.

[22] E. Cherkashin, S. Badmatsyrenova, A. Popova, I. Vladimirov. “An Optimal Control Module of Sustainable Natural Resources Consumption Control Synthesis for Decision Support Systems,” Proc. of 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2014 - Proceedings, 2014, p. 1100-1105.

Knowledge as a product

Goran Sučić
University of Split / Faculty of Humanities and Social Sciences
goran.sucic@ffst.hr

***Abstract* - The paper compiles and consolidates an analysis of investment in the development of education in the last 20 years. Comparison with the results conducted by the Institute for Social Research in Zagreb and the Center for Research and Development of Education 2005/6., and the results of research that I conducted in 2010 on a small sample and the comparison with the results of research in 2020. The aim of research was to examine the opinion of teachers and students about current curriculum, and in this research relevant data refer to the use of IT techno-logy by teachers and students in educational process.**

Research conducted in 2020. on a sample of 860 teachers and students in primary schools and high schools in counties Splitsko-dalmatinska, Dubrovačko-neretvanska, Šibensko-kninska and Zadarska, and students of the Faculty of Philosophy in Split who gave answers to the questions whose results will simultaneously indicate the direction of change and the need for faster implementation of digital technology in educational process.

The analysis of research results in the process of 20 years aims to show the dynamics of accomplished changes in the educational system, and higher education in the field of digital technology usage and modernization of the educational process in the period of 10 years.

Keywords: knowledge, education, development, techno-logy, IT.

I. Introduction

The process of teaching entails activity of an individual which results in acquisition od certain knowledge, skills and habits, attitudes and ideologies. In developed societies main developmental resource is precisely knowledge as a result of the process of teaching which we call „the human capital“ in knowledge-based economy. It is pointless to talk about importance of learning in institutions that offer formal education and in those that offer informal education which is becoming prevalently more present (to so-called „Lifelong learning“).

Taking into consideration that today's society and technology develop at lightning speed, without lifelong learning it is impossible to be up to date with the new trends, system innovations, and individual as such is not able to rise up to challenges which are set as an imperative demands of a modern man. Average timeframe in which certain knowledge becomes obsolete has never been shorter so it is clear why such tremendous pressure is put on permanent development and specialization in the context of ongoing education and lifelong learning. Once acquired knowledge or skill are not enough for the entire working life of an individual, but rather represent the cheapest way of acquiring competitive advantage because every investment in knowledge and skills development as a base capital and axis of development becomes the most lucrative way of investing into quality of products and services. [1, page 553]

Lifelong education term is often replaced with lifelong learning term but those two terms are not synonymous. Lifelong education entails only organized forms of learning while lifelong learning is much wider concept which entails deliberate, spontaneous and non-organized acquisition of knowledge. [2, pp 111–150]

II. Investment and Staff education

In recent few decades the tendencies to continue education during and after working life have come to the forefront. On the job training is becoming exceedingly more popular. The focus is gradually shifting from education of the younger generations to education and training of the adults from the ages of 25 to 60 or even older [3, page 44]. It is necessary to distinguish terms „education“ and „training“ where education represents the widest notion of general knowledge and skills as well as competences of the individual, but it can also be considered as a process of general acquisition of the knowledge and skills which form the basis for further development of knowledge and skills. Training can be defined as staff preparation for more successful completion of tasks or for some future specific skills which will be required from that very same staff where the aforementioned training is oriented towards work and less so on the actual individual. It is a system of organized program of choreographed physical or intellectual activity and it can entail exercise of some psychological traits. Need for training is a consequence of series of changes which happen daily in some company or institution. [4]

Methods of training managers by using work experience are based on acquiring managerial experience in doing managerial tasks and in the actual practice and experience where a manager has the ability to work under pressure of reality all the while learning from their own mistakes. Methods of manager training by using out of work experience are used as methods of experience acquisition by playing roles, solving managerial problems, simulation, training, etc. The training is work based while career is oriented towards the individual and their potential. Career is mutually connected, successive sequence of tasks, positions and work experience of a certain individual during their working life which follows changes in preferences, attitudes, experience and behavior of an individual. [5, page 831]

Staff development is a complex process of acquiring new knowledge, skills and abilities which enable perspective individual to take over more complex task after having completed more basic ones. The process of complex task takeover is defined as staff development. [1]. Career development is a process of series of interlinked and synchronized partial individual and organizational activities in which an individual and organization are partners in improvement and development of individual career. [5, page 832]

III. Theoretical review of Educational system investment trends in some European countries

Educational institution outlay (shown as GDP) shows in which quantity certain country puts priority on education in relation to entire distribution of funds. Around 60% of expenditures for educational institutions, more specifically, around 3,7% of entire GDP in OECD countries is dedicated to primary, secondary and postsecondary education without university-level education. For example, relative to their GDP, Iceland spends almost double the amount Greece does on education. Furthermore, university level education entails almost one third of the combined OECD expenditure for educational institutions (2,0% from GDP). More people are acquiring their high school and university diplomas than ever before, and in many countries the spread has been accompanied with mass financial investments. For all educational levels together, public and private investment in education is increased in all countries for at least 8% in 2005, and increase in OECD countries is on average around 42%.

Naturally, expenditure increase rate for educational institutions has changed greatly during 21st century. On average, all levels education expenses grew at a much lower rate in 2000 (around 10%) and increased significantly more than GDP between 2000 and 2005 (around 24%).

All level educational institution expenditures have increased in the last five years in 7 out of 28 OECD countries. On a tertiary level, during 2005, educational expenditures have increased at a much higher rate than the GDP growth rate. Only Belgium and Ireland had GDP growth rate higher than educational institutions expenditure rate. The biggest spending on educational institutions happened in Denmark and Iceland in 2009. 7 out of 28 OECD countries (according to the accessible statistics) spend less than 5% of GDP on educational institutions, and Greece and Russia spend 4,2, or rather 3,8%.

Differences in spending on educational institutions is the most obvious in preschool level. It varies between 0,2% GDP in Ireland to 0,8% GDP or more in Denmark, Hungary and Iceland. Differences on preschool level can be explained mostly by participation rates among young children, but sometimes they reflect the extent of private education in early childhood which is covered by indicator of investment in preschool education. In Ireland, for example, majority of education in early childhood is delivered in private institutions which are not yet included in official statistical data. GDP percentage that is spent on tertiary institutions in Belgium, France, Iceland, Switzerland and the Great Britain is below OECD average. In Switzerland, GDP portion reserved for tertiary institutions is one of the highest expenditure rates per student stemming from relatively low tertiary enrollment rate and high GDP.

If we were to compare the aforementioned data from European countries and research conducted during 70's and 80's, it is clear to see that education in the USA incentives and develops only two abilities; reproduction of facts and solving well defined problems. According to National Assessment of Education Progress, majority of students has not developed cognitive skills. Only 5-8% of seventy years old shows problem solving capabilities, synthesizes data, reads analytically and thinks critically. According to estimations, only 12 – 15% USA population is able to read with ease, apply simply mathematical operations, and use geographical map and draw charts.

After those educational institutions investment trends review, tendency and developmental outcome of a certain country is apparent in the context of investments in general educational and university-level educational system. Analyzed data from the previous chapter which extrapolate on education a le expenditure date back to 2005 – 2010 and as a consequence have a status quo in educational system without any major qualitative shifts. When we factor in the data that many of the aforementioned countries have not increased investments into general educational and university-level educational system in the last ten years (2010 – 2020) it is obvious to see why significant qualitative shifts cannot be expected.

Comparison of the old and new research

In 2005/6 Research was conducted by the Institute for Social Research in Zagreb and Centre for Research and Development of Education in which participated 2134 middle school teachers, 1134 elementary teachers, 2674 pupils and 120 principals in 120 schools in Croatia. The goal of the research was to examine their opinion about current curriculums. In this chapter there will be focus on implementation of the informational technology from the teachers' and students' perspective during teaching - learning process.

This research is to be compared with the similar research that I conducted in 2010/2011 for the purpose of comparing the overall IT usage trends in the span of 5 years among the high school students and university students. Similar research was conducted in 2020 and it included sample size of 860 subjects. Contrastive analysis points to significant changes in the educational system.

Right from the start, the results indicated that older teachers statistically use IT technology significantly less than younger teachers do, as well as new sources of information and tools while preparing for class and during class.

Information about *IT and digital usage* show that teachers rarely use it during class. Personal computer rarely or almost never uses 54% of teachers, Internet 61%, and 1/3 of teachers never use e-mail. According to aforementioned, majority of teachers rarely uses digital technology for teaching purposes. It is very rarely used as an *Information source in lesson preparation.* As it has already been pointed out, IT usage is directly correlated with the age of the teacher. However, testing the significance of discrepancy has shown that IT usage depends on the region and the location in which the school is located as well as equipment of the school.

These data show urgence of improving the quality of the mandatory educational system in Croatia. In recent research one of the reasons was insufficient equipment of the institutions in educational system but contemporary results give a different impression showing impressive levels of institutional equipment but a significant lack of educational competence. This datum, discovered during 2020 research, implies the urgency of IT education of the older teachers that work in schools located in poorly developed regions, villages, towns and cities including.

A. The 2020 research

As it has been mentioned before, this 2020 research entails 860 elementary, middle and high school teachers of the Splitsko-dalmatinska County, Dubrovačko-neretvanska County, Šibensko-kninska County and Zadarska county as well as preschool education students and elementary school education students from the Faculty of Philosophy in Split. The subjects were submitted to extensive surveys regarding IT usage in teaching –learning process. Methods that were implemented during research are the following: curriculum analysis, observational method, survey method, interviews, descriptive method, and comparative method. Surveys used during the research were the same or similar ones from the previous, earlier researches.

B. Goals and tasks

- To test whether pupils, students, elementary, middle and high school teachers and university professors have the tendency to deviate from the traditional- theoretical approach in teaching in favor of using more contemporary approach in teaching which includes extensive IT usage.
- To test what kind of attitudes pupils, students, teachers and professors have towards educational improvement and more significant usage of IT in educational system via lifelong self-educational process.

Within this context, the goal was to actualize the process of faster abandonment of the traditional teaching in favor of more contemporary teaching through the means of IT in teaching- learning process.

IV. RESEARCH HYPOTHESES

H1: Elementary, middle and high school teachers as well as university professor do not have positive attitude towards leaving traditional- theoretical teaching in favor of contemporary teaching process with IT heavy usage.

H2: Younger elementary, middle and high school teachers as well as university professors express the need for better education and IT implementation in educational system through the process of self-education and lifelong education, while older elementary, middle and high school teachers and university professors are not ready to adopt aforementioned changes.

V. RESEARCH RESULTS AND DISCUSSION

Characteristics of the answers of the elementary, middle and high school teachers as well as university professors’ subsample

Elementary, middle and high school teachers as well as university professors' answers to survey questions

TABLE I. FREQUENCY OF THE SURVEY ANSWERS. IF THE SCHOOL IS TO ORGANIZE EDUCATION WITH THE PURPOSE OF BETTER QUALIFICATION OF THE EMPLOYEES WHEN IT COMES TO IT USAGE DURING THE TEACHING PROCESS, WOULD THEY ACCESS THE PROCESS OF EDUCATION?

	Frequency	%
Yes	450	**53,32**
No	370	44,02
Total	820	97,67
Missing answer	40	4.65
Total	860	100,0

Therefore, significant concern is evident in the new results which show teachers' lack of interest in education and permanent self-training. Therefore, over 44% of the subjects is not interested in education for the purpose of being more qualified in the field of IT usage during the teaching - learning process. This data is concerning and it does not leave us indifferent but, on the contrary, shows the need of making vital changes in this segment.

Due to shortcomings and existing oversights as well as poor lifelong education curriculum realization many changes need to be implemented in the form of lifelong learning project which will measure the direction in which the society will develop as a learning society. [6].

A. Average achievements of the subjects according to survey items

Average achievements of the teachers according to survey items are expressed by arithmetic mean and standard deviations. Considering the framework of possible answers was between 1 to 5, those values denote the strength of answers 1 having the lowest value and 5 having the highest.

TABLE II. ARITHMETIC MEAN, STANDARD DEVIATIONS, MINIMUM AND MAXIMUM VALUE.

	Minimum	Maximum	Arithmetic mean	Std. Deviation
Ever since I was a student little has changed in the field of IT usage	1,00	5,00	2,9400	1,24380
Pupils and students attend class: TV, CD, computer, Internet....	1,00	5,00	3,4286	1,31633
Schools should be more welcoming to the many technological changes.	1,00	5,00	4,1200	,95521
Pupils have the option of choosing elective class according to their own personal interest.	3,00	5,00	1,7595	1,06476
The class i teach doesnot keep up with the contemporary content because the conditions for e-teaching are non existent.	1,00	5,00	1,8286	1,31633
pupils are happy with the choice of textbook	1,00	5,00	1, 9311	1,13645
Pupils often use their textbooks.	1,00	5,00	2,3880	1,18615
Teachers are pleased with the choice of textbooks.	1,00	5,00	4 1377	1,21873
Curriculum needs thorough changes in the context of new learning technologies.	1,00	5,00	4,8843	,75236
Cabinets and classrooms are informatically equipped.	1,00	5,00	4,2921	1,09238
Better IT equipment would offer higher quality teaching.	1. 00	5.00	1,9912	1,30402
I would enroll into extracurricular IT education with the purpose of having better quality teaching.	1.00	5.00	3,3301	,84587
I believe in better quality teaching based on IT based teaching.	1.00	5.00	3,7166	1,20861
It is necessary that teachers know how to use computer or that necessary IT experts are available in case they need to help teachers out.	1.00	5.00	3,7764	1,21327
introducing IT tools in the teaching-learning process would mean that students are encouraged to think critically and creatively.	1.00	5.00	2,4324	1,18361
IT as an obligatory course would be much more interesting if it was integrated with other classes and fields of knowledge.	1,00	5,00	4,8936	3,61854
Employees need to be actively involved in innovating teaching-learning process of the given school or university.	1.00	5.00	4,6221	1,49027
Psycho-physiological distress can be reduced with better implementation of the contemporary IT devices in the teaching-learning process.	1.00	5.00	3,1952	1,30402

Results show that teachers on average show the highest agreement on the following questions:

- Introducing digital technologies in the teaching process would have a beneficial and encouraging effect on pupils' creative and critical thinking
- Teachers are not sufficiently educated in the IT field
- Psycho-physiological distress can be reduced with better implementation of the contemporary IT devices in the teaching-learning process.

From the above shown analysis regarding investing in the educational system in certain European countries, the systemic negligence of investing in the process of modernization of the teaching system is apparent as well as insufficient willingness to change in the direction of education and implementation of IT tools in class. Additional problem can be seen in not so insignificant resistance to change.

TABLE III. ARITHMETIC MEANS, STANDARD DEVIATIONS, MINIMUM AND MAXIMUM ANSWER VALUES.

	Minimum	Maksimum	Aritmetička sredina	Std. Devijacija
Maths	1,00	7,00	4,1233	3,72080
History	1,00	7,00	**5,9965**	2,53691
Foreign languages	1,00	7,00	5,1651	8,60374
Biology	1,00	6,00	**4,9544**	1,64496
Religious education	1,00	7,00	3,1223	8,06536
Arts	1,00	8,00	3,9652	2,38502
Other	1,00	5,00	1,6256	81,81738

According to students – pupils' assessment, interest of the students' is higher proportional to representation of contemporary forms of teaching in specific subjects.

Comparing these data to the ones conducted by Institute for Social Research 10 years ago in Zagreb, it is evident that contemporary 2020 results denote how 92% of students and pupils uses the Internet for information research, 84% uses IM as the main mean of communication (64% of adults), 91% communicates with friends or family via electronic means, 92% uses current messaging system as a mean to stay in touch with friends, 86% prefers using the Internet instead of telephone.

New data indicate significant rise in usage of technology compared to the previous period which can loosely be interpreted as a significant progress in the last decade. Unfortunately, that cannot be attributed to the beneficial changes in the teaching system as those are minimal or non-existent especially when compared to field of IT, internetization and globalization in the past 10 years.

What naturally follows from this discussion is that the first hypothesis is entirely rejected which denotes a change in the pupils and students' attitude compared to that of 10 years ago. Pupils and students have changed their attitudes towards rejecting traditional ways of teaching in favor of contemporary process aided with IT. Other hypothesis, however, is entirely accepted and shows bigger agreement and need for quality education with the purpose of implementation IT tools in educational system through the process of self-education and lifelong education.

VI. Conclusion

Continuous professional development is imperative to achieving professional functioning in a forever-changing societal and technological context that has been imposed by the 21st century. [7]

It is not correct to believe that knowledge acquired by initial education is enough for quality professional work throughout entire working life.

Lifelong learning is an answer to the crisis of ever faster obsolescence of knowledge and skills which is caused by fast-paced rhythm of technological and societal changes. Lifelong learning is also very important for development of new informational and communicative technologies.

Instead of present unsuitable approach, IT based learning needs to cover:

- Software tutoring during which teachers guide students through matter covered with tutoring for evaluation and reflection of knowledge which offers automatic answer assessment.

- Software simulations where students and pupils in „ virtual reality "complete tasks for which they should be qualified to perform both in reality of is work place and in real time.

- Hypertext and hypermedia should be available as well as other navigational systems and links with relevant content of the multimedia nature such as text, drawing, picture, movie, sound...

- Computer tools for communication and collaboration support during studying in which fall distance learning like Moodle or Black Board, Workflow Management systems, CMS, Internet-Relay- Chat, e-mail, Mailing lists, News forums, video conferences...

Development of the contemporary educational systems that will remove all resistance to change is based on the concept of *lifelong education* and *the society that studies.* Listed concepts alongside cognitive area emphasize the significance of development and other pupils' areas, especially motivational, social, and emotional.

References

[1] Cerović, Z., Hotelski menadžment**.** Opatija**:** Fakultet za turistički i hotelski menadžment, 2003.

[2] Puljiz, I., Živčić, M., Pojmovnik politika obrazovanja odraslih, Zagreb, 2009.

[3] Erceg, C.J. and A. T. Levin,. "Imperfect Credibility and in ation Persis-tence". Journal of Monetary Economics 50, 2003.

[4] Čičin-Šain, D. "Sustav organiziranja i organizacijske strukture" http://www.unizd.hr/portals/4/nastavni_mat/2_godina/menadzment/menadzment_11.pdf, Retrieved 25/03/2018.

[5] Bahtijarević-Šiber, F., Management ljudskih potencijala. Zagreb: Golden marketing, 1999.

[6] Cerović, Z., "Hotelski menadžment", Fakultet za turistički i hotelski menadžment, Opatija, 2003.

[7] Antikainen, A. (ed.), Transforming a Learning Society: The case of Finland. Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien: PeterLang, 2005/

[8] Fatović, M., Časopis za pedagogijsku teoriju i praksu, Školski vjesnik, Vol. 65 No. 4, 2016.

[9] Altrichter, H., Der Lehrerberuf: Qualifikationen, strukturelle Bedingungen und Professionalität. In W. Specht & J. Thonhauser (Hrsg.), Schulqualität. Entwicklungen, Befunde, Perspektiven, Innsbruck, Studienverlag, 1996.

[10] Baranović, B, Što nastavnici misle o nastavnim programima za obvezno obrazovanje-rezultati istraživanja, Zagreb, Institut za društvena istraživanja i Centar za istraživanje i razvoj obrazovanja, 2005.

[11] Lučić, K., Odgojne znanosti vol. 9, br. 1, 2007.

[12] Previšić, V., Pedagogija i metodologija kurikuluma, u Kurikulum, Zavod za pedagogiju Filozofskog fakulteta, Zagreb, Školska knjiga, 2007.

Analyses on Open-source LoRaWAN Simulators GitHub Project Statistics

N. Kermeci *, M.. Mazalica **and D. Dobrilovic**
* TD Wealth, Wellington West, Toronto, Canada
** University of Novi Sad / Technical Faculty "Mihajlo Pupin" Zrenjanin, Zrenjanin, Serbia
nkermeci@rogers.com; milica.mazalica@tfzr.rs; dalibor.dobrilovi@tfzr.rs;

Abstract - Open source software (OSS) is an extremely popular type of software. GitHub is a social coding platform, which allows sharing changes made in the code. GitHub is a git-based control system for versioning control. LoRa and LoRaWAN technology are groups of LPWAN (Low-Power Wide Area Network) technology, and with their features, they are suitable for long-range IoT and smart applications. This paper presents the analyses of GitHub based project of LoRaWAN simulators. An effective tool such as GitHub statistics can offer information about projects, their updates, popularity, etc. This paper is focused on LoRaWAN simulators representing its popularity and estimated growth in the future. The research results of this paper provide the best possible recommendation for the tool which can be used to simulate LoRaWAN networks, as well as the directions for building own simulation tools.

I. INTRODUCTION

Nowadays the Open source software (OSS) is an extremely popular type of software among the community of programmers and computer users. GitHub is OSS forge founded in 2008 to simplify code sharing. [1] After foundation GitHub is growing up fast, and nowadays is the largest repository for OSS project and hosts the world's largest collection of open-source software. [3] The major reason which made GitHub so successful is the fact that project hosting is based on the popular git system.

GitHub is a git-based control system that integrates several features for social coding. The user of GitHub can star (like) the hosted project and in this way, they show their interest or satisfaction whit the hosted project. The success of GitHub brings competitiveness to the open-source market, and as the result, offers an opportunity for projects to compete for the user's attention. [3]

This paper presents the analyses of GitHub based project of LoRAWAN simulators. An effective tool such as GitHub statistics can offer detailed information on projects, their updates, and popularity. The paper is focused on LoRaWAN technology because of its popularity today as well as estimated growth in the next five years.

This paper is structured as follows. In the Introduction section, basic information about open source software and GitHub platform is given, next more information about GitHub and his importance is given in section GitHub Platforms. The next section is about LoRa and LoRaWAN technology which is in focus in this paper, LoRaWAN simulation, and LoRaWAN GitHub project statistics are given in sections IV and V. The results, analysis, discussion, and concluding remarks are given at the end of this paper.

II. GITHUB PLATFORM

GitHub is a social coding platform for sharing open-source code which allows third parties to share their changes to the code. The social coding site GitHub share open-source code for software and website development, distribution of data sets, and research results. [2]

Open collaboration is a new form of innovation in the public sector. Online platforms are used by Government organizations to collaborative create or contribute to public sector innovations with the help of external and internal problem solvers. [2]

The paper [2] shows, that working with the code and discovering issues is a much more active form of social coding or open collaboration, but it is less frequently observable in the U.S. federal government. Paper [1] analyzes the OSS projects of the GitHub portal and mines success rules for such projects. The paper [3] provides a study on the meaning, characteristics, and dynamic growth of GitHub stars. The research results show a list of recommendations to open source project managers and to GitHub users and Software Engineering researchers.

Basic GitHub concepts include:

- Initializing of a Repository.
- Cloning of a Repository.
- Adding and updating the context.
- Browsing differences and change history.
- Branching, merging, and expanding.

III. LORA AND LORAWAN TECHNOLOGY

Low-Power Wide-Area (LPWA) networks are a group of relatively modern technologies designed for low power, low bit-rate wireless communication with low power consumption. With their features, they are suitable for the Internet of Things (IoT) and smart city applications. LPWAN technologies can be divided into two groups. LoRa and LoRaWAN, together with SigFox, Ingenu, Weightless, and more belong to the category of non-cellular LPWAN wireless communication network protocols. The other group is cellular LPWANs such as NB-IoT and LTE-M.

LoRa and LoRaWAN can be considered a single protocol stack where LoRa is a radio modulation technology operating at the physical layer, and LoRaWAN defines the communication protocol and system architecture for the network. LoRa and LoRaWAN operate in the unlicensed spectrum in several bands. In Europe, LoRa operates at 433 and 868 MHz, in North America LoRa operates in 914 MHz, and in Asia, it operates in 430 MHz bands.

In Fig.1 and Table I the estimation of LPWAN technology growth in millions of connections is shown.

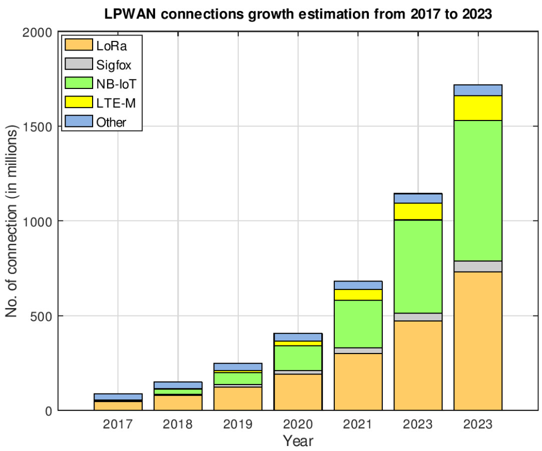


Figure 1. LPWAN connections growth estimation

As it is shown in Table 1. LoRa/LoRaWAN is the most applied LP-WAN technology these days, but in the future, NB-IoT will gradually become more popular. It is an estimation that in 2023 NB-IoT will have an increase of 739.8 million connections and LoRa/LoRaWAN will have 730.69 million connections worldwide.

TABLE I. LPWAN CONNECTIONS GROWTH ESTIMATION 2017-2023 IN MILLIONS OF CONNECTIONS

Year/Tech.	LoRa	SigFox	NB-IoT	LTE-M	Other
2017	46.39	2.46	5.43	0.86	32.4
2018	78.34	6.16	27.42	3.86	34.12
2019	123.33	11.93	64.94	10.51	36.59
2020	190.76	19.94	129.58	25.93	39.9
2021	299.06	30.12	252.08	56.04	43.94
2022	470.69	42.93	491.19	89.83	49.22
2023	730.69	58.05	739.8	132.75	55.7

Despite estimations that LoRa/LoRaWAN technology will be outnumbered by NB-IoT connections soon, it will have an important share of the market. Based on this estimation, the importance of developing applications for the simulation of LoRa/LoRaWAN technology is very clear. Besides, it is necessary to analyze their popularity, promptness, etc.

IV. LORAWAN SIMULATION

Because of the number of implemented LoRa/LoRaWAN technology systems, their good characteristics and high applicability in modern ICT systems such as IoT, smart cities, etc., the LoRaWAN simulation tools have greater importance,

There is a variety of simulation tools applicable to LoRAWAN simulation. Some of the most popular tools are OMNET++ and ns-3 simulator.

OMNET++ is a modular, simulation library and framework designed for building network simulations. FLoRa (Framework for LoRa) is an open-source simulation framework LoRa simulations based on the OMNeT++ [25, 26].

NS-3 [21], as well as its predecessor NS-2 [22], are discrete-event network simulators designed for research and educational use. They are free software, licensed under the GNU GPLv2 license, and both tools are publicly available for research and development applications for a number of networking technologies and LoRAWAN among the other as well. [23, 24]

Besides described popular simulating platforms, a certain number of LoRaWAN simulations is built on SimPy/Python platform [28, 29, 30]. This paper is focused on LoRaWAN simulation open-source project hosted on GitHub platform no matter the tool or programming language used for building.

V. LORAWAN SIMULATION GITHUB STATISTICS

This chapter presents a statistic comparison of GitHub projects related to LoRaWAN simulation. The project comparison includes comparison by popularity and maturity metrics as the most important parameters. All this information is collected with the GitHub Stats tool. [14] The selection of the project for the comparison is made by GitHub search with key phrases "LoRaWAN simulation" and "LoRaWAN simulator". A total of 27 projects are found and included in the research.

TABLE II. LORAWAN SIMULATORS GITHUB PROJECT DETAILS

Project	Date Created	Primary Language	Last Push	Watchers
Bulk-LoRa	Nov 29, 2018	Perl	4 months	1
chirpstack-simulator	Feb 20, 2020	Go	a month	5
Flora	Jul 04, 2018	C++	2 years	1
FREE	Dec 13, 2018	Python	3 months	1
IQ-sample-based-simulator-for-LoRaWAN	Jul 25, 2020	Python	2 months	0
Lds	Nov 20, 2018	Go	a month	8
Lora	Apr 23, 2020	C#	4 months	1
LoRa_LBT_simulator	Jul 14, 2017	Python	3 years	1
LoRaEnergySim	Jan 11, 2018	Python	2 months	3
lora-motes-emulator	May 30, 2018	Python	4 months	2
lora-pktgen	Apr 19, 2017	Go	4 years	4
lora-sensor-simulator	Aug 10, 2019	Rust	10 months	1
Lorasim	Mar 26, 2017	Jupyter Notebook	9 months	5
LoRaSim	Sep 04, 2020	Python	2 days	1
LoRaSim	May 15, 2019	Python	2 months	0
LoRaSIm-Multiple-Applications-Support	Jun 19, 2018	Python	2 years	0
LoRaSimulator	Jul 22, 2017	Java	3 years	2
Lorasimulator	Jul 04, 2018	Python	2 years	1
LoRaSimulator2	Nov 10, 2017	Python	3 years	4

Lorawan	Sep 30, 2017	C++	a month	19
LoRaWAN-SIM	May 11, 2020	Perl	3 months	1
LoRaWanSimu	Nov 26, 2018	C++	2 years	0
Lpwansimulation	Jan 23, 2017	MATLAB	4 years	8
nimbuslorasim	Jan 08, 2018	Python	3 years	2
ns-3-dev-with-signetlabdei-lorawan-module	Nov 24, 2018	C++	a year	1
Simlorasf	Nov 24, 2018	Python	8 months	1
Tesi-LoRaWANSimulator	Oct 05, 2017	Java	3 years	1

Table II presents projects which are compared and their general repository information such as date created, primary language, last push at, and the number of watchers. The most watchers have lorawan project with 19, and lds and lpwansimulation with 8. For the most updated project, the last push is 10 days ago, and the latest last push is 4 years ago. The primary language used for these projects are as follows: Python with 12 projects or 42.86%, C++ with 4 projects or 14.29%, Go with 3 projects or 10.71%, Java and Perl with 2 projects and 7.14%, and MATLAB, C#, Jupyter Notebook, Rust, and C with 1 project or 3.57% of all projects.

The first created project was lpwansimulation created in January 2017, and the latest project is IQ-sample-based-simulator-for-LoRaWAN created in June 2020.

TABLE III. GITHUB TOP 10 STARS STATISTICS

Project	Total Stars	Average stars/day	Max. stars/day
Lorawan	81	0.07	2
Lpwansimulation	30	0.02	1
lora-pktgen	19	0.01	1
Lds	19	0.02	1
lorasim	15	0.01	1
lora-motes-emulator	14	0.01	3
chirpstack-simulator	13	0.05	4
FREE	9	0.01	1
LoRaEnergySim	5	0	1
Tesi-LoRaWANSimulator	2	0	1

The popularity of a project is measured by several watchers, stars, and forks. In Table II are shown stars for the top 10 projects. Project lorawan (LoRaWAN ns-3 module) is the project with the most stars - total 81.

Table VI shows forks for the top 10 compared projects and projects with the most forks are lorawan (57) and lora-pktgen (24).

TABLE IV. GITHUB TOP 10 FORKS STATISTICS

Project	Total forks	Avg. forks/day	Max. forks/day
lorawan	57	0.05	1
lora-pktgen	24	0.01	1
Lds	15	0.02	1
lpwansimulation	14	0.01	1
lorasim	10	0	1
chirpstack-simulator	8	0.03	1
lora-motes-emulator	5	0	1
FREE	3	0	1
LoRaEnergySim	3	0	1
lorasimulator	2	0	2

The maturity of projects is measured by the number of commits (Table V), the number of issues (Table VI), and the number of pull requests (Table VII).

The number of top 10 total commits for these projects is given in Table V. The project with the most commits is LoRaSimulator2 (166), followed by lds (67) and lorawan (65).

Table VI shows the top 10 issues by the projects, also the number of total issues, average, and maximum issues per day. The projects with the highest number of total issues, average and maximum issues per day are lorawan, lds, and LoRaEnergySim, respectively.

TABLE V. GITHUB TOP 10 COMMITS STATISTICS

Project	Total commits	Avg. commits/day	Max. commits/day
LoRaSimulator2	166	0.11	34
Lds	67	0.16	7
Lorawan	65	0.16	39
Tesi-LoRaWANSimulator	41	0.02	12
LoRaWAN-SIM	39	0.08	10
lora-motes-emulator	35	0.07	3
LoRaSimulator	34	0.02	13
LoRaSim	21	0.05	8
simlorasf	19	0.03	3
chirpstack-simulator	18	0.04	5

TABLE VI. GITHUB TOP 10 ISSUES STATISTICS

Project	Total issues	Avg. issues/day	Max. issues/day
lorawan	68	0.06	2
lds	17	0.02	2
LoRaEnergySim	9	0	2
lora-motes-emulator	3	0	1
lpwansimulation	2	0	1
lorasim	1	0	1
chirpstack-simulator	1	0	1
FREE	1	0	1
lora-pktgen	0	0	0
lorasimulator	0	0	0

Table VII shows the top 10 projects with the number of total pull requests, average and maximum requests per day. The projects with the highest and average number of total pull requests per day are lorawan.

TABLE VII. GITHUB TOP 10 PULL STATISTICS

Project	Total pull requests	Avg. pull requests/day	Max. pull requests/day
Lorawan	25	0.02	6
LoRaSimulator2	9	0	1
Basicstation	8	0.01	3
Lds	8	0.01	1
LoRaSimulator	5	0	2
LoRaEnergySim	1	0	1
Lorasim	1	0	1
lora-motes-emulator	0	0	0
lpwansimulation	0	0	0
chirpstack-simulator	0	0	0
FREE	0	0	0

The GitHub statistics analyses are given in the next chapter.

VI. RESULTS ANALYSES

If we compare all of the 27 presented projects, seven projects will have better statistical indicators comparing to others. These seven most popular projects are shown in Fig. 2.

Project *lorawan* (LoRaWAN ns-3 module) seems to be most popular with 19 watchers, 8 stars, and 57 forks. It can be also considered as the most mature according to [14] with 68 issues and 25 pulls, as well as the most updated one with 65 commits, and with the last change dating within one month. This is an ns-3 module designed to perform simulations of a LoRaWAN network [9, 10].

The software is licensed under the terms of the GNU GPLv2 (same as ns-3). There are a number of publications dealing with the simulator [11, 12, 13]. This module contains a series of classes and examples focused on modeling the modulation and medium access technology of a LoRaWAN network. It has a simple physical layer model and it can support simulations with many devices that access the wireless channel infrequently.

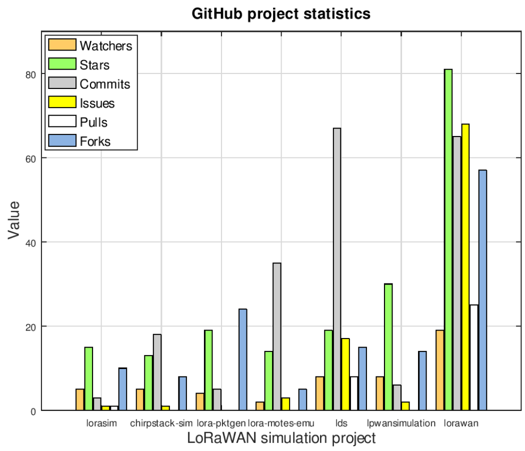


Figure 2. GitHub project statistics

Detailed data about each of the seven projects shown in Fig. 2 are given in Table VIII.

TABLE VIII. GITHUB PROJECTS SUMMARY STATISTICS

Project	Watchers	Stars	Commits	Issues	Pulls	Forks
lorasim	5	15	3	1	1	10
chirpstack-simulator	5	13	18	1	0	8
lora-pktgen	4	19	5	0	0	24
lora-motes-emulator	2	14	35	3	0	5
lds	8	19	67	17	8	15
Lpwansimulation	8	30	6	2	0	14
lorawan	19	81	65	68	25	57

After the LoRaWAN ns-3 module, three projects have similar statistical parameters: lora-pktgen (4 watchers, 9 stars, and 24 forks), lds (8 watchers, 19 stars, and 15 forks), and lpwansimulation (8 watchers, 30 stars, and 15 forks). Other projects such as chirpstack-simulator and lora-motes-emulator have a significant number of commits but are slightly less popular having a lower number of watchers, stars, and forks compared to the described four projects.

Lora device simulator (lds) is a program for the simulation of LoRaWAN devices. This program supports all bands and configurations LoRaWAN versions 1.0 and 1.1 and is compatible with the newest versions of ChirpStack Network Server modules (v3 test) and lorawan-server.[15] This is a utility program to simulate LoRaWAN devices and for the ChirpStack project, it acts as a chirpstack-gateway-bridge middleman, publishing and receiving messages through MQTT.

Project *lpwan* simulation is used to simulate Low Power Wide Area Networks (LP-WAN) and especially for non-cellular LP-WANs LoRaWAN and Sigfox. [17]

LoRaWAN Packet Generator (GW Simulator) is a command-line tool for the generation of UDP packets that can be sent from the PC host to the LoRa Network Server. It is designed to simulate LoRaWAN gateway and send the UDP packages defined by the "Gateway to Server Interface Protocol" defined in Semtech document ANNWS.01.2.1.W.SYS. Project lora-pktgen has features of both the LoRa node and LoRa Gateway devices. It can generate packages from a user-defined payload, adding service information in JSON that is injected by Packet Forwarder during the packet traversing through a gateway. This allows LoRa Network Server to get a properly shaped UDP packet, like the one that would come from a real hardware device. [18] In that way, lora-pktgen can be used to test the LoRa Network Server deployments without need for having expensive hardware such as gateways and nodes and the rest of the network set-up.

LoRa Motes Emulator is a tool designed for testing the LoRa server, and for end devices emulation. It supports LoRaWAN 1.0.2. & 1.1 protocol and requires Payton (3.6 mandatory) and Ubuntu. [19]

ChirpStack Simulator is an open-source simulator for the ChirpStack open-source LoRaWAN® Network-Server stack [16]. It simulates several devices and gateways, which are automatically created during the start of the simulation. The project has been developed together with TWTG, an ISO 9001-certified company with industrial-grade solutions and a roster of international clients in industry, safety, and energy. [21]

VII. CONCLUSION

This paper presents the analyses of GitHub based project of LoRaWAN simulation. For collection information of LoRaWAN technology is used GitHub stats tool. The results and analyses show that the most popular LoRaWAN simulation project is LoRaWAN ns-3 module. This project is followed in popularity with lora-pktgen, lds, and lpwansimulation. The most used programming language for the simulation is Python with 12 projects or 42.86% of all projects, followed by C++ with 4 projects or 14.29%, and Go with 3 projects or 10.71%

Considering the importance of LoRaWAN and non-cellular LP-WAN technologies, and their growth in a number of connections in the future it is really important to analyze and observe the best possible tools for their simulation. The simulation of the aforementioned technologies is important in order to facilitate and make more effective the implementation process of these technologies in the future IoT systems.

The results of this research give the best possible recommendation for the tool which can be used to simulate LoRaWAN networks, as well as the directions for building own simulation tools.

REFERENCES

[1] F. Chatziasimids, I.Stamelos, Data collection, and analysis of GitHub repositories and users, 2015 6th International Conference on Information, Intelligence, Systems, and Applications (IISA). doi:10.1109/iisa.2015.7388026.

[2] I. Mergel, Open collaboration in the public sector: The case of social coding on GitHub, Government Information Quarterly, Vol. 32, Issue 4, 2015, pp 464-472, https://doi.org/10.1016/j.giq.2015.09.004.

[3] H. Borges, M. Tulio Valente, What's in a GitHub Star? Understanding Repository Starring Practices in a Social Coding Platform, Journal of Systems and Software,Volume 146, 2018,pp 112129, https://doi.org/10.1016/j.jss.2018.09.016.

[4] AN1200.22 LoRa Modulation Basics, Available at https://www.semtech.com/uploads/documents/an1200.22.pdf, Semtech, Accessed 2020/02/05.

[5] Lora-alliance, Retrieved at https://lora-alliance.org/sites/default/files/2018-07/lorawan1.0.3.pdf, Accessed February 2020.

[6] S. O'Dea, Global LPWAN connections 2017-2023, by technology, Feb 27, 2020, Statista, Available at https://www.statista.com/statistics/880822/lpwan-ic-market-share-by-technology/, Accessed September 2020.

[7] D. Magrin, M. Capuzzo and A. Zanella, "A Thorough Study of LoRaWAN Performance Under Different Parameter Settings," in IEEE Internet of Things Journal, vol. 7, no. 1, pp. 116-127, Jan. 2020, doi: 10.1109/JIOT.2019.2946487.

[8] GitHub, https://github.com , Accessed September 2020.

[9] GitHub, LoRaWAN ns-3 module, https://github.com/signetlabdei/lorawan, Accessed September 2020.

[10] LoRaWAN ns-3 module, https://signetlabdei.github.io/lorawan-docs/models/build/html/lorawan.html#usage, Accessed September 2020.

[11] D. Magrin, M. Capuzzo and A. Zanella, "A Thorough Study of LoRaWAN Performance Under Different Parameter Settings," in IEEE Internet of Things Journal. 2019. Link.

[12] M. Capuzzo, D. Magrin and A. Zanella, "Confirmed traffic in LoRaWAN: Pitfalls and countermeasures," 2018 17th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), Capri, 2018. Link.

[13] D. Magrin, M. Centenaro and L. Vangelista, "Performance evaluation of LoRa networks in a smart city scenario," 2017 IEEE International Conference On Communications (ICC), Paris, 2017. Link.

[14] Compare Similar GitHub Projects with this GitHub Stats Tool, https://dev.to/nebulagraph/compare-similar-github-projects-with-this-github-stats-tool-1n3e, Accessed September 2020.

[15] GitHub, LoRa device simulator, https://github.com/iegomez/lds, Accessed September 2020.

[16] ChirpStack, https://www.chirpstack.io/, Accessed May 2020.

[17] GitHub, lpwan simulation, https://github.com/maartenweyn/lpwansimulation, Accessed September 2020.

[18] GitHub, lora-pktgen, https://github.com/donadonny/lora-pktgen, Accessed September 2020.

[19] GitHub, lora-motes-emulator, https://github.com/houluy/lora-motes-emulator, Accessed September 2020.

[20] GitHub, ChirpStack simulator, https://github.com/brocaar/chirpstack-simulator, Accessed September 2020.

[21] The Network Simulator - ns-3, [Online] retrieved in May 2020, https://www.nsnam.org/

[22] The Network Simulator - ns-2, [Online] retrieved in May 2020, https://www.isi.edu/nsnam/ns/

[23] Brecht Reynders, Qing Wang, and Soie Pollin, 2018. A LoRaWAN Module for ns-3: Im-plementation and Evaluation. In Proceedings of the 2018Workshop on ns-3 (WNS3 2018), (June 2018), Surathkal, India. ACM, New York, NY, USA, 61-68. https://doi.org/10.1145/3199902.3199913

[24] D. Magrin, M. Centenaro and L. Vangelista, 2017. Performance evaluation of LoRa net-works in a smart city scenario, In Proceedings of 2017 IEEE International Conference on Communications (ICC), Paris, 1-7, doi: 10.1109/ICC.2017.7996384.

[25] M. Slabicki, G. Premsankar and M. Di Francesco, 2018. Adaptive configuration of lora networks for dense IoT deployments, In Proceedings of NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, Taipei, 1-9, doi: 10.1109/NOMS.2018.8406255.

[26] Luca Sciullo, Angelo Trotta and Marco Di Felice, 2020. Design and performance evaluation of a LoRa-based mobile emergency management system (LOCATE), Ad Hoc Networks, Vol. 96, 2020, 101993, https://doi.org/10.1016/j.adhoc.2019.101993.

[27] J. Markkula, K. Mikhaylov and J. Haapola, 2019. Simulating LoRaWAN: On Importance of Inter Spreading Factor Interference and Collision Effect, In Proceedings of ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 1-7, doi: 10.1109/ICC.2019.8761055.

[28] SimPy, Team SimPy, [Online] retrieved in May 2020. https://simpy.readthedocs.io/en/latest/

[29] Martin Bor, Utz Roedig, Thiemo Voigt and Juan Alonso, 2016. Do LoRa Low-Power Wide-Area Networks Scale?, In Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, November, 59–67, https://doi.org/10.1145/2988287.2989163

[30] Thiemo Voigt, Martin Bor, Utz Roedig, and Juan Alonso, 2017. Mitigating Inter-Network Interference in LoRa Networks, In Proceedings of the International Conference on Embedded Wireless Systems and Networks, EWSN '17, (February 2017), 323–328.

Improving Competitiveness of Domestic SMEs with Cloud Computing

Mihalj Bakator*, Dragan Ćoćkalo*, Dejan Đorđević*, Cariša Bešić**, Miloš Vorkapić***
* University of Novi Sad, Technical Faculty "Mihajlo Pupin", Đure Đakovića bb, 23000, Zrenjanin, Republic of Serbia
** University of Kragujevac, Technical Faculty Čačak, Svetog Save 65, 32000 Čačak, Republic of Serbia
*** University of Belgrade, Institute of Chemistry, Technology and Metallurgy (ICTM) - Center of Microelectronic Technologies (CMT), Belgrade, Serbia Njegoševa 12, 11000 Belgrade
mihalj.bakator@uns.ac.rs; dragan.cockalo@tfzr.rs; dejan.djordjevic@tfzr.rs; car.besic@gmail.com; worcky@nanosys.ihtm.bg.ac.rs

Abstract - The fourth industrial revolution - Industry 4.0 and the globalization of markets has put a tremendous amount of pressure on domestic enterprises when it comes to achieving and maintaining competitive ability on the market. Modern ICT are becoming a necessity in the modern way of conducting business. Such technologies are the "carriers" of the fourth industrial revolution, and enterprises have to adapt to the changes that these technologies bring to the business world. There are a lot of factors which influence the level of competitiveness of an enterprise. In this paper the application of cloud computing and its potential in improving competitive ability of domestic enterprises is addressed. The main goal of this paper is to analyze the issues and challenges of domestic enterprises' competitiveness. Additionally, the potential of cloud computing solutions and their effect on the domestic business environment is discussed. Further, guidelines and propositions for enterprises are provided.

INTRODUCTION

The globalization of markets and their constant fragmentation and segmentation has brought challenges to enterprises when it comes to achieving competitiveness and maintaining that competitive position on globalized markets. The rapid spread of information-communication technologies (ICTs) has added additional "pressure" as enterprises have to adapt to the changes these technologies brought to the market [1]. Namely, ICTs are the cornerstone of the fourth Industrial revolution - Industry 4.0. This revolution is characterized by modern technologies which include (but are not limited to) social product development; Radio Frequency Identification (RFID); 3D printing; smart sensors; wireless sensors; cloud based solutions (manufacturing, computing, logistics etc.); Internet of Things (IoT); Internet of Value (IoV); cyber security; advanced robotics, and Big data analytics [2]. Some of these technologies are narrowly specialized while others are more generic and can applied in various industry settings. Depending on the goals which are aimed to be achieved, the adequate technology solution is introduced. Such solutions have the potential to positively affect business performance and competitiveness. Modern enterprises aim at implementing ICTs into their strategies and business operations [3].

Further, as Serbia is a country in transition, the challenges of competitiveness within the Industry 4.0 framework are even more pronounced. Domestic enterprises lack productivity, adequate product and service quality, the manufacturing and tools are outdated, and there is a lack of adequate modern management tools and techniques application [4]. Now cloud computing as an on-demand network model where computing resources are shared, has a strong influence on various industries. It is viewed as a highly adaptable, applicable and scalable concept in the domain of ICTs. When it comes to conducting business within the framework of the fourth industrial revolution, enterprises have to consider the application of cloud computing, or better say, to move from traditional systems to cloud enterprise systems. Does such a transfer bring success? It depends. Some of the main factors which determine if a cloud enterprise system will bring improved business performance and competitiveness include financial support, compatibility, support from vendors and pressure within a certain industry [5].

In this paper the potential of cloud computing technologies in improving the competitive ability of domestic enterprises is analyzed. The main goal is to provide a basis for future research and to discuss potential guidelines and propositions regarding the improvement of competitiveness of domestic enterprises. The paper consists of four main sections (excluding the *Introduction* and *Conclusion* sections). The first section analyzes the competitiveness of domestic enterprises and the present trends of the global economy. The second section outlines the potential use-cases of cloud computing in domestic enterprises. The third section discusses possible future scenarios regarding the competitiveness of the domestic economy. The fourth section discusses guidelines and suggestions for improving the competitive ability of domestic enterprises.

I. COMPETITIVENESS OF DOMESTIC ENTERPRISES AND THE GLOBAL ECONOMY

Domestic enterprises are facing barriers when it comes to achieving and maintaining the competitiveness on globalized markets. The main reasons for such issues include low product and service quality, inadequate application of knowledge management, inadequate application of modern management tools and techniques, obsolete manufacturing equipment, and inadequate organizational infrastructure [6]. Further, the level of investments from GDP (in %) into research and development is around 0.93%. This is not sufficient, if we compare it to West European countries who invest 2% of their, several times larger, GDP [7].

According to the Competitiveness report issued by the World Economic Forum, Serbia ranks 72nd (from a total of 141 surveyed countries). In comparison, Croatia is ranked 63rd; Slovenia 35th; North Macedonia 82nd; Bosnia and Hercegovina 92nd; Montenegro 73rd; Romania 51st; Hungary 47th; Austria 21st; and Germany 7th [8]. From here, it is evident that Serbia is not in a competitive position, even if only the neighboring countries are taken into consideration. In addition, Serbia fell from it 65th position in 2018 to 72nd in 2019. Further, some of the main issues of the domestic economy include the lack of customer sophistication, lack of ability to rely on professional management, inadequate protection of private property, inadequate legal system efficiency and judiciary, low market dominance levels, lack of effective protection of intellectual property and inadequate cooperation between employees and employers. Now, on a positive side, some of the main factors which affected the development of Serbia regarding its competitive ability include low inflation rate, improved transport infrastructure, a relatively stable financial system, and short waiting times when it comes to starting own business [8].

Furthermore, as noted earlier outdated manufacturing equipment with the mean age of 25 years, is a big handicap for enterprises who want to compete on the international market. In addition, foreign investors are investing in Serbia mainly due to its cheap and qualified labor force, thus these investors don't bring modern technologies but rather labor intensive workplaces. Such investments don't affect competitive ability in a significant manner, as they don't focus new technologies, nor research and development. In the next section, the potential of cloud computing technologies is addressed.

II. CLOUD COMPUTING APPLICATIONS IN ENTERPRISES

In the modern business environment, enterprises have to adapt to changes which are brought by dynamic market shifts and the rapid spread of modern ICTs. Cloud-based solutions are more effective compared to traditional enterprise systems as they can integrate various sub-solutions (cameras, sensors, CCTV, wireless networks, application platforms) into one managing platform. Cloud computing represents the distribution of computing resources in accordance of demand. Resources can be in the form of software, hardware, or data processing [9].

The application of cloud computing technologies in domestic enterprises can increase productivity by enabling effective remote work of employees. Cloud computing can increase storage capacity, computing performance and reduce manufacturing costs. Therefore, a cloud-based computing solution can be proved useful for small and medium-sized enterprises (SMEs). Depending on the needs of the enterprise a specific cloud computing solution can be introduced [10].

Software-as-a-Service (SaaS) cloud solutions which host on-demand applications can be applied in enterprise resource planning systems. This may include payment services, invoice automation, central managing software which integrates several other software solutions etc. In addition, SaaS solutions are cheaper, and require a fixed monthly or annual subscription plan which includes maintenance and security as well [11]. Cloud-computing was found to improve collaboration and interaction between employees, customers and suppliers; to reduce maintenance time in the IT department; improve business continuity, backups and catastrophic system failure recovery; scalability of load capacity; access to technical knowledge; and environmental and ecological friendliness [12]. Further, cloud computing can be used for application testing on software development enterprises. In learning enterprises, cloud-based solutions can offer immense scalability just for a fraction of a cost. In the domain of science, cloud-based solutions can offer a central research hubs which increase R&D activities [13].

Cloud computing technologies can be also applied in the domain of human resource management (HRM) in the form of on-demand payments, and alliance and remote work with other SMEs [14]. In the same study it was noted that cloud computing can applied for auditing, videos, e-mail consultation, expert consultation, collaboration, forum discussions. All of these application are available for employees, employers, teams of external services, staff at departments, experts, and globalized institutions. Other integrations of cloud computing can include logistics, distribution, manufacturing, accounting, marketing and finance. Generally, an enterprise shouldn't implement cloud computing solution just for the sake of implementing one. But rather, an organic problem should be defined and evaluated. From here, the most appropriate cloud solution can be implemented. This includes mainly subscribing to cloud computing providers (Amazon, Google, Microsoft Azzure etc.) as these don't require upfront resources and they are maintained by the provider, thus securing high levels of up-time. Some of the less complex cloud-based applications include Gmail, Google Calendar, Google Drive, Dropbox and other apps. Such a scenario lacks the application of advanced cloud computing solutions.

Additional applications in enterprises can be in the form of enterprise collaboration which can help improve logistics, and reduce distribution costs; developing clusters for distributing valuable information and knowledge in order for every enterprise to achieve its maximum on the market; developing databases with the goal to effectively improve timely decision making; developing a platform for enterprise networking; implementing applications for real-time purchase orders; reducing bureaucratic procedures and real-time sale order following. Now, what can be achieved with cloud computing and its application in domestic enterprises and how can it affect competitiveness?

Cloud-based solutions by reducing operational costs and by increasing productivity opens the doors to a more competitive pricing strategy as well as room for improvement in the quality department. Some of the noted cloud-based solutions, such as cluster formation and platform for enterprise networking can increase the potential of innovations which further can increase competitiveness.

Cloud computing is not necessarily an innovation as a technology, but rather the innovation part comes from its capability to integrates various services (platforms, software, hardware etc.) and it can be applied for user interconnected to a main virtual computing environment where every user has access to specific resources in accordance with network allocated permissions [15]. This indicates, that the main "power" of cloud computing is not the infrastructure it lies on, but rather the vast amount of features which are available for enterprises. Certainly, virtualization and grid computing are two fundamental concepts on which cloud computing is based [16].

From here, it can be argued that domestic enterprises should consider implementing some of the noted cloud-based solutions which have the potential to increase productivity, quality and reduce operating costs. As it was mentioned earlier, cloud computing has an indirect influence on innovation development. In order for cloud-based enterprise solutions to have a significant impact on competitiveness of the domestic economy, the implementation and application of cloud-based technological improvements should be systematic on a national scale. This is because one or several enterprises are simply not enough to create a broader movement towards these modern ICTs solutions.

Now, how far have domestic enterprises come when it comes to cloud-based solutions? Unfortunately, the data is scarce in this domain. The only reliable data is from 2017, and this additionally raises concerns about how cloud-based solutions are managed on a national level. According to Eurostat [17] in Serbia 15% of SMEs apply cloud computing services over the Internet (26% in the EU); 5% of SMEs use e-mail cloud computing (18% in the EU); 4% of Serbian SMEs use cloud computing software (14% in the EU); only 5% of SMEs apply cloud-based data storage (18%); and only 3% SMEs use a cloud-based computing software. According to this data, domestic SMEs lack behind enterprises in the EU when it comes to cloud-based solution in an office setting. In addition, it can be assumed, that the lack of cloud-based solutions in manufacturing enterprises are even more pronounced. A recap and discussion of potential future trends regarding national competitiveness from the aspect of cloud computing implementations is given in the next section.

III. Competitiveness Scenarios Regarding Cloud Computing Application in Enterprises

When it comes to competitiveness of the domestic economy, four distinct scenarios can be discussed when it comes to achieving future competitiveness ranks on the international market. The first scenario (Scenario A) is when a large percentage of SMEs would implement advanced cloud-based solutions. This would include the development of R&D clusters where SMEs can share knowledge, and effectively distribute valuable information from which every participant in the cluster can benefit in the form of business performance or market performance. In this scenario national competitiveness ranks can experience noticeable improvements. This scenario is the least probable, as the matter of wide adoption of cloud-based solutions depends on two factors which are relatively slowly executed: the first factor is government legislation and support, and second is proneness to change and improvement of enterprises.

The second scenario (Scenario B) is also a positive one. Here, the impact on the competitiveness rank is not so high. However, this scenario doesn't rely heavily on wide adoption of cloud-based solutions, but rather it requires moderate adoption rates. Therefore, the probability rate is higher. This scenario would require from SMEs to consider the first steps of integrating and applying modern ICTs, which includes cloud computing as well.

The third scenario (Scenario C) is neutral result, where there is no change in competitiveness rank over the next several years. This would likely happen if no advancement in modern ICTs implementation and application is introduced. A "no change" scenario indicates a halt in innovation, R&D and overall the lack of market penetration potential. Such scenario lack the application of advanced cloud computing solutions.

Finally, the fourth scenario (Scenario D) is a result of downgrading existing ICT infrastructures. Now, this scenario can be the bi-product of a "no change" scenario as well, due to the rapid development of new technologies and as old technologies are becoming obsolete. In the next section guidelines and suggestions for competitiveness improvement from the aspect of modern ICTs and cloud-based solutions are discussed.

IV. Guidelines and Suggestions

Based on the analyzed literature in the domain of cloud computing applications and its potential to improve competitiveness, the following guidelines and suggestions for improving business performance and competitiveness of domestic enterprises are proposed:

- domestic enterprises should consider applying not only cloud-based solutions but other complementary ICTs which alongside cloud computing can improve various business activities (manufacturing, logistics, distribution, accounting etc.). Cloud-based solutions can offer real-time maintenance and control over production processes.
- modern ICTs should be used for innovation as well as new product development. Cloud-based platforms with integrated applications for networking and data-sharing between SMEs could lead to new partnerships and the creation of value with cumulative efforts of all parties.
- developing cloud platforms for enterprise networking can help increase the number and size of clusters;
- application of cloud-based solutions should be considered on a national level as such

technologies have the potential to increase national competitiveness. An effective e-government platform would speed-up bureaucratic procedures for getting business licenses, taxing, and business reporting. There are a lot of cloud-based solution providers who specialize in e-government and crisis-level data back-ups. An effective bureaucratic system is important metric for foreign investors and for local individuals and groups as well, who want to start their business.

In sum, it is evident that domestic enterprises have to adapt to the changes of the digitalized and globalized market. Cloud-based technologies may offer relatively cost-effective solutions for improving business performance, which may further translate into improved competitiveness.

V. Conclusion

The globalization of markets and the fourth industrial revolution has brought challenges to enterprises, especially to domestic enterprises when it comes to achieving and sustaining a competitive position on the market. The lack of productivity, quality, modern manufacturing equipment, and modern management technique and tool application, "handicaps" domestic enterprises on the international market. This paper provides a concise overview on the potential of cloud-based solutions in domestic enterprises. It can be concluded, that domestic enterprises have to adapt to the changes on the globalized market. As it was discussed, cloud-based solutions can be a relatively low-cost and effective solution. Certainly, these technologies are an adequate “step” into the right direction when it comes to applying modern ICTs.

Now, the main limitation of this paper is the lack of survey data regarding cloud computing application in domestic enterprises. However, the paper provides adequate insight into this domain and offers a good starting point for future research. For this future research it is recommended to conduct a meta-analysis of existing studies and to compare the effectiveness of various cloud-based solutions in specific enterprise environments.

References

[1] D. Đorđević, D. Ćoćkalo, M. Bakator, S. Bogetić, M. Vorkapić, and C. Bešić, C., “Cloud Computing Model for Achieving Competitiveness of Domestic Enterprises” In Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing, pp. 239-249, 2020. Springer, Cham.

[2] Y. Lu, and X. Xu, "Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services," Robot. and Comp.-Integ. Manu., vol. 57, pp. 92–102, 2019. doi:10.1016/j.rcim.2018.11.006,

[3] M. Vidas-Bubanja, S. Bogetić, and I. Bubanja, "International standards: An important component of a successful digital transformation of the national economy," Jou. of Eng. Manag. and Comp. (JEMC), vol. 9, no. 1, pp. 72-81, 2019.

[4] M. Arsić, "Dugoročne posledice ekonomskog sloma privrede Srbije tokom 90-ih godina 20-og veka," Ekonomske ideje i praksa, vol. 22, pp. 25-41, 2016.

[5] Y. Liu, X. Xu, L. Zhang, L. Wang, and R.Y. Zhong, "Workload-based multi-task scheduling in cloud manufacturing," Robot. and Comp.-Integ. Manu., vol. 45, pp. 3–20, 2017. doi:10.1016/j.rcim.2016.09.008

[6] M. Kovačević, "Uzroci deindustrijalizacije i ograničenja za reindustrijalizaciju. U knjizi Deindustrijalizacija u Srbiji - Mogućnosti revitalizacije industrijskog sektora," Beograd: Institut ekonomskih nauka, pp.11-44, 2014.

[7] M. Radulescu, A. Fedajev, C. Sinisi, C. Popescu, and S. Iacob, "Europe 2020 Im-plementation as Driver of Economic Performance and Competitiveness. Panel Analysis of CEE Countries," Sustainability, vol. 10, no. 3, p. 566, 2018. doi:10.3390/su10020566

[8] World Economic Forum (WEF). "The Global Competitiveness Reports 2019," Available at: http://reports.weforum.org (Accessed on 10.04.2020.)

[9] P. Dašić, J. Dašić, and B. Crvenković, "Primena video nadzora kao cloud servisa (VSaaS) u rudarstvu," U: Zborniku radova, vol. 2, pp. 22-23, 2017.

[10] C. Keesookpun, and H. Mitomo, "Cloud computing adoption and determining factors in different industries: a case study of Thailand," Int. Jou. of Inf. Tech. and Man., vol. 13, no. 4, p. 243, 2014. doi:10.1504/ijitm.2014.065631

[11] S. L. Saini, D. K. Saini, J. H. Yousif, and S. V. Khandage, "Cloud computing and enterprise resource planning systems," In Proceedings of the world Congress on Engineering, vol. 1, pp. 681-684, 2011.

[12] E. O. Yeboah-Boateng, and K. A. Essandoh, "Factors influencing the adoption of cloud computing by small and medium enterprises in developing economies," Int. Jou. of Emerg. Sci. and Eng., vol. 2, no. 4, 13-20, 2014.

[13] A. Sulistio, C. Reich, and F. Doelitzscher, (2009). Cloud Infrastructure & Applications – CloudIA. Cloud Computing, 583–588, 2009. doi:10.1007/978-3-642-10665-1_56

[14] X. L. Wang, B. Z. Wang, Y. Y. Li, and Y. Xu, "Cloud computing in human resource management (HRM) system for small and medium enterprises (SMEs)," The Int. Jou. of Ad. Manu. Tech., vol. 84, no. 1-4, pp. 485–496, 2016. doi:10.1007/s00170-016-8493-8

[15] N. A. Sultan, "Reaching for the “cloud”: How SMEs can manage," Int. Jou. of Inf. Man., vol. 31, no. 3, pp. 272–278, 2011. doi:10.1016/j.ijinfomgt.2010.08.001

[16] N. Carr, "The Big Switch: Re-Wiring the World, from Edison to Google," New York & London: W.W. Norton, 2009.

[17] Eurostat, "Cloud computing services use in SMEs," 2018 Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_cic ce_use&lang=en (Accessed on 12.13.2019.

Applications of Convolutional Neural Networks

S. Mihajlović*, D. Ivetić**, I. Berković*
* Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
** Faculty of Technical Sciences, Novi Sad, Republic of Serbia
mihajlovicsinisa25@gmail.com, ivetic@uns.ac.rs, ivana.berkovic62@gmail.com

Abstract - This paper describes the role and architecture of a convolutional neural networks. CNNs are widely used in computer vision and natural language processing. In computer vision, it is mainly used for face recognition, scene labelling, image classification, action recognition, and human pose estimation. In the field of speech, it is used to recognize and classify text for natural language processing.

I. Introduction

One of the most popular algorithms for deep learning is convolution neural network (CNN). Convolutional networks are modeled on the visual system in humans, in terms of connections between neurons used in CNN [1]. The work of Hubel and Wiesel has shown the action of the cells of the visual cortex of the cat's brain on various stimuli. Hubel and Wiesel found that cortex neurons that are close to each other are activated in the presence of lines at one angle, while another group of neurons is activated when the angle is changed. Also, some groups of neurons recognize edges, depending on the angle at which they are in the receptive field. Their work was an inspiration in modeling convolutional neurons network [3].

CNNs are a biologically inspired version of the Multilayer Perceptron (MLP). According to their architecture, CNNs can be classified into deep neural networks, to which the term Deep Learning (DL) is associated. DL is a relatively new area of machine learning, which is expanding. After many years of stagnant development in the field of AI, deep neural network architectures, which can be applied in various domains, have shown great importance and contributed to solving seemingly, for one machine, unsolvable problems. CNN consists of one or more convolutional layers and optionally one or more Fully Connected (FC) layers, which can be found in conventional multilayer neural networks. CNNs are designed to take advantage of working with 2D structures, such as images or speech signal inputs, while recent studies show that they achieve significant results with 3D structures as well, even at the expert level. CNNs are the basis for the implementation of many state-of-the-art solutions in the field of computer vision (Computer Vision - CV). Deep CNNs are considered to have become the "mainstream" and subject of research efforts and have been observed to have achieved significant advantages in various tests [2]. The name "convolutional neural network" indicates examples of mathematical operations called convolution. Convolution is a special type of linear operator that convolutional neural networks apply on at least one layer [3]. Convolutional filters are used to sharpen and blur images, as well as to detect edges. From the aspect of CNNs, the application of edge detection algorithms, which detect objects in the image, is important. Where classical neural networks show their disadvantages, CNNs show their advantages. For example, when images are used as input data, an image measuring 200 × 200 pixels is converted into 40,000 neurons in the input layer, which is a huge number and it is practically impossible to train fully connected neural networks. For a color image, 3 channels representing the base colors (RGB) should also be considered, giving 120,000 neurons in the input layer. The idea of CNNs is to lay more layers to reveal the essential properties of the input data. In line with this idea, convolutional filters are applied to the image to extract useful features and create their maps. By applying filters to the input image, in addition to detecting significant characteristics, the resolution is also reduced. The disadvantage of CNNs is the need for significant hardware resources, as well as the fact that previous models have been trained with a large number of images, but recent results suggest that high performance can be achieved with a relatively small number of samples [2].

II. CNN Architecture

The following paper describes the architecture of convolutional neural networks. There may be a monochrome or color image at the input of the convolutional neural network. If it is a color image, it displays it in three dimensions: height, width, and depth, where the depth dimension corresponds to the color channel (red, green, blue). A convolutional neural network can be represented as a series of layers, with each layer transforming the result of the previous layer using certain differentiable functions and producing an output called a feature map [3].

Convolutional networks are trainable multistage architectures with each stage consisting of multiple layers. The input and output of each stage are sets of arrays called as feature maps. In the case of a colored image, each feature map would be a 2D array containing a color channel of the input image, a 3D array for a video and a 1D array for an audio input. The output stage represents features extracted from all locations on the input. Each stage generally consists of a convolution layer, nonlinearity, and a pooling layer. A single or multiple fully connected layers are present after several convolution and pooling layers [4]. Figure 1 shows the CNN architecture for character recognition in an image.

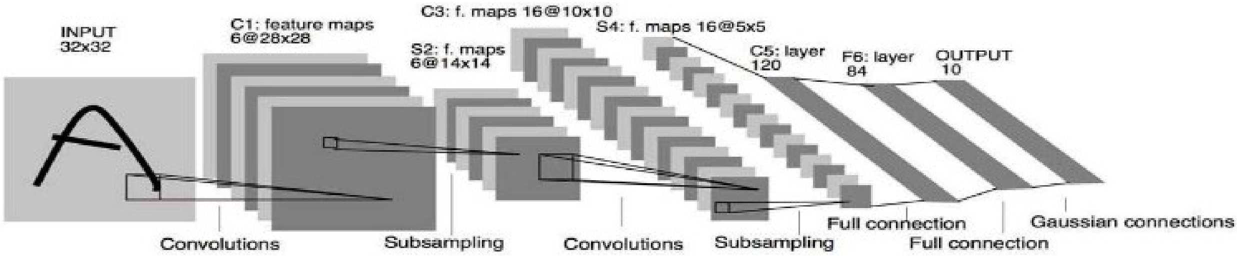


Figure 1. CNN architecture [16]

A. Convolution Layer

The difference between CNN and other types of neural network architecture is reflected in the layer where the convolution takes place, after which it got its name. The convolutional layer is the basic building block of CNN and performs many very demanding calculations. The convolutional layer realizes the basic operation of training network neurons. In its most general form, convolution is an operation of two functions, whose arguments are real values [2].

The convolutional layer is the basis of convolutional neural networks and performs most of the data transformation. It usually consists of two parts. In the first part, the input activation folder, which represents the image in the first layer of the network, enters convolution with a certain number of filters which recognize different characteristics. After that, in the second part, one of the activation functions acts on the obtained result. The most effective action on various tasks showed the ReLU activation function. Each convolutional layer contains a pre-selected number of filters, the so-called fixed core dimensions, which store the values of weight W. The convolution of a filter with a feature map takes place only at spatial coordinates width and height of the feature map and represents Hadamard's product of filter and small part of the input image, the so-called of the receptive field. In the case of a color image, the same filter is generally applied to each color channel, and the results are summed by the feature map channels. The convolutional layers at the beginning of the network learn to recognize details and structures. The first layer of the network can learn to recognize the edges of objects in images. The deeper the convolutional layer is in the neural network, the more it learns to recognize more complex features. Thus, the second layer of the network can recognize simple geometric shapes, while the last layer can recognize faces, dogs, cars, and the like. The weight values stored in the filters represent the parameters of the model and can be learned by the network [3].

B. Non-linearity Layer

The purpose of the normalization layer is to speed up the training process. In general, the values of input data to the network need to be normalized, and this in the case if the values of one type of data are in the range of 0 to 1, while the values of the second data type range from 0 to 1000. In this case, the values of these two variables are required reduce to the same range. The normalization layer is within the convolutional neural networks added after the convolutional and ReLU layers. Reason it has been shown that the training process can be accelerated if we also perform normalization on the activations of individual hidden layers (i.e. normalization of our property maps) and not just over the input pictures. Normalization is done in such a way that all property folders subtract their mean and then each of the values in divide these maps by the standard deviation of those property maps. Of course, this normalization is done only for the purpose of value regulation weight values of individual layers, and as such should not be further propagated through the network. For this purpose, a layer of normalization on its own it also gives the output the parameters γ and β that the network needs to learn. Parameter γ it is necessary to multiply all property folders. The parameter β is added to the result of this operation [1].

C. Pooling Layer

Often a pooling layer is added between successive convolutional layers. The role of the pooling layer is to reduce the dimensionality of the matrices obtained by various convolutions, and thus the number of parameters which the network needs to learn [1]. The two most common types of pooling are max-pooling and average-pooling. Max-pooling with a filter size 2 × 2 and step 2, pooling takes place by taking the local maximum value of the pixel on receptive field. Figure 2 shows an example of max-pooling operation with a 2 × 2 filter and step 2.

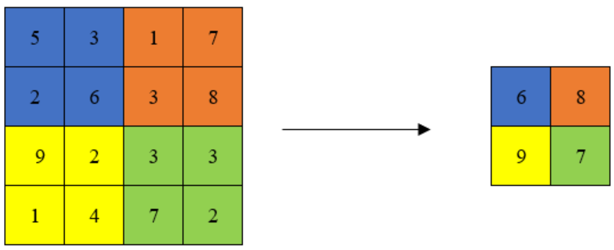


Figure 2. Illustration of max-pooling operation with a 2 × 2 filter and step 2

Average-pooling with a filter of size 2 × 2 and step 2, pooling takes place by taking the arithmetic mean of the values of all elements of the receptive field. Figure 3 shows an example of average-pooling operation with a 2 × 2 filter and step 2. Figure 3 shows an example of average-pooling operation with a 2 × 2 filter and step 2.

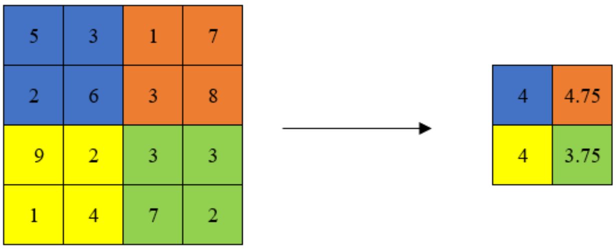


Figure 3. Illustration of average-pooling operation with a 2 × 2 filter and step 2

The pooling layer is very useful in preserving invariance towards small one's pixel translations. The invariance itself to translation means yes, if input we translate the image for some small amount, the value of the output after pooling it will not change significantly. This feature has proven to be very useful when it matters if a feature is in the image, but it does not matter where it is results [3].

D. Fully Connected Layer

The fully connected (FC) layer is located at the very end of the convolutional neural network. It is composed of neurons completely connected to the previous 3D volume, as is the case with classical neural networks. Each neuron is connected to each element of the activation map from the previous one layer. In practice, several FC layers at the very end of the neuron are often used networks. The fully connected layers are followed by the result, usually shown in vector form [3].

III. Applications of CNN

A. Computer Vision

Convolutional neural networks are used for identification. Instead of entering the image into the neural network as a single grid of numbers, it is divided into overlapping tiles of the image, and each of them is drawn into a small neural network.

CNNs can be trained in multiple phase architectures, with the inputs and outputs of each phase consisting of sets of arrays called feature maps. If the image entry is in color, each feature map is a 2D array that contains the color channel of the input image with the video or a volumetric image that is a 3D array. Each extracted feature at all locations at the input is represented by a feature map at the output. Each phase consists of a filter bank layer, a non-linearity layer, and a feature pooling layer. CNN consists of one, two, or three such three-layer phases, followed by a classification module.

Face recognition: there are several facial recognition issues:

- face recognition in images.
- focusing on each face and if the quality is low or the face is shown in different poses.
- recognizing unique characteristics comparing the recognized characteristics with those in the database and determining the name of the person.

Faces represent a complex, multi-dimensional, visual stimulus which was earlier presented using a hybrid neural network combining local image sampling, a self-organizing map neural network and a convolutional neural network. The results were presented using Karhunen-Loe`ve transform in place of the self-organizing map which performed almost as well (5.3% error versus 3.8%) and a multi-layer perceptron which performed poorly (40% error versus 3.8%) [6]. Google's FaceNet and Facebook's Deep-Face are based on CNNs. DeepFace models the face in 3D and aligns it to look like a frontal face. The normalized input is fed into one convoluted pooling filter followed by three locally connected layers and two fully connected layers. DeepFace achieves excellent performance, its representation is not easy to interpret because the faces of the same person are not necessarily grouped during the training process. FaceNet defines the triple loss function on the presentation, which makes the training the learning process learn to group the represented persons of the same a person [7].

In the area of facial recognition, CNNs could be used for distance learning. The role of CNNs would be to recognize the facial expressions of students, and thus analyze the satisfaction and understanding of students of what is taught based on whether they are smiling, frowning, amazed and the like.

Scene labelling: each pixel is marked with the object category sign to which it belongs in the scene marking. Clement Farabet and co-workers proposed a method that uses multiscale a convolutional network that brought record accuracy to Sift Flow Dataset (33 classes) and Barcelona Dataset (170 classes) and almost record accuracy at Stanford Data set in the background (8 classes). Their method produced 320 × 240 image tagging in less than a second, inclusive separation of properties [8]. The recurrent architecture for a convolutional neural network suggests a sequential array of networks that share the same set of parameters. The network automatically learns to smooth out its own intended labels. As the size of the context increases with built-in iteration, the system identifies and corrects its own errors. A simple and scalable detection algorithm that pleases mean average precision (mAP) by more than 30% called R-CNN: Regions with CNN characteristics are combined region suggestions with CNN characteristics [9].

Image classification: Because CNNs have a common learning ability and ability classifier, they produce better classification accuracy compared to other methods when operating on large-scale datasets. Several significant works have been done to improve the accuracy of classifications by reducing the filter size or expanding the mesh depth. Multi-column deep neural networks (MCDNN) can outperform all image classification methods and demonstrate that pre-training is not required while reducing the error rate by 30-40%. Unsaturated neurons and efficient GPU implementation of convolutional operation resulted in a winning rate of top 5 test errors. Hierarchical deep convolutional neural networks (HDCNN) are based on the intuition that some classes in image classification are confusing than other classes. It builds on conventional CNNs that are N-way classifiers and follows the module for classification strategy and design. Image classification systems dealing with fine-grained images are based on the concept of unique characteristics. Categorization can be performed with at least monitoring settings where only the class label is given. This is accomplished using the attention of a CNN trained for the categorization task [5].

Action recognition: There are difficulties in developing action recognition systems, such as translations and distortions of characteristics in different patterns belonging to the same action class. The three-dimensional receptive field structure of the modified CNN model provides translation invariant feature extraction capability, and the use of shared weight also reduces the number of parameters in the action recognition system [10]. At Stanford University, they proposed the development of standard methods used in SIFT-based visual recognition, proposed by Lowe, and HOG proposed by Dalal and Triggs, using the Independent Spatial Analysis (ISA) algorithm that is an upgrade of the Independent Component Analysis (ICA) used in natural image statistics. The ISA algorithm to learn invariant spatiotemporal characteristics from uncategorized video data was applied on the Hollywood2 and YouTube action datasets and it resulted in 53.3% and 75.8% accuracy percentage, respectively [11]. Temporal pyramid pooling is based on CNN, which is used to recognize action, reduces the risk of missing different frames, and requires less data training and gives better results. The position-based convolutional neural network (PCNN) descriptor is used to recognize human actions in videos [12]. By performing 3D convolutions, the 3D CNN action recognition model extracts functions from the spatial and temporal dimensions, and records motion information encoded in multiple adjacent frames. An example in a real-world environment where 3D CNN outperforms 2D CNN is intelligent video surveillance, customer attributes, and shopping behavior analysis. The 3D CNN model is most effective when the number of positive samples is smaller [5]. To enable higher performance, calculations must be increased to allow for large data sets and to speed up model training. 3D models of deep learning with extended connectivity using CNNs are used. This is achieved by using multi-core CPUs and GPUs by achieving parallelism of data and models through parallel model preparation. It has been found that 3D CNN code is best scaled on processors if the convolution step is applied by a highly parallel FFT-based method, and thus GPU-like performance is achieved using OpenMP [13].

Human pose estimation: The goal of human posture assessment is to determine the position of human joints relative to images, image sequences, depth images, or skeletal data, as provided by motion capture hardware. Due to the wide range of human silhouettes and appearances, the difficult lighting, and the crowded background, the assessment of human is a very challenging task. The assessment of posture before the era of deep learning was based on the discovery of body parts, for example using pictorial structures. Using deep learning methods in assessing human posture, they can be divided into holistic and partial methods depending on the way the input images are processed. Holistic processing methods accomplish their task in a global way and do not explicitly define a model for each individual part and their spatial relationships. Holistic methods are inaccurate in the high-precision region due to difficulties in learning direct regression of complex pose vectors from images. Part-based processing methods focus on the individual discovery of parts of the human body, followed by a graphical model that includes spatial information. Instead of training the network using the full image, local work patches and CNN training background patches are used to learn the conditional probability of work presence and spatial relationships. Multiple resolution CNN is designed to perform heat map probability regression for each body part, followed by an implicit graphical model to further improve joint consistency [14].

B. Natural Language Processing

Speech is a series of signals, and its recognition is one of the most important tasks in NLP. Currently, CNN is used to classify sentences, categorize topics, sentiment analysis, and many other tasks.

Speech recognition: CNNs have been used in speech recognition and have given better results than deep neural networks (DNNs). Domains in which CNN gives better results than DNN are:

- Noise robustness,
- Distant speech recognition,
- Low-footprint models,
- Channel-mismatched training-test condition [15].

When the researchers applied a CNN on 1000 hours of Kinect distance, they obtained a 4% Word Error Rate Reduction (WERR) compared with a DNN on a similar size. Microsoft developed Kinect, and it introduces a series of motion sensors that allow users to interact with their computers through character and voice commands. Kinect distance is the distance that the device supports which is 1.2 to 3.5 m. The polling should be done on the local frequency region to increase the robustness of the CNN. Fewer parameters are used to extract low-level features to avoid over-fitting. Automatic speech recognition (ASR) is a technology that converts human speech into spoken words [5].

Distant Speech can be captured by either a Single Distant Microphone (SDM) or Multiple Distant Microphone (MDM). Some of the problems are multiple audio sources and continuous noise. Accurate results in Distant Speech were obtained using the Recurrent Neural Network (RNN). The incorporation of the attention mechanism in the RNN contributed to the result being obtained in one step. Combining the RNN and CNN approaches gives a better output. CNN is used for frame level classification and RNN is used with Connectionist Temporal Classification for decoding frames in a sequence of phenomena. Speech Emotion Recognition (SER) has an important application in human-centered signal processing. CNN is used to learn the prominent features of SER. CNN was trained for SER in two phases. In the first stage, local invariant features (LIF) are learned using sparse auto-encoder (SAE) and in the second stage, LIF is used as an input to salient descriptive feature analysis. The developed system proved to be robust and stable in complex scenes. Handling unwanted noise is a major challenge in speech recognition. Researchers have built systems across years that are not affected by noise signals. The finest pooling, padding, and input feature map selection techniques were deployed [4].

Text classification: The key task for natural language processing (NLP) is text classification. In their natural form, sentences have a complex structure, both sequential and hierarchical, which are essential for their understanding. CNN has achieved top performance in sentence modeling. An appropriate CNN architecture is important for text classification. One convolution layer is used for sentence modeling, but there is a possibility of several layers of convolution. Multiple convolutional layers help to single out high-level abstract features. The pooling operation helps the network to cope with variable sentence lengths. Max-pooling is used to retain the most important information that makes up the sentence. Max-pooling cannot distinguish whether a characteristic occurs in one of the rows only once or more times and ignores the order in which the characteristics occur [14]. There are methods of embedding words such as Word2vec and using 100 billion words from Google news that are publicly available, Word2vec is trained. In NLP, the main features of CNNs, location invariance and local compositionality, do not apply like they do in computer vision applications. In a sentence, the place of the word is very important, while in the case of pixels, those that are close to each other can belong to the same object and can be connected. In sentences this is not true because words that are close to each other do not have to have the same meaning, and therefore they may not be related [5].

In the field of text classification, CNNs could be used to identify knowledge gaps by performing an analysis based on test results. After the analysis of the tests, it would be suggested what is needed to learn better, as well as additional literature, for the result to be better.

IV. Conclusion

Convolutional networks are modeled on the visual system in humans. It represents the connection between neurons used in CNN, and according to their architecture can be classified into deep neural networks. CNNs consist of one or more convolutional layers and are designed to work with images or speech signal. The disadvantage of CNN is the need for significant hardware resources, as well as that previous models are trained with many images. Convolutional networks are trainable multistage architects with each stage consisting of multiple layers. The input and output of each stage are sets of arrays called as feature maps.

The convolutional layer realizes the basic operation of training network neurons. In its most general form, convolution is an operation of two functions, whose arguments are real values. The purpose of the normalization layer is to speed up the training process. Often a pooling layer is added between successive convolutional layers. The role of the pooling layer is to reduce the dimensionality of the matrices obtained by various convolutions, and thus the number of parameters which the network needs to learn. The fully connected layer is located at the very end of the convolutional neural network. It is composed of neurons completely connected to the previous 3D volume, as is the case with classical neural networks.

CNNs are widely used today and their importance is great. CNN has wide application in computer vision as well as in natural language processing. It has been shown that the accuracy of classification in newer, very deep CNNs is very high. One of the consequences of that fact is the application of deep CNNs in medicine, but at the level of research and pilot studies for now. The power of recognizing tumor tissues using CNNs is at approximately the same level as that of an expert. Problems, such as detecting lies from the respondents' faces, are something that researchers are intensively dealing with. Many large companies invest heavily in research in the field of DL / CNNs. Deep architectures have proven useful even for solving some specific problems, such as autonomous playing of computer games by a machine or recognizing the activities of people from video. The whole development of DL domains is accompanied by the development of very powerful computers, whose performance is growing day by day, which is crucial for the future work of researchers in this field.

Acknowledgment

Ministry of Education, Science and Technological Development, Republic of Serbia financially supported this research, under the project number TR32044: “The development of software tools for business process analysis and improvement”, 2011-2020.

References

[1] E. Šemšić, Korištenje konvolucionih neuronskih mreža za potiskivanje šuma na slici, Elektrotehnički fakultet u Sarajevu, februar, 2019.

[2] M. Dabović, I. Tartalja, Duboke konvolucijske neuronske mreže - koncepti i aktuelna istraživanja, Konferencije za elektroniku, telekomunikacije, računarstvo, automatiku i nuklearnu tehniku, ETRAN 2017, Kladovo, 05. do 08. juna 2017.

[3] K. Džomba, Konvolucijske neuronske mreže, Zagreb, jul, 2018.

[4] A. Bhandare, M. Bhide, P. Gokhale, R. Chandavarkar, Applications of Convolutional Neural Networks, International Journal of Computer Science and Information Technologies, Vol. 7, No. 5, pp. 2206-2215, 2016.

[5] A. S. Shamsaldin, P. Fattah, T. A. Rashid, N. K. Al-Salihi1, The Study of The Convolutional Neural Networks Applications, UKH Journal of Science and Engineering, Vol. 3, No. 2, pp. 31-40, 2019.

[6] Lawrence, Steve, et al, Face recognition: A convolutional neural network approach, IEEE transactions on neural networks, Vol. 8, No. 1, pp. 98-113, January 1997.

[7] A. Voulodimos, N. Doulamis, A. Doulamis and E. Protopapadakis, Deep Learning for Computer Vision: A Brief Review, Hindawi, Computational Intelligence and Neuroscience, February 2018.

[8] Farabet, Clement, et al., Learning hierarchical features for scene labeling, IEEE transactions on pattern analysis and machine intelligence, Vol. 35, No. 8, pp. 1915-1929, 2013.

[9] P. O. Pinheiro and R. Collobert, Recurrent Convolutional Neural Networks for Scene Labeling, ICML, 2014.

[10] K. Ho-Joon, J. S. Lee and Y. Hyun-Seung, Human action recognition using a modified convolutional neural network, International Symposium on Neural Networks, Springer, Berlin Heidelberg, 2007.

[11] Q. V. Le, et al, Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis, Computer Vision and Pattern Recognition, pp. 3361-3368, 2011.

[12] P. Wang, et al, Temporal pyramid pooling based convolutional neural networks for action recognition, arXiv preprint, April 2015.

[13] M. S. Rajeswar, et al, Scaling Up the Training of Deep CNNs for Human Action Recognition, Parallel and Distributed Processing Symposium Workshop (IPDPSW), pp. 1172-1177, 2015.

[14] J. Gu, et al, Recent Advances in Convolutional Neural Networks, Pattern Recognition, Vol. 77, pp. 354-377, May 2018.

[15] J. Huang, J. Li and Y. Gong, An analysis of convolutional neural networks for speech recognition, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4989-4993, 2015.

[16] S. Cheng, Convolutional Neural Networks - Deep Learning Lecture 4, School of ECE, University of Oklahoma, January 2017.

Deepfake videos and their Impact on Privacy and Security

Mihalj Bakator, Dragica Radosav
University of Novi Sad, Technical Faculty "Mihajlo Pupin", Đure Đakovića bb, 23000, Zrenjanin, Republic of Serbia
mihalj.bakator@uns.ac.rs; dradosav@tfzr.rs

Abstract - Modern digital technologies have advanced at a tremendous rate. There are a lot of positive socio-economic outcomes from the emergence of advanced ICTs. However, there are certainly negative sides of them as well. Through generative adversarial networks, sophisticated video editing, where individuals faces are swapped, have become available to the average computer user. These deepfake videos are becoming more and more convincing for the untrained eye, and if the development of this technology advances further, soon the naked eye would not be able to identify if a video is real or not. This kind of video editing "power" is often used to harm individuals, groups or even countries through fake news, reports, "real" video footages. Unfortunately, deepfake technology is rather used in an unethical manner, or better say, it harms people, and more often than not, these are criminal activities which can't be fully stopped. In this paper the emergence of deepfake videos and their harmful effect on privacy and security is addressed.

I. Introduction

Modern information-communication technologies (ICTs) have improved a large number of business and social-economic constructs. The application of modern ICTs in the business world has a tremendous impact on how modern business is conducted [1]. From a social-economic standpoint, the rise of social media and other Internet-based applications has brought people closer and increased the number of interconnections between individuals and/or groups. Through these interconnections the distribution and sharing of various information has intensified in an astronomical manner. Alongside with the numerous positive aspects of modern ICTs and their rapid development, there are certainly negative sides as well. Unethical, morally wrong, and criminal activities are on the rise on the Internet. Among the wide variety of malicious actions on the web, a fairly new digital technology product has emerged which can and has a large negative impact on privacy and security. This digital technology is based on machine learning, and it is knowns as deepfake videos - hyper realistic videos edited with artificial intelligence and can depict someone do or say things which have never occurred. Such videos have emerged in the form of fake news, and explicit videos which are harmful for society and human democracy. Deepfake videos appeared on the Internet in 2017, and as technology progressed and consumer-grade computers had more and more computing power, such deepfake videos became more and more sophisticated [2]. There are videos and images produced which are so error-free that the average untrained eye (and even trained eyes) can't detec if the video or image is fake or manipulated.

There are evidence-based indications that deepfake videos that distribute false information may have serious consequences and complications in society. Some of the main concerns of the negative impact of deepfake videos are in the tourism industry in the form of destination and brand image deterioration as fake news can lead to massive word-to-mouth negative reviews. Additionally, various hoaxes can spread and affect almost any individual, group, country or any other subject or form which are news-worthy [3].

The basis of deepfake videos is artificial intelligence, and it consists of mainly swapping faces of an individual with another individual through a generative adversarial network algorithm. This technique is used against other people through distributing fake content in the form of fake news or other type of videos. A few years prior processing videos with deep algorithms was not easily manageable and accessible to everyday computer users. As technology progressed, easy to use applications for creating deepfake videos have emerged [4]. These applications are used to create harmful video content, hence ethical and moral questions arise. This issue of deepfake videos affects entrepreneurs, politicians, managers, celebrities as well as other professions if they are targeted. These attacks can take a large tool on emotional and social quality of life of the victim. Even economies are at risk, as fake news can affect various industries.

In this paper the negative impact of deepfake videos on security and privacy is analyzed. The main goal of this paper is to review and concisely present the burning issue of deep learning technology use in video editing and malicious actions, which can affect individuals and groups from a financial, emotional or social aspects. The paper consists of three main sections (excluding the Introduction and Conclusion sections). The first section addresses the technology behind deepfake videos and discusses its advancement. The second section addresses the impact of deepfake videos on privacy and security and its micro and macro effects on society. In addition, potential methods for "fighting" deepfake videos are discussed. In this way a concise overview is developed and an adequate basis for future research is created. Finally, in the third section future trends regarding the challenges which are brought by deepfake videos are discussed. After the three main sections, conclusions are drawn and guidelines for future research are proposed.

II. Deepfake videos and the technology behind them

Generative adversarial networks (GANs) as a tool for automated video and audio editing allows the creation of high quality tampered video content, which is distributed on social media [5]. The main technology behind deepfakes is deep learning, where deep neural networks (DNNs) reminiscent brain neurons with their interconnectedness. These artificial neurons with the DNN individually can process simple calculations and operations, but with their interconnectedness, these artificial neurons are capable of complex computations which may recognize person's faces from pixels. Now, if a DNN is untrained, it would correctly identify a face only by chance. However, a trained DNN has an improved capability when it comes to recognizing faces. In order for a DNN to be effectively trained a large set of training data is required. That is why the most frequent targets of deepfakes are celebrities [6]. The main difference between deep learning-based video manipulation from other tools and techniques, is that deepfake videos are way more photorealistic. There is an increasing percentage of newly developed videos and are harder and harder to identify if they are fake or not. In addition, there are easy to use apps which make it simple for the average computer user to create deepfake content [7, 8)]

Further, the main deep architecture which identifying a face is called an autoencoder which has its sub-parts: the encoder, latent space and decoder. Particular facial characteristics are compressed into thousands of pixels and the encoder transfers the image into these compressed pixels into the information bottleneck where the autoencoder can learn more facial characteristics rather than memorizing every single input. With deepfake videos, the "trick" with autoencoders is to train two autoencoders - one for each of the two persons. Form here there are several types of deepfakes. These are [9]:

- photo deepfakes (face and body swapping);
- audio deepfakes (voice swapping and text-to speech),
- video deepfakes (face swapping, face-morphing, full body puppetry) and,
- audio + video deepfakes (lyp-syncing).

The above noted content can have negative effects on individuals on a micro level, and on relations between countries on a macro level. The development of deep learning algorithms has certainly affected the Internet and how online content is consumed and perceived. This DNN-based fake content is receiving well-deserved attention from the media, scientists and governments, as there are tremendous risks and threats for individual and national security [10].

It can be expected that DNNs will further develop, and it is matter of time when the majority of deepfake content will be indistinguishable from real content (even with deepfake content detection). This will put pressure on forensic teams for digital crimes and digital content.

III. Impact on privacy and security

Deepfakes can seriously threaten not only individual's security, but also public and national security as well. In 2019 deepfake videos appeared which were so detailed and sophisticated that it was impossible to determine if they were fake or altered [11]. Deepfakes that are intended for malicious actions, so called "dark deepfakes" can be used to execute timely sabotage of an individual, group, company, country or other type of targets. Through these deepfakes public opinion can be manipulated and democracy can be negatively affected if deepfake attacks are aimed at elections [9]. It is concerning that in the digital age, there is no way of securing private information to be collected (purposefully or accidentally) and the processed and used for various marketing, targeting advertising or even harmful content which further leads to phishing, blackmailing and other types of attacks. As noted, fake news and misleading online content can help individuals and groups achieve commercial or political gains in an unethical, illegitimate and criminal way [12].

Other harmful application of GANs is the manipulation of images and placing individuals into explicit content. The subject can be anybody who posts pictures of themselves on social media and can be a target as revenge from someone who has malicious intent. Security and privacy are severely compromised with such technology available to the masses and these apps are even a larger threat as they are relatively easy to use. Now, there are techniques for preventing malicious intent regarding deepfakes, where the algorithms are adversely attacked [13]. Another solution for detecting deepfake videos and photos are so called "Patch&Pair" Convolutional Neural Networks (PPCNNs). This approach slices the image of a face into patches and detect unnatural textures and shapes in local regions of interest, as such regions are often displayed by deepfake images [14].

Why do deepfakes have an impact? Unfortunately, humans don't really care about hard facts and their psychological reasoning is affected by internal motives, objectives and beliefs. Even if a deepfake video is proven fake, sub-beliefs remain, thus these deepfake videos influence the public, regardless if they are proven not real later on [15]. It is evident that deepfakes, and overall artificial intelligence has a tremendous impact on social norms and structures, and affects consumers, customers, viewers and induces the lack of trust and loss of control in democratic situations [16].

Furthermore, as deepfakes will continue to evolve and to advance it is necessary to address them in a preventive manner. This means developing methods which timely and effectively detect fake videos which are created by DNNs. Some of the deepfake video detection algorithms search for inconsistencies with blinking and head posture. Other detection methods involve signal-level artifacts where the synthesized faces are split into video frames, or data-driven methods where DNNs are trained to detect specific artifacts [17]. An another approach utilizes convolutional neural networks (CNNs) to extra features on a frame-level basis. From here, the extracted features are used for training recurrent neural networks (RNNs). These RNNs then learn to detect and classify videos if they are subject to manipulation [18]. This way a semi-prevention approach is introduced. Another approach based on CNNs was able to detect deepfake videos 84% of the time and this was mainly due to glitches and other noticeable artifacts [19]. In the same research, saving and uploading videos into a protected server is a way of saving the integrity of the content.

Further, there is the attribution based confidence (ABC) metric for detecting manipulated deepfake videos. The ABC metric is practical. It does not need the traning data and validation data, only the original videos, and it detects with a confidence of 0.94 [20].

Based on the analyzed literature it is evident that deepfake videos are getting more and more sophisticated, that they pose a threat to privacy and security, and that the underlying technology is advancing at a tremendous rate. This leads to the development of various detection and prevention methods with the goal to ensure trust and integrity of online content and reduce the harmful effects which may arise from deepfake content.

IV. Future trends

Based on the analyzed literature in the domain of deepfake technology and the impact deepfake content has, the following challenges, changes and trends can be expected in the future:

- Deepfake content will evolve into undetectable realistic content.
- Enterprises specialized in detecting deepfake content will rise.
- Various political, social and economic attacks can be expected. As the technology improves, the frequency of these attacks will increase.
- A "race" between deepfake creation technology and deepfake detection technology development is taking place.
- Deepfake content increases risk to society, democracy and privacy. Extortion, fake accusations, and other type of attack will increase in frequency, not only among celebrities and public figures, but also individuals who are not public figures.

Deep learning technology will further evolve, and improve. This improvement is almost exponential. Just a few years ago, in 2017 the first deepfake videos and photos appeared on the Internet. Today, there are apps which allow individuals without technical skills and expertise to create deepfake content. These application have limitations and there are methods which can detect current deepfake content. Certainly, deepfake content is a threat with which the modern political and socio-economic domains haven't faced before. Such content is one of the negative sides of modern digital technologies.

V. Conclusion

Deep learning, more precisely DNNs have brought threats and uncertainty when it comes to trusting content which is distributed on the Internet. Deepfake content and the technology behind are the future of digital technology, and there should be safeguard mechanisms which can prevent the harmful effects of such technologies. Deploying permanent mechanisms is counter-productive as deep learning, and deepfake content is improving and evolving at a fast rate. Therefore, flexibility and adaptability is an imperative when it comes to annulling potential threats which may come from deepfake content. It can be concluded, that preventive measures are a necessity in this modern day and age when it comes to content where individuals or groups can be targeted by cyber attacks in it various forms (deepfake, stealing personal data, extortion etc.). In this paper, a concise overview was provided on the trending issue of deepfake technology.

The main limitation of the article is the lack of survey data on just how deepfake content can affect enterprises, individuals or groups. However, the paper still gives a solid basis for future research in this domain. Future research should aim at analyzing the speed of deep learning technology development and its application for creating deepfake content. In addition, the number of malicious attacks through deepfake content should be addressed and discussed. This way, this current paper will be broaden and narrowed at the same time, providing a significant outlook on how will deep learning technology evolve and affect social, political and economic activities and trends.

References

[1] M. Bakator, D. Đorđević, and D. Ćoćkalo, "Developing a model for improving business and competitiveness of domestic enterprises," Jo. of Eng. Man. and Comp. (JEMC), vol. 9, no. 2, pp. 87-96, 2019.

[2] M. Westerlund, "The Emergence of Deepfake Technology: A Review," Tech. Inn. Manag. Rev., vol. 9, no. 11, pp. 40-53, 2019. http://doi.org/10.22215/timreview/1282

[3] A. O. J. Kwok, and S. G. M. Koh, "Deepfake: a social construction of technology perspective," Curr. Iss. in Tour., pp. 1–5, 2020. doi:10.1080/13683500.2020.1738357

[4] D. Yadav, and S. Salmani, "Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network," International Conference on Intelligent Computing and Control Systems (ICCS), 2019. doi:10.1109/iccs45141.2019.9065881

[5] P. Korshunov, and S. Marcel, "Vulnerability assessment and detection of Deepfake videos," In The 12th IAPR International Conference on Biometrics (ICB) (pp. 1-6), 2019.

[6] J. Kietzmann, L. W. Lee, I. P. McCarthy, and T. C. Kietzmann, "Deepfakes: Trick or treat?," Bus. Hor., 2019. doi:10.1016/j.bushor.2019.11.006

[7] N. S. Ivanov, A. V. Arzhskov, and V. G. Ivanenko, "Combining Deep Learning and Super-Resolution Algorithms for Deep Fake Detection," 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2020. doi:10.1109/eiconrus49466.2020.9039498

[8] M. Koopman, A. M. Rodriguez, and Z. Geradts, "Detection of deepfake video manipulation," In The 20th Irish Machine Vision and Image Processing Conference (IMVIP), pp. 133-136, 2018.

[9] H. K. Hall, "Deepfake Videos: When Seeing Isn't Believing Cath," UJL & Tech, vol. 27, p. 51, 2018.

[10] L. Yuezun, Y. Xin, P. Sun, Q. Honggang, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3207-3216, 2020.

[11] H. Jahankhani, S. Kendzierskyj, N. Chelvachandran, and J. Ibarra, "Cyber Defence in the Age of AI, Smart Societies and Augmented

Humanity," Adv. Sci. and Tech.s for Sec. App., 2020. doi:10.1007/978-3-030-35746-7

[12] A. Mendelson, "Security and Privacy in the Age of Big Data and Machine Learning," Computer, vol. 52, no. 12, pp. 65–70, 2019. doi:10.1109/mc.2019.2943137

[13] Y. Chin-Yuan, H.W. Chen, S. L. Tsai, S.D. Wang, "Disrupting Image-Translation-Based DeepFake Algorithms with Adversarial Attacks," The IEEE Winter Conference on Applications of Computer Vision (WACV) Workshops, pp. 53-62, 2020.

[14] X. Li, K. Yu, S. Ji, Y. Wang, C. Wu, and H. Xue, "Fighting Against Deepfake: Patch&Pair Convolutional Neural Networks (PPCNN)," In Companion Proceedings of the Web Conference 2020 (WWW '20). Association for Computing Machinery, New York, NY, USA, pp. 88–89, 2020. doi:https://doi.org/10.1145/3366424.3382711

[15] J. Fletcher, "Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance," Thea. Jou., vol. 70, no. 4, pp. 455-471, 2018. doi:10.1353/tj.2018.0097.

[16] K. M. Manheim, and K. Lyric, "Artificial Intelligence: Risks to Privacy and Democracy," Yale Jou. of Law and Tech, no. 2018-37. Available at SSRN: https://ssrn.com/abstract=3273016

[17] S. Lyu, " Deepfake Detection: Current Challenges and Next Steps," 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2020 doi:10.1109/icmew46912.2020.9105991

[18] S. Dobbie, "Challenge of biometric security for banks," Biometric Tech. Today, vol. 2020, no. 3, pp. 5–7, 2020. doi:10.1016/s0969-4765(20)30037-0

[19] H. R. Hasan, and K. Salah, "Combating Deepfake Videos Using Blockchain and Smart Contracts," IEEE Access, pp. 1–1, 2019. doi:10.1109/access.2019.2905689

[20] S. Fernandes, S. Raj, R. Ewetz, J. S. Pannu, S. K. Jha, E. Ortiz, I. Vintila, and M. Salter, "Detecting Deepfake Videos Using Attribution-Based Confidence Metric," The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 308-309, 2020.

Data Interoperability in Higher Education Information Systems: Case of Diploma XML Generator

Ljubica Kazi, Dušanka Milanov, Marko Blažić,
Aleksandra Stojkov, Milica Mazalica, Maja Gaborov, Igor Vecštejn
University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia
ljubica.kazi@gmail.com, dusanka.milanov@gmail.com, markoblazic93@gmail.com,
sahskica.s@gmail.com, milicamazalica@gmail.com, majagaborov26@gmail.com, igor.vecstejn@gmail.com

Abstract – In modern world of communications and IT-based integration of business processes, very important aspect is related to data interoperability. Today, integration between institutions' and companies' information systems is implemented by using web services, where data transfer is based on common data formats such as XML and JSON. This paper presents interoperability at one institution level, in context of integrating software results from different vendors. Particular domain, addressed in this paper, is related to document management in higher education institution and illustrated with the example of processing data needed for diploma printing. The specific case of integrating different data sources and formats, as well as the software tool "Diploma XML generator" will be described.

I. Introduction

Data collection and document management are very delicate in the case of higher education, particularly when the diploma data are to be collected, processed and printed. Usually, all software functions related to the whole life-cycle of students are integrated within the students' office information system, or moreover, higher institution information system.

Generally speaking, it would be beneficial if all business processes could be supported by appropriate software functions in the information system of an institution or an enterprise. Usually, a business process consists of activities that are performed in collaboration of many different organization units. Using information technologies enables integration of activities, by collecting, processing and sharing data. However, real-world practice shows that not all business process activities are supported by software solutions and some business process are organized by having particular activities supported by software from different vendors. Therefore, data interoperability among software tools is needed, in aim to enable smooth operation and integration of activities within a business process.

This paper presents results of integration of different software tools in aim to support the process of data collection, transformation and processing for graduated students' diploma printing. Particular case has been shown with the example of University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia (abbreviated: @TFZR). This paper describes the particular case of the diploma data processing and software tool "Diploma XML generator", which integrates data from students' office database export (students' data) with other relevant data (such as study plan and courses data). The tool produces output in the form of XML, which has the structure suitable to be integrated with InDesign diploma template files.

II. Related work

"Interoperability has been a basic requirement for the modern information systems environment." [1] It could be analyzed with four basic levels (scope): department, enterprise, regional and global. There are several fundamental characteristics of interoperability, such as distribution, heterogeneity and autonomy. [1]

"One of the leading areas where integration and sharing is in high demand is education, particularly e-learning." [2] Major interoperability efforts in e-learning include standards and protocols for educational content dissemination. [2]

Another area of interoperability application is related to digital cultural content. "The key to the interoperability of digital cultural content is consistency", which "is achieved through the use of standards – codified rules and guidelines for the creation, description and management of digital resources". [3]

The paper [4] addresses interaction between information environments (such as libraries) and instructional/learning environments, particularly in higher education. It positions the research in the context of Global Learning Consortium and alignment of standards to bridge the gap between library information services and emerging e-learning environments.

Learning objects, instructional design and architectures, as well as digital libraries have been integrated within the scope of the paper [5] and their interaction and interoperability has been examined in the context of higher education. Learning objects are presented with metadata schemes, which are presented with XML and RDF specifications. Particularly wide-recognized standard is IEEE LOM (Learning Objects metadata) standard from year 2002.

"E-Government continues to be recognized as a key strategy for improving government services and the effectiveness of public policies and programs. A key component of e-government initiatives is the ability of multiple government and non-government organizations to share and integrate information across their traditional boundaries. E-Government interoperability represents a set of multidimensional, complementary and dynamic capabilities needed among these networks of organizations in order to achieve successful information sharing." [6] Interoperability in administration, particularly in e-Government faces challenges of different solutions to information integration and process integration. "Information integration centers around facilitating the flow of information, i.e. providing access to structured informational resources across technical and organizational border, in order to enable new services based on a virtually shared information environment". [6] Information integration brings data together, physically or logically, which makes possible for applications to make use of the relevant data in the enterprise. In process integration, areas relevant for e-Government are business process integration, cross-organizational workflow management and web services, focusing on the value for the process client (e.g. a citizen using an e-Government service) [7].

Interoperability could be considered a feature within integrated university information systems, particularly with such systems that support process workflow and document management. In the paper [8], web based information system is created in aim to enable automated workflow business process management and document management at integrated university. The system is created during restructuring organization of university by shifting from independent faculties to centralized university. Special emphasize has been given to interoperability and composition of web services. Another important application area is integration and data interoperability between higher institutions information systems. Such an example has been presented in [9], with particular implementation based on cloud computing.

III. The case of Diploma Processing and Data Integration

This paper presents particular case of document management in higher education institution related to collecting, integration and processing of data needed for diploma printing. The particular case is related to University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin.

Figure 1 presents business process model as set of interrelated activities and artifacts, as well as the organization units, i.e. working positions in the whole process of diploma preparation. The model is given in the form of UML activity diagram, where swim-lanes are related to working positions of vice-dean for education, students' officer, students' office software admin and diploma software admin. Vice-dean for education is responsible for keeping and maintaining records about study plan and subjects ("study programme"), as well as general institution data. This position is in charge for periodical initiation of data export from students' office database. Students' office employees (students' officers) are in charge for entering and maintenance of data in students' records database. They enter data about all relevant students' activities, including enrollment, exams and graduation. Especially important is the role of students' office software admin, which is able to start export from the students' records data. Finally, diploma software admin integrates general data and specific data related to each student and transforms the collected data in the form suitable to be entered into "Diploma XML generator" software tool. Then, the output XML set for each student is generated by the "Diploma XML generator" software tool.

Figure 2. presents Unified Modeling Language (UML) component model of an established system (consisting of multiple software and their outputs) for data collection, transformation, integration and finalization related to diploma printing.

IV. Diploma XML Generator

Bologna process brought many innovations in data related to study process, such as ECTS (European Credit Transfer System) points to be addressed to each subject and one-semester subjects. It also introduced new types of diploma – basic diploma (containing general study programme and student's data) and diploma supplement (having details related to exams and additional student's information, such as awards, entry procedures and points, as well as descriptions of study programme goals and professional status of graduates). Since ECTS system enables students' exchange programmes (transitions from one to another university, supported by Erasmus + mobility grants and others), one of innovations in diploma-related document management is creating diploma and diploma supplements written in two languages (local language used at dominant at regular classes and English language). All these innovations directed existing students' office information systems to upgrade and include new data and business rules in supporting software solutions.

First years for University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia to have new form of diploma/diploma supplement prepared and issued for the large number of graduates, urged the need for creation of particular software tool "Diploma XML generator". This tool was created as a C# desktop application by Ljubica Kazi in year 2014 and has been maintained and used by the author ever since. The tool integrates data from various sources (Microsoft Excel files) and transforms them into a set of XML files that have the structure precisely appropriate to be integrated with InDesign diploma templates. This tool enables creating XML outputs for Bachelor and Master diploma – basic diploma and diploma supplement written in Serbian and English language.

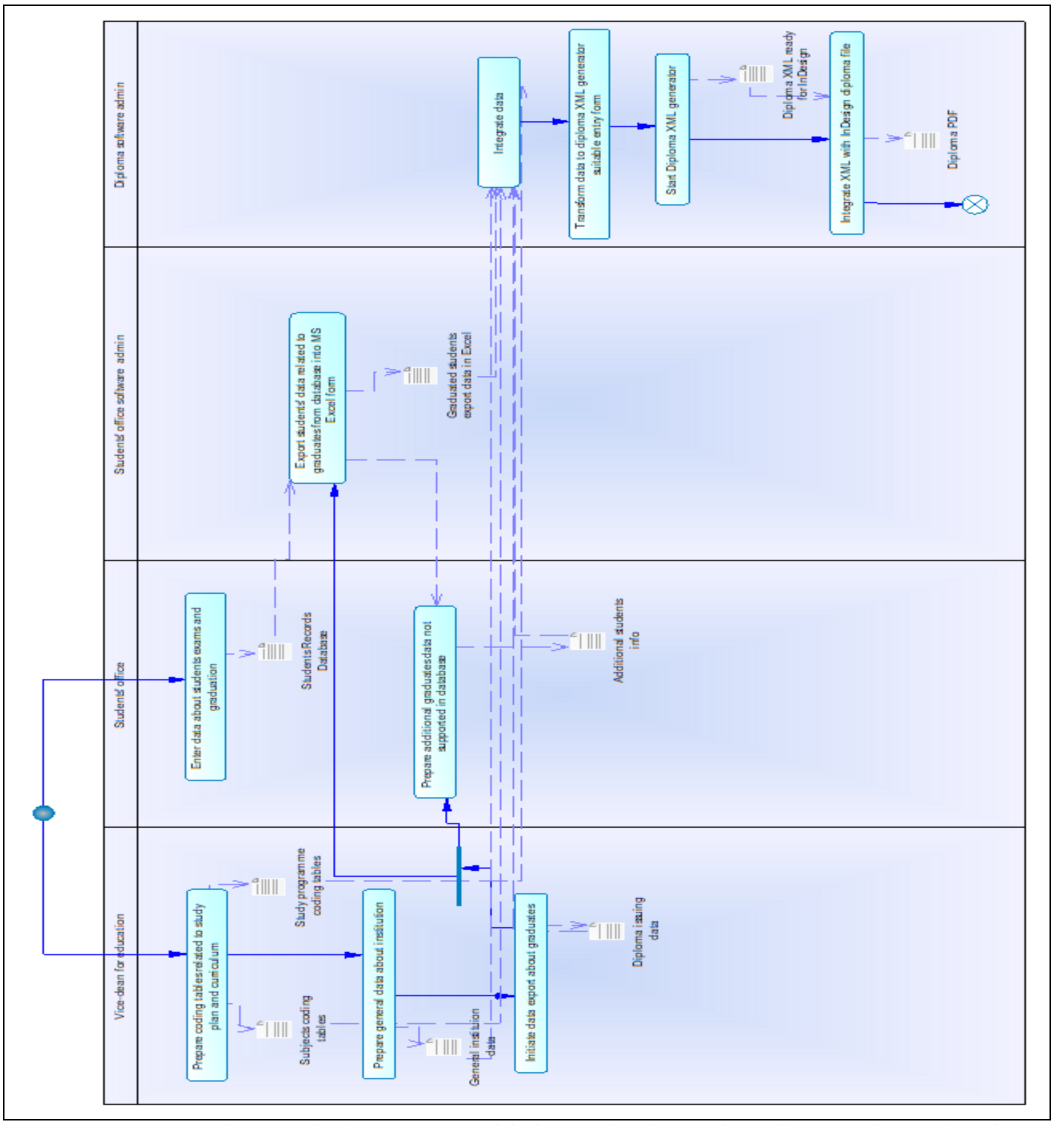


Figure 1. UML activity diagram as business process model of organizing diploma data collection and processing @TFZR

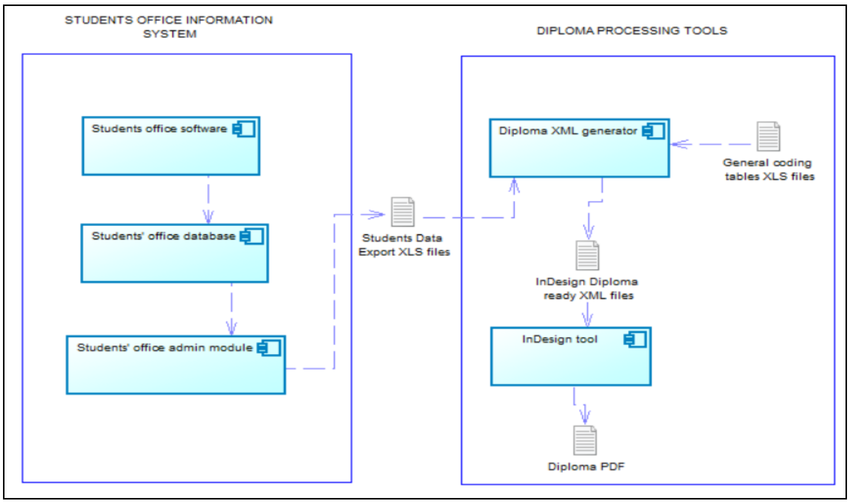


Figure 2. UML component diagram related to all used tools in diploma data processing @TFZR

Figure 3 presents user interface of the tool “Diploma XML generator”.

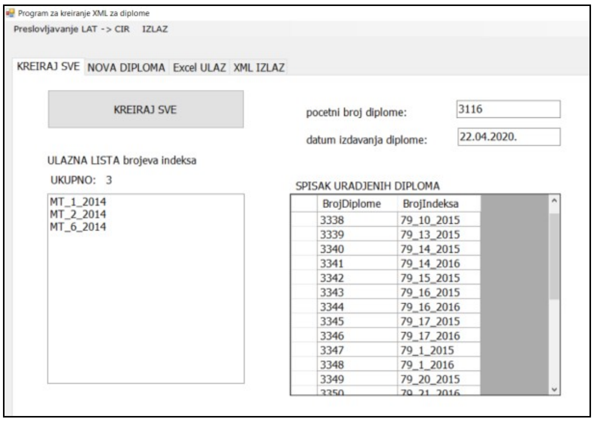


Figure 3. User interface of Diploma XML generator software tool

Example of MS Excel input for particular student is presented at Figure 4.

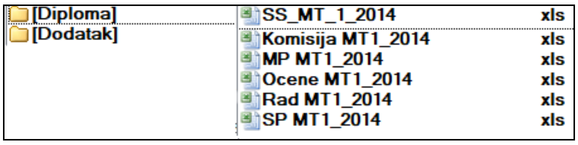


Figure 4. Part of input data to Diploma XML generator

Figure 5 presents folder with several generated diploma XML sets – for each students there is one folder with four XML files – diploma (Serbian, English) and diploma supplement (Serbian, English).

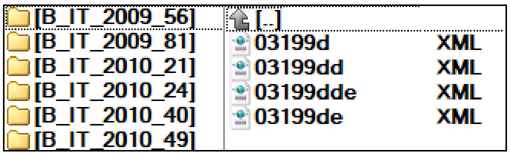


Figure 5. Output from Diploma XML generator – list of folders, where each student has one folder with four XML files

Figure 6. presents an example of generated XML file contents.

Figure 6. Example of generated XML for diploma and diploma supplement in Serbian language

V. Quality issues in diploma data integration and XML processing

Considering the significance of a diploma as a document related to future employment of graduates, the quality of data included in diploma is crucial. The printed diploma is the most important product in the process, but the process organization, people engaged and tools used in their creation influence the result.

There were several issues during the diploma creation process:

1. Changes with including new accreditation data brings new or updated study plans (lists of subjects), changes in subjects titles, number of class hours and ECTS points. Sometimes, the change in study plan, that are included in new accreditation, does not influence changes in subjects titles, but only in their teaching hours and ECTS points.
2. Students’ officers abilities to use new information technologies appropriately (insufficient knowledge and skills to use advanced options in tools, such as Microsoft Excel, Microsoft Word etc.) influence quality of initial data entry. Some of them were ready for new challenges, including advanced options in software tools, and even more – to perform inspection and correction of generated diploma XML.
3. Awareness of importance of data pre-processing phase, i.e. collecting, checking and completing “general and coding-tables data” before “Diploma XML generator” tool usage - It is of a great value to avoid multiple errors in generated XML files due to insufficient paying attention to check the core general data that are to be applied and integrated with each students’ particular data.
4. Changes of diploma forms (general design and data structure pattern) and adding new rules to diploma data structure - brings the need to change the software tool functionality and output form.
5. Existence of particular sub-cases within study plans (more precisely named modules) requires automated detection of data which direct recognition of special cases and categorization of a student into a specific module, with applying divergent automated behavior resulting in appropriate category of XML output.
6. Diversity of coding tables (UTF-8 with Cyrillic letters for Serbian version), as well as language-specific conventions, such as decimal places separator in average marks (comma or dot).
7. Optimization of processing algorithm – there was the need to speed-up processing of each students’ diploma, because of the number of input and output data. In this particular case, average number of graduates (diplomas) to be processed for each year at University of Novi Sad, Technical Faculty “Mihajlo Pupin” Zrenjanin, Serbia is 300 multiplied by 4 (four files for each student) gives total number of diplomas/diploma

supplements (in Serbian and English language) to be created and printed.

8. Typing errors from students' office employees (officers) in regular data entry lead to possible unrecognizable elements, such as course (subject) code. In such cases, software is unable to recognize particular data and to integrate them with general data and the diploma XML for that particular student will not be generated.

VI. CONCLUSION

Interoperability as an issue arises in integration of information systems of different institutions and companies, as well as within a single institution having diversity of software solutions to support business processes. Modern solutions that enable interoperability include web services, specific data formats (such as XML and JSON), as well as integration by using cloud services. Particularly intensive need for integration is addressed in e-Government and e-Learning areas. Worldwide-recognized solutions have also special applications in higher education institutions.

Aim of this paper is to address the interoperability problems in specially focused area of document management at a higher education institution, by presenting a case related to diploma data collection, transformations and processing. The particular case has presented the specially created tool "Diploma XML generator" that integrates raw data (from relational database exported into the form of Microsoft Excel files) with general data (related to study plan and courses) needed to complete data set for diploma generating. This tool, after integration, performs transformation of data and finally, creates output in XML form precisely adjusted to fit into InDesign diploma document template. This example shows how software from students' office information system could be integrated with software that creates output for InDesign. This way, three software applications were integrated and interoperability was established with data that were used for integration (Microsoft Excel files and XML files).

The "Diploma XML Generator tool" is not a final solution in this particular case of University of Novi Sad, Technical Faculty "Mihajlo Pupin" Zrenjanin, Serbia. The tool was created in 2014 and maintained until the latest changes in year 2020. During software maintenance, the software was made more flexible to adjust to new requirements and special cases. It has been also made more general (scalable to any study plan, module and courses) and adaptable to future expected changes with automated detection and categorization of data (such as a module detection). Still, the developed tool could be improved in structural (code organization) and non-functional (performances, user interface) aspects.

REFERENCES

[1] A. P. Sheth, "Changing focus on interoperability in information systems: From system, syntax, structure to semantics", book chapter in: M.F. Goodchild, M.J. Egenhoger, R. Fegeas and C.A. Kottman (eds): Interoperating Geographic Information Systems, Kluwer, 1998.

[2] M. Hatala, G. Richards, T. Eap and J. Willims, "The Interoperability of Learning Object Repositories and Services: Standards, Implementations and Lessons Learned", International conference WWW 2004, May 17-22, 2004, New York, ACM, USA

[3] T. Gill and P. Miller, "Re-inventing the wheel? Standards, interoperability and digital cultural content", D-Lib Magazine, Vo. 8, No 1, ISSN 1082-9873, January 2002.

[4] N. McLean and C. Lynch, "Interoperability between Library Information Services and Learning Environments – Bridging the Gaps", a joint white paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information, Pp.1-15, 2004.

[5] M.A. Marzal, J. C. Prado, M.J.C.Ruiz and A.C Cervero, "Learning objects, instructional architectures and digital libraries: New perspectives on interoperability", 5th International conference on Information Communication Technologies in Education ICICTE 2004, Samos, Greece.

[6] T.A. Pardo, T. Nam and G. B. Burke, "E-Government interoperability", Social Science Computer Review, Vol 30, No. 1, February, 2012.

[7] R. Klischewski, "Information Integration or Process Integration? How to Achieve Interoperability in Administration", International Conference eGovernment, 2004.

[8] B. Fetaji, M. Fetaji and M. Ebibi, "Sofware Engineering Interoperable Environment for University Process Workflow and Document Management", International Journal of Computer and Information Engineering, Vol. 4, No 4, 2010.

[9] L.M Ribeiro, R. H. Pereira, O. Pacheco, M. Bernardes and R. T. Marins, "Interoperability between information systems of Portugese Higher Education Institutions", International Conference European University Information Systems EUNIS, Athens, Greece, 2016.

Image Classification Using Convolutional Neural Networks

S. Mihajlović [*], D. Ivetić[**], I. Berković[*]
* Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
** Faculty of Technical Sciences, Novi Sad, Republic of Serbia
mihajlovicsinisa25@gmail.com, ivetic@uns.ac.rs, ivana.berkovic62@gmail.com

Abstract - In this paper, the classification of images on the created dataset is presented. Examples of the role of convolutional neural networks are given, as well as the results achieved. Tensorflow, Keras API, Python and convolutional neural networks were used for the practical part of image classification. To classify images, it is necessary to create dataset, to load images and labels, to perform a pre-process for each image, to define the layers of neural network architecture. After defining the layers of the neural network architecture, it is necessary to train the network, and to obtain the results, it is necessary to test the network. The tested network showed an accuracy of 78.57%.

I. Introduction

Image classification is one of the more important problems in computer vision. Traditional approaches to machine learning were used to classify images. Nowadays, a convolutional neural network is used, which has shown excellent in computer vision problems, and most of all in image classification. The goal of image classifications is to perform a classification based on visual content based on input images. Most researchers relied on hand-crafted features, HOG or SIFT, to describe the image in a discriminatory way. There is an SVM classifier that is applied to the selected characteristics to make a final decision. When there are too many images it is difficult to find their characteristics. That is why deep networks are used. When the input dimension is too large to use, the deep neural networks takes a long time to train. Convolutional neural networks solved this problem by dividing weights by the convolution method, and thus improved the performance of classification for different datasets [1].

CNN model design and training involves a large amount of trial and error. Existing visualization work on CNN mainly focuses on what neurons have learned, connections in the network and network structure, as well as training information. Existing CNN visualization techniques lack the ability to investigate and compare the difference in parameters/weights of two model images. Model performance usually improves over the time of the training process. One can know the training status from accuracy and loss information, but one cannot know what happens to the network parameters and how they affect the performance of the CNN model. Better visualization of the parameter update trend speeds up the training process and improves model performance. CNN model training involves learning parameters for a model from a training dataset. The performance of the model is largely related to the learned parameters. There can be millions of parameters in the model, which makes it difficult to find important parameters. Different training processes can sometimes give different results due to the differences introduced by random initialization and interpretation of different training processes is not feasible [2].

CNN is named after the convolution, a widely used operator used in image and signal processing. In these areas, it is about convolutional filters that are used to sharpen and blur images, as well as to detect edges. From the aspect of CNNs, the application of edge detection algorithms, which detect objects in the image, is important. Where classical neural networks show their disadvantages, CNNs show their advantages. The idea of CNNs is to lay more layers to reveal the essential properties of the input data. In accordance with this idea, convolutional filters are applied to the image to extract useful characteristics and create feature maps. By applying filters to the input image, in addition to detecting significant characteristics, the resolution is also reduced. The disadvantages of CNNs is the need for significant hardware resources, as well as the fact that previous models have been trained with a large number of images, but recent results suggest that high performance can be achieved with a relatively small number of samples [3].

II. Related Work

The convolutional neural network (CNN) is state-of-the-art for performing image classifications. CNN or ConvNet is a special type of multilayer neural network inspired by the mechanism of the optical system of living beings. Hubel and Wiesel found that animal cortex cells detect light in a small receptive field. LeCun et al. introduced the practical model of CNN and developed LeNet-5. Training by backpropagation algorithm helped LeNet-5 recognizing visual patterns from raw pixels directly without using any separate feature engineering mechanism. The performance of CNNs such as high-resolution image classification is limited by the lack of large training data, the lack of a better regularization method, and inadequate computing power. Today, there are larger datasets with millions of tagged high-resolution data in thousands of categories such as ImageNet, LabelMe, etc. With the advent of powerful GPU machine and better regularization method, CNN delivers outstanding performance on image classification tasks. A large convolutional neural network called AlexNet designed by Krizhevsky et al. in 2012 demonstrated excellent performance on the ImageNet large-scale visual recognition challenge (ILSVRC). The success of AlexNet has become an inspiration to various CNN models such as ZFNet, VGGNet, GoogleNet, ResNet, DenseNet, CapsNet, SENet etc. in the coming years [4].

Classification is a process that does not fully utilize the potential discriminatory power of what is learned from CNN. This leads to the need for fusion of characteristics from multiple layers. A method of combining characteristics from several layers in given CNN models is proposed. Learned CNN models with training images are again used to extract characteristics from multiple layers. This proposed fusion method is estimated according to the datasets of the reference image classification, CIFAR-10, NORB and SVHN. In these cases, the proposed method was shown to improve the existing performance of existing models by 0.38%, 3.22% and 0.13%, respectively [1].

In the context of machine vision, image recognition is the ability of software to identify objects, places, people, writing and actions on images. Computers use machine vision technologies with a camera and artificial intelligence software to accomplish the task of image recognition. To perform many machine visual tasks, image recognition is used, such as image content tagging, image content search to guide autonomous robots, self-driving cars, and random avoidance systems. The human brain easily recognizes objects, while computers have difficulty with the task. Image recognition software requires deep machine learning. Performance is based on the complexity of CNN because a specific task requires a huge amount of computing power. Compared to other region proposal classification networks (fast RCNN) which perform detection on various region proposals and thus end up performing prediction multiple times for various regions in an image, Yolo architecture is more like FCNN (fully convolutional neural network) and passes the image (n*n) once through the FCNN and output is (m*m) prediction. Thus, the architecture is splitting the input image in m*m grid and for each grid generation 2 bounding boxes and class probabilities for those bounding boxes. Note that bounding box is more likely to be larger than the grid itself [5].

An image recognition algorithm based on the ensemble learning algorithm and neural network convolution structure (ELA-CNN) is used to solve the problem because the classical neural network classifier (CNN) may be more prone to errors or unreliable prediction. In order to improve the ensemble learning effect, improve the transfer of characteristics, extract deeper characteristics and the characteristics of multiple scales, the network structure uses different model structures of the main algorithms. Bagging training method is used in the training process, that is, different learners use different data sets to ensure the learning differences. The prediction result of all classifiers is used to obtain the final recognition of the image target according to the decision algorithm. The algorithm was simulated using the open data set CIFAR-10. Experimental results show that the proposed algorithm has high recognition accuracy. The recognition rate of the test set is 98.89%, while the recognition result is safer and more reliable [6].

An algorithm for the detection and identification of white blood cells, the idea is to build convolutional neural network based on the available white blood cell image database, which classifies white blood cells into one of five types. The construction of the model was done offline, i.e. it was made before the analysis of images of new blood samples. Blood sample images are segmented to isolate all white blood cells. Segmented white blood cell images are then brought to a pre-trained classifier input and predict into one of five classes (eosinophils, segmented neutrophils, monocytes, lymphocytes, or unknown). At the very end, the cells are marked on the original image along with the corresponding class. When using supervised learning algorithms, it is very important to have well-marked data for model training, i.e. for each measured sample of input quantities there is a corresponding value of the output quantity. After all images were tagged and classified, values were obtained. After all the cells are marked, using the Python script, the images of the detected cells of the desired size are cut with the cell type contained in the image name. The labeled images thus obtained are used as a learning set in the process of constructing a classifier based on a convolutional neural network [7].

Information about a person's health can be found based on nails. There are many diseases that can be seen through the condition and shape of the nails. The research was conducted in order to detect abnormalities on the nails on the axis of digital images. The convolutional neural network method was used to detect abnormalities. The Tensorflow Inception-V3 model of architecture with the method of learning transfer was used for the research, and the obtained results of experiments with an accuracy of 95.24% [8].

III. USED TECHNOLOGIES

Colaboratory, or "Colab" for short, allows you to write and execute arbitrary Python code through a browser, and is especially suitable for machine learning, data analysis, and education. The code is executed in a private virtual machine for each account. Colab is a hosted Jupyter notebook service that does not require settings, and at the same time provides free access to computer resources including the GPU.

Tensorflow is Python's open source numerical computing library that makes machine learning faster and easier. Tensorflow can train and power deep neural networks to classify handwritten numbers, recognize images, and more. Tensorflow applications are Python applications themselves. Actual mathematical operations are not performed in Python but as high-performance C++ binaries. Tensorflow applications can run on a local machine, cluster in the cloud, iOS and Android devices, CPU, GPU and TPU.

Keras is a deep learning API written in Python, launched at the top of the Tensorflow machine learning platform. It was developed with a focus on enabling rapid experimentation. Being able to move as quickly as possible from idea to result is key to good research. Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is extremely readable, and its language construction and object-oriented approach help developers write clear and logical code. Convolutional neural networks are one of the most popular algorithms for deep learning, a type of machine learning in which a model learns to perform classification tasks directly from images, videos, text, or sound.

IV. EXPERIMENT

The aim of the experiment is to classify the images on the created dataset. Tensorflow, Keras API, Python and convolutional neural networks were used for the practical part of image classification. The images show Sinisa and Zeljka. Images are of different qualities and resolutions.

A. Find images and label them

It is necessary to create a dataset containing images of Sinisa and Zeljka for training CNNs. The created dataset contains 128 images, of which 60 are images of Sinisa and 68 images of Zeljka. Each image must be renamed to mark the images. In the images where Sinisa is, the name of the image must be changed to Sinisa 1, Sinisa 2,..., as well as where Zeljka is, the name of the image must be changed to Zeljka 1, Zeljka 2,... Figure 1 shows a part of the created dataset containing images of Sinisa and Željka for CNN training.

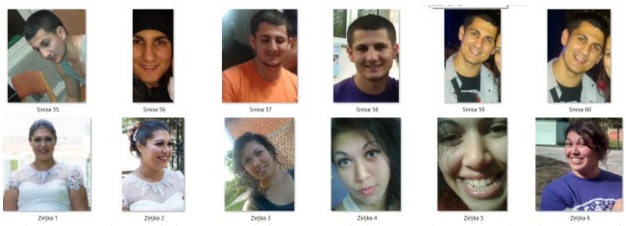


Figure 1. View of the part of the created dataset that contains images of Sinisa and Zeljka for training CNNs

B. Load images and labels into Python

Google Colab opens, and a new notebook is created called ClassificationSinisaOrZeljka.ipynb. After that it is necessary to install Tensorflow. Listing 1 shows the code for successfully installing Tensorflow.

LISTING I.

```
!pip install tensorflow
```

Google Colab is connected to the Google Drive, which stores the created notebook in the Colab Notebooks folder. Images used for a dataset must be packaged. The packaged images are called training.zip and are uploaded to Google Drive in the Colab Notebooks folder. It is necessary to allow access to Google Drive to access a dataset called training.zip. To allow access, an authorization code must be entered, by clicking on the received link, the authorization code is obtained, which should be copied and confirmed. A training file called training.zip has been defined. The training file must be unpacked on the z: partition located online, to use the images as a defined data set. Listing 2 shows the code for allowing access to the Google Drive, defining the training file, and unpacking the training file.

LISTING II.

```
from google.colab import drive
drive.mount("/content/drive")

train_file = "/content/drive/My Drive
/Colab Notebooks/trening.zip"

import zipfile

with zipfile.ZipFile(train_file, 'r') as z:
    z.extractall()
```

When unpacking images that are in training.zip, it is necessary to define the path to the folder that contains the images that represent the dataset. The OpenCV library is used to load images as a pixel array and is stored in a list. Since the corresponding label is in the file name, a numeric label can be extract for each image, 1 is assigned if the label is Sinisa, otherwise it is 0 for Zeljka. Listing 3 shows the code to load images and assign a value of 1 or 0.

LISTING III.

```
import os
import cv2

def load_image(file_path):
    return cv2.imread(file_path)

def extract_label(file_name):
    return 1 if "Sinisa" in file_name else 0

train_path = "./trening/"
image_files = os.listdir(train_path)
train_images = [load_image(train_path + file) for
file in image_files]
train_labels = [extract_label(file) for file in
image_files]
```

C. Preprocess the image data

It is necessary to pre-process the data because raw images are used, which are different dimensions and sizes, and the neural network requires that each sample has the same dimensions and sizes. It is necessary to crop the image, and to change the width and height of the image to a fixed value, and it is changed to 200 pixels. Images can be in different colors, and this adds more noise and useless complexity to the model, so it is necessary to convert each image to a grayscale that removes three RGB layers and gives only one pixel brightness value and finally returns the normalized image. Listing 4 shows the code for cropping the image, changing the width and height of the image to 200 pixels, and converting the image to grayscale.

LISTING IV.

```
def preprocess_image(img, side=200):
    min_side = min(img.shape[0], img.shape[1])
    img = img[:min_side, :min_side]
    img = cv2.resize(img, (side,side))
    img = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
    return img / 255.0
```

It is necessary to display some images before and after to verify that the data processing has been done successfully. Listing 5 shows the code for displaying images before and after the pre-processing.

LISTING V.

```
def preprocess_image(img, side=200):
    min_side = min(img.shape[0], img.shape[1])
    img = img[:min_side, :min_side]
    img = cv2.resize(img, (side,side))
    img = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
    return img / 255.0 import matplotlib.pyplot as
plt

preview_index = 13
plt.subplot(1,2,1)
plt.imshow(train_images[preview_index])
plt.subplot(1,2,2)
plt.imshow(preprocess_image
(train_images[preview_index]), cmap="gray")
```

Figure 2 shows the image before and after pre-processing the data.

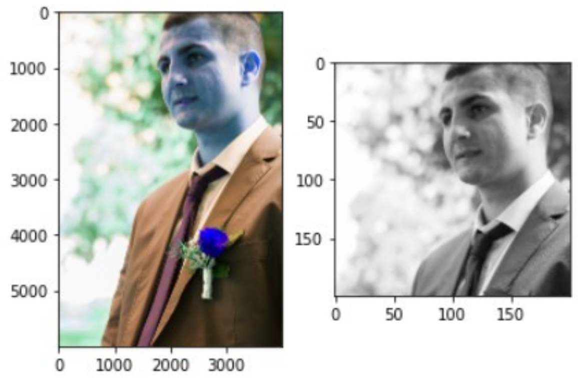


Figure 2. Image display before and after pre-processing data

If the dataset is very large and takes up a lot of RAM, care should be taken when creating new copies of the data. That is why pre-processing was done for each image. The images are converted to grayscale and therefore it is necessary to expand the last dimension into one channel to use the convolutional layers in the neural network. Listing 6 shows the pre-processing code for each image and expanding the dimension into one channel.

LISTING VI.

```
for i in range(len(train_images)):
    train_images[i] =
preprocess_image(train_images[i])

import numpy as np

train_images = np.expand_dims(train_images, axis=-
1)
train_labels = np.array(train_labels)
print(train_images.shape, train_labels.shape)
```

D. *Define the neural network architecture*

It is necessary to define the layers that will make up the architecture of the neural network. Tensorflow is imported first and it is checked which version is installed. Tensorflow 2.3.0. is installed version. Using the high-level Keras API, a sequential model is created that takes in a sequence of layers to transform the input data into classification predictions. The list starts with a convolutional layer, with 16 filters, kernel size of three by three, padding equals same and the ReLU activation function. The shape of the input sample must be determined for any first layer of the Keras model. A max-pooling layer is added, to reduce the sample output values in height and width by half with pool size and stride two by two. Then it is necessary flattened the output to pass it to a couple of dense layers and then a final dense layer that outputs a probability for each of the two output classes. It is necessary to determine the optimizer and the type of loss function that is sparse categorical cross-entropy, if the labels are only classes of numbers, and the accuracy is shown as the metrics. Listing 7 shows the code for defining the layers of the neural network architecture.

LISTING VII.

```
import tensorflow as tf
print("Tensorflow:", tf.__version__)

layers = [
    tf.keras.layers.Conv2D(filters=16,
kernel_size=(3,3), padding="same",
activation=tf.nn.relu,
input_shape=train_images.shape[1:]),
    tf.keras.layers.MaxPool2D(),
    tf.keras.layers.Conv2D(filters=32,
kernel_size=(3,3), padding="same",
activation=tf.nn.relu),
    tf.keras.layers.MaxPool2D(pool_size=(2,2),
strides=(2,2)),
    tf.keras.layers.Conv2D(filters=64,
kernel_size=(3,3), padding="same",
activation=tf.nn.relu),
    tf.keras.layers.MaxPool2D(pool_size=(2,2),
strides=(2,2)),
    tf.keras.layers.Conv2D(filters=128,
kernel_size=(3,3), padding="same",
activation=tf.nn.relu),
    tf.keras.layers.MaxPool2D(pool_size=(2,2),
strides=(2,2)),
    tf.keras.layers.Conv2D(filters=256,
kernel_size=(3,3), padding="same",
activation=tf.nn.relu),
    tf.keras.layers.MaxPool2D(pool_size=(2,2),
strides=(2,2)),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(units=512,
activation=tf.nn.relu),
    tf.keras.layers.Dense(units=256,
activation=tf.nn.relu),
    tf.keras.layers.Dense(units=2,
activation=tf.nn.softmax)
]

model = tf.keras.Sequential(layers)
model.compile(optimizer=tf.optimizers.Adam(),
loss=tf.losses.SparseCategoricalCrossentropy(),
metrics=[tf.metrics.SparseCategoricalAccuracy()])
```

After defining the layers of the neural network architecture, the architecture of the sequential model and the number of training parameters for each layer, i.e. the sum of training parameters, can be presented. Listing 8 shows the code to display the architecture of the sequential model. Figure 3 shows the architecture of the sequential model and the number of training parameters for each layer.

LISTING VIII.

```
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 200, 200, 16)      160
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 100, 100, 16)      0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 100, 100, 32)      4640
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 50, 50, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 50, 50, 64)        18496
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 25, 25, 64)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 25, 25, 128)       73856
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 12, 12, 128)       0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 12, 12, 256)       295168
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 6, 6, 256)         0
_________________________________________________________________
flatten (Flatten)            (None, 9216)              0
_________________________________________________________________
dense (Dense)                (None, 512)               4719104
_________________________________________________________________
dense_1 (Dense)              (None, 256)               131328
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 514
=================================================================
Total params: 5,243,266
Trainable params: 5,243,266
Non-trainable params: 0
```

Figure 3. Sequential model architecture and number of training parameters for each layer

E. *Train the network*

A fit function model is used with images for training and labels, by number of epochs and batch size. The batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters. The number of epochs is a hyperparameter that defines the number times that the learning algorithm will work through the entire training dataset. It is necessary to determine the number of epochs and in this case, it is 50 and the batch size is 3. The number of images contained in the dataset is divided by the batch size value. In this case it is 128/3, and the result is 43. The learning algorithm will go through 43 images from the dataset 50 times. After training, the weights of the model are saved. Listing 9. shows the training code.

LISTING IX.

```
model.fit(train_images, train_labels, epochs=50,
batch_size=3)
model.save_weights("model.tf")
```

V. RESULT AND DISCUSSION

The model is trained and shows 100% accuracy on the training data, but the most important thing is how the unseen images of Sinisa and Zeljka will be classified. To get the results it is necessary to test the model. The model is tested with images numbered 1, 2, 3, … which show Sinisa and Zeljka. Test images are uploaded through the Google Colab library from where the images are located. Number of uploaded images is 14. Listing 10. shows the code for uploading images.

LISTING X.

```
from google.colab import files
uploads = files.upload()
```

To get the names of the images, only the keys can be taken, and then a list of images to test is obtained by calling the image loading function. Each loaded image must be pre-process, and a new model with the same layers is created, and the weights that were previously saved are loaded. The probability of prediction can then be obtained by calling the model predict function in the image list. At the end, all images and their predicted class label are shown, to see how successfully the trained model has classified the new images of Sinisa and Zeljka. Listing 11. shows the code for testing and displaying classified images from a predicted class label.

LISTING XI.

```
eval_images = [preprocess_image(load_image(file))
for file in uploads.keys()]
eval_model = tf.keras.Sequential(layers)
eval_model.load_weights("model.tf")
eval_predictions =
eval_model.predict(np.expand_dims(eval_images,
axis=-1))

cols = 4
rows = np.ceil(len(eval_images)/cols)
fig = plt.gcf()
fig.set_size_inches(cols * 4, rows * 4)
for i in range(len(eval_images)):
    plt.subplot(rows, cols, i+1)
    plt.imshow(eval_images[i], cmap="gray")
    plt.title("Sinisa" if
np.argmax(eval_predictions[i])==1 else "Zeljka")
    plt.axis('off')
```

In this case, the trained model successfully predicted 11 out of 14 images, and erred 3, which represents an accuracy of 78.57%. Figure 4 shows the results of classified images from a predicted class label after testing.

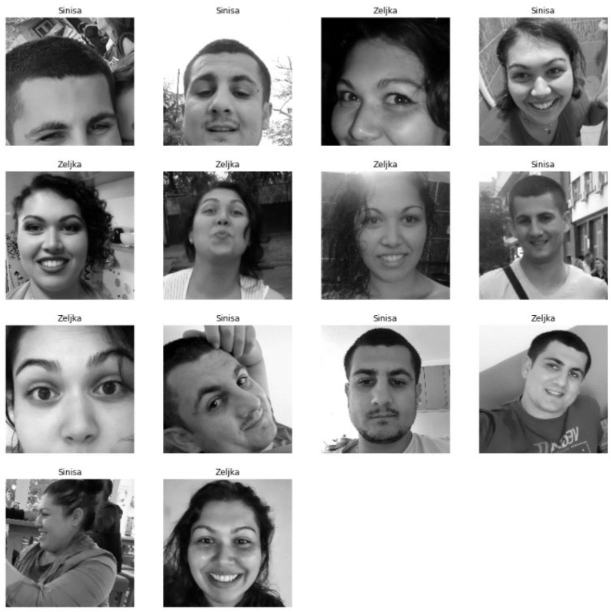


Figure 4. Result of classified images from a predicted class label after testing

VI. Conclusion

The classification of images was performed on the created data set. The dataset consists of 128 images of Sinisa and Zeljka. An appropriate label is added to each image, i.e. if Sinisa is in the image, the value 1 is returned, and if it is Zeljka, then the value is 0. After that, a pre-process of the image was done, which meant that each image had to be cropped, its height and width changed to 200 pixels, and the image converted to grayscale. The image before and after the pre-process data is then displayed. The layers of neural network architecture, optimizer and type of loss function that are sparse categorical cross-entropy are defined, if the labels are only classes of numbers, and the accuracy is shown as the metrics.

The created model consists of 5.243.266 parameters that make up the training parameters. After defining the architecture of the neural network, it is necessary to train, but before that, the number of epochs and batch sizes are determined. It is difficult to determine the number of epochs and batch size because it affects the duration of training, as well as the accuracy of the trained data. In this case, the learning algorithm went through 43 images from the data set 50 times. The model is trained and shows 100% accuracy on the training data, but the most important thing is how the unseen images of Sinisa and Zeljka will be classified. The model is tested with 14 images, and each image must be pre-process. In this case, the trained model successfully predicted 11 out of 14 images, and erred 3, which represents an accuracy of 78.57%.

Some further improvements would be to expand the dataset with more images, because in this way the network has more information and would give better results. In addition, the architecture of the neural network can be changed and improved, as well as reducing the number of epochs and increasing the batch size.

The number of images that would make up the test data can be increased. The most important thing that could be improved is that the trained model has higher accuracy in percentages, and that it has as few errors as possible.

Acknowledgment

Ministry of Education, Science and Technological Development, Republic of Serbia financially supported this research, under the project number TR32044: “The development of software tools for business process analysis and improvement”, 2011-2020.

References

[1] J. Yim, J. Ju, H. Jung and J. Kim, Image Classification Using Convolutional Neural Networks With Multi-stage Feature, Robot Intelligence Technology and Applications 3, Advances in Intelligent Systems and Computing Vol. 345, 2015.

[2] H. Zeng, H. Haleem, X. Plantaz, N. Cao and H. Qu, CNNComparator: Comparative Analytics of Convolutional Neural Networks, arXiv, 15. October 2017.

[3] M. Dabović, I. Tartalja, Duboke konvolucijske neuronske mreže - koncepti i aktuelna istraživanja, Konferencije za elektroniku, telekomunikacije, računarstvo, automatiku i nuklearnu tehniku, ETRAN 2017, Kladovo, 05. do 08. juna 2017.

[4] F. Sultana, A. Sufian, P. Dutta, Advancements in Image Classification using Convolutional Neural Network, arXiv, 8. May 2019.

[5] N. S. Lele, Image Classification Using Convolutional Neural Network, International Journal of Scientific Research Computer Science and Engineering, Vol.6, Issue.3, pp.22-26, June 2018.

[6] W. Mo, X. Luo, Y. Zhong and W. Jiang, Image recognition using convolutional neural network combined with ensemble learning algorithm, Journal of Physics: Conference Series, Vol. 1237, Issue 2, 2019.

[7] F. Novoselnik, Klasifikacija slika metodama dubokog učenja, Univerzitet Josip Juraj Štrosmajer u Osijeku, Fakultet elektrotehnike, računarstva i informacionih tehnologija, Osijek, 2017.

[8] M. Yani, B. Irawan, and C. Setiningsih, Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry’s Nail, Journal of Physics: Conference Series, Vol. 1201, 2019.

Using Cassandra in the Internet of Things

Djerdj Srdanov*, Dalibor Dobrilovic** and Goran Gecin**
* Lead Database Engineer, Knoxville, Tennessee, USA
** University of Novi Sad / Technical Faculty “Mihajlo Pupin”, Zrenjanin, Serbia
sdjerdj@yahoo.com; goran.gecin@tfzr.rs; dalibor.dobrilovic@tfzr.rs;

Abstract – Considering the complexity of IoT systems and the large quantities of data gathered, transferred and stored in these systems, choosing the proper data store for those environments becomes important and challenging task. Recognizing the importance of the data store in IoT layered architecture is a critical aspect of the design. This paper gives the overview of the most popular data store systems these days, with the focus on NoSQL databases. Also, this paper outlines the principles of the CAP theorem and applying the same principles to modern IoT systems. Finally, as a contribution, this paper makes a proposal how to integrate a NoSQL database, particularly Cassandra, in a typical IoT system. This proposal is based on the research and experience of the authors, giving the synthesis of scientific research and professional practical experience.

I. Introduction

Nowadays, we are witnesses of rapid wireless technology development, the appearance of variety of new communication standards, and significant growth in numbers of connected devices to the Internet. This growth is influenced by the low cost of the hardware and the minimization of the components. It allows the implementation of the connectivity modules in many different devices. The idea of connecting various devices to the Internet allows easier monitoring and feedback, and allows automation of tasks. This feature also allows interconnection of larger number of different devices in order to create complex systems. This idea led slowly to the creation of Internet of Things.

Internet of Things (IoT) is a paradigm that aims to enhance the “quality of life” of its users with the utilization of automatic data collection and data processing to perform actions in real world through remote sensing and monitoring [1].

This paper has the structure as follows: Section I – Introduces the topic and defines some key points; Section II – Gives short summary of related work; Section III – Explains the architecture of the IoT systems; Section IV Discusses the topic of data stores; Section V – Deals with SQL/NoSQL databases and IoT systems; Section VI - Focuses on CAP theorem and its relation to the NoSQL databases, and Section VII – Explains the usage of Cassandra NoSQL database in IoT systems. Finally, the Conclusion section provides short summary of the paper.

II. Related Work

In this section a brief overview of several research papers is given with the focus on implementation of database systems in IoT.

In [2] the big data storage technologies are explained. They are designed from ground-up to deal with large data volume and to process it in real time. These systems deal with increase in volume by introducing new nodes in the system for extra storage space and processing power. There are different technologies which aim to satisfy different goals. Some of the differentiating characteristics are: data model for storing, existence of schema, APIs for data access, architecture models, and CAP theorem compliance. One of the recent advancements in big data storage technologies is their ability to process data in-memory. These systems use RAM and flash storage to further increase responsiveness of the system.

In [3] among the number of different architectures, the Fog-based IoT architecture is considered. This architecture consists of 7 layers and one of the layers is the *storage layer*. The role of this layer is to store the data in different format as required and distribute it as needed with a suitable protocol.

In [4] the detailed analyses of RDBMs and NoSQL data storage engines in big data systems are presented. The comparison of different aspects NoSQL data bases is also given. According to the authors, the amount of data and number of concurrent users that need to be handled in big data and IoT systems has challenged relational DBMS as prime technology for storing data.

Main problems identified were enforcement of schema against data and strict adherence to ACID properties of a transaction. Solution was found in NoSQL databases that remove requirement for schemas and simplify transaction processing. They allow storing of unstructured and complex datasets.

Main differentiating factors among NoSQL databases are their data models or the way in which they organize stored data (four main types are: key-value stores, column stores, document stores and graph stores). Also, according to the same authors, there are other key aspects that need to be considered when choosing a NoSQL database. These aspects are how they operate their partitions, how they handle data replication and how they achieve consistency.

III. IoT and its Layered Architecture

In order to talk about the role of a database system in IoT it is necessary to properly analyze IoT architecture and the position of database systems in it. Although there are many different approaches, architecture of an IoT system in enough details can be presented in five layers: perception, transport, middleware, application and business layer [3, 5]. The five layered architecture is given in Fig. 1.

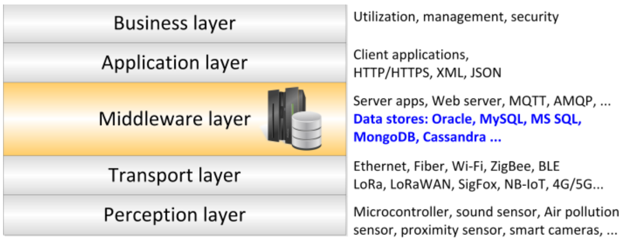


Figure 1 Five-layered IoT Architecture and DB position

On the perception layer different sensors gather data from the physical environment, target objects and their surroundings relevant to the system. The collected data packages are continuously transferred through the network layer to upper layers. Systems can monitor multiple parameters per single object and usually this data includes: temperature, humidity, vibration, motion, noise, air pollution, location and so on.

The transport layer (or the network layer), has to provide means of secure transmission of data between perception and middleware layers. It can be based on wired, but is much more commonly based on wireless standards such as Wi-Fi, Bluetooth, Bluetooth Low Energy, ZigBee, LoRa, etc.

The middleware layer handles the storage of all collected data points, processing it and then automatically taking actions based on the results. It also serves as a bridge between devices and applications. It provides access to the data for higher layers utilizing standard interfaces and services. This lets developers to focus on main functionality without having to worry about how the infrastructure was implemented [6]. As pointed out previously in this paper data store engines are located in this layer as well.

The application layer implements and manages application specific services relevant to the user. These applications are smart health, smart home, smart city, etc.

The business layer is responsible for managing the whole IoT system and all of its applications. It analyzes historical data and makes suggestions for future strategies [5, 7].

IV. DB Engines

The current market trends in the field of database systems should also be considered as part of the scope of this research. As a reference for this consideration the DB-Engines Ranking site has been used. DB-Engines Ranking site currently lists more than 340 database management systems ranked by their current popularity. The top 10 database rankings in two periods - October 2019 and October 2020 are given in the Table I [8].

The rank of databases or systems popularity is calculated using the following parameters: number of mentions of the system on websites (Google and Bing), general interest in the system (frequency of searches in Google Trends), frequency of technical discussions about the system (number of related questions and the number of interested users on the well-known IT-related Q&A sites), number of job offers (job search engines Indeed and Simply Hired), number of profiles in professional networks (LinkedIn), and relevance in social networks (Twitter tweets).

TABLE I. TOP 10 DATABASE SYSTEMS BY DB RANKING

Rank		**DBMS**	**Database Model**	**Score**		
Oct 2020	**Oct 2019**			**Nov 2020**	**Oct 2020**	**Nov 2019**
1.	1.	Oracle	Relational, Multi-model	1345.00	-23.77	+8.93
2.	2.	MySQL	Relational, Multi-model	1241.64	-14.74	-24.64
3.	3.	Microsoft SQL Server	Relational, Multi-model	1037.64	-5.48	-44.27
4.	4.	PostgreSQL	Relational, Multi-model	555.06	+12.66	+63.99
5.	5.	**MongoDB**	Document, Multi-model	453.83	+5.81	+40.64
6.	6.	IBM Db2	Relational, Multi-model	161.62	-0.28	-10.98
7.	7.	**Elasticsearch**	Search engine, Multi-model	151.55	-2.29	+3.15
8.	8.	**Redis**	Key-value, Multi-model	155.42	+2.14	+10.18
9.	11.	SQLite	Relational	123.31	-2.11	+2.29
10.	10.	**Cassandra**	Wide column	118.75	-0.35	-4.47
11.	9.	Microsoft Access	Relational	117.23	-1.02	-12.84

Table I shows the top 10 rated database systems in the October of 2019 and 2020. Besides well-know SQL databases (relational, multi-modal) several other types of databases are listed. These other, non-relational types of databases will be briefly explained in the following text. [8, 9, 10, 11]

Most of these other, non-relational data stores can be classified as:

- Key-value stores,
- Column stores,
- Document stores,
- Graph stores, and
- Search engines.

A. Key-value Stores

Key-value stores are probably the simplest form of database management systems. Data items are keys that have a set of attributes. Relevant data (key and value) is always stored together. Key-value stores can only store pairs of keys and values, as well as retrieve values when a key is known. Because of the simplicity, these systems are normally not adequate for complex applications. However this simplicity makes such systems attractive in certain circumstances. Resource-efficient key-value stores are often used in embedded systems or as high performance in-process databases. Popular key-value stores are Amazon's Dynamo DB, Riak, and Voldemort, while many popular caching technologies act as key-value stores, including Oracle Coherence, Redis and MemcacheD.

B. Column Stores

Column stores are also frequently known as wide column stores (or *extensible record stores*). They store data in records with an ability to hold very large numbers of dynamic columns. In these databases the column names and the record keys are not fixed, a record can have billions of columns. This is the reason why wide column stores can be seen as two-dimensional key-value stores. Google's BigTable is considered to be the origin of this class of databases. There are schema-free like document stores type databases, but with the different implementation. Most popular examples are: Cassandra, Hbase, Accumulo and Microsoft Azure Cosmos DB.

C. Document Stores

Document stores (or *document-oriented database systems*) are characterized by their schema-free organization of data. In these databases the records do not need to have a uniform structure, different records may have different columns. The types of the values of individual columns can be different for each record. The basic unit of storage in the database is a complete document, often stored in a format such as JSON (Javascript Option Notation), BSON (Binary JSOB), XML, or YAML. Document databases are recognized as powerful, flexible and agile tools for storing huge amount of data. Most popular examples are: MongoDB, Couch DB, Simple D etc.

D. Graph database

Graph DBMS (graph-oriented DBMS) represent data as graphs with nodes and edges. Edges are relationships between nodes. Both nodes and edges can have properties. They allow easy processing of data in that form, and simple calculation of specific properties of the graph, such as the number of steps needed to get from one node to another node. Most popular examples of these databases are: Neo4j, FlockDB, Polyglot, Microsoft Azure Cosmos DB, ArangoDB, OrientDB, and Virtuoso.

E. Search Engines

Search engines are NoSQL database management systems dedicated to the search for data content. In addition to general optimization for this type of application, the following features are also supported: support for complex search expressions, full text search, stemming (reducing inflected words to their stem), ranking and grouping of search results, geospatial search and distributed search for high scalability. Most popular examples of these databases are: Elasticsearch, Splunk, and Solr, MarkLogic, and Algolia.

V. DATABASES AND IOT

The RDBMS commonly operate through the transactions. Their role is to keep database in a valid state after each operation. This is done by adhering to ACID features (Atomicity, Consistency, Isolation, and Durability). Side effect of applying this principle is that any operation that modifies data takes longer time to complete. The system has to wait until it receives confirmation of successful data entry to continue processing.

This situation will be exacerbated when there are multiple nodes, because all nodes have to be synchronized before the operation is completed, thereby time taken will be increased considerably. This makes the traditional RDBMS less suitable for deployment in IoT systems. In scenarios when responsiveness of the systems is critical, massive amount of data needs to be written and processed, the usage of traditional RDBMS would cause long response times.

In order to tackle this problem NoSQL database systems don’t enforce consistency as strongly. They provide what is known as BASE approach, which stands for: Basically available – system is always operational; Soft state – system is not always consistent; Eventual consistency – node synchronization will be performed at a later time [4].

As it has been explained briefly in the previous section, most NoSQL databases operate using one of the following data models to organize data [12],

- Key-value stores,
- Document stores,
- Column stores,
- Graph stores.

Important aspect of a NoSQL database is how it handles data replication. Replication is a process in which copies of data are distributed to other nodes. This process synchronizes nodes and brings them to a consistent state, which in turn gives the system fault tolerance. This approach allows the system to operate even if a node fails. There are different ways to implement replication, for instance in Cassandra replication factor is applied to all data items, which specifies how many copies of data will exist among other nodes. To contrast Cassandra’s approach, MongoDB uses master-slave roles, so master has to propagate updates to all of its slave nodes.

For consistency Cassandra uses a quorum system, where write operations are completed only when a defined quorum is achieved. Quorum in Cassandra is achieved when more that 50% acknowledgement is the same (for example 2 out of 3 nodes). On the other hand MongoDB provides option for immediate consistency so master immediately replicates changes to slave nodes. This behavior in Cassandra can be achieved in with the quorum all mode.

From the data structure perspective due to their static schema approach, the RDBMS are additionally less suitable for IoT applications. On contrary, the non-relation, schema free, no joins, easy replication support, horizontally scalable, etc. [13] makes the NoSQL databases much better choice for IoT systems.

In order to find the answer, if NoSQL perform better than SQL in all application scenarios, the research [13] compares SQL and NoSQL databases for a small scale IoT application of water sprinkler system and investigates whether NoSQL performs better than SQL in different scenarios.

The comparison study is based on the time taken to execute Select and Insert queries against varying number of records and threads. In the case where the number of records have increased, also the load on the system has increased, further increasing the response time. According to the results of the study each database has its own pros and cons. In some scenarios, MongoDB provided better response time compared to MySQL, but MySQL responses were more uniform as compared to MongoDB. Therefore, the authors conclude that choice of a database for IoT depends on which query is mostly used and the type of workload of the rest of the application.

Figure 2 presents the typical IoT system with the placement of distributed data storages in the Middleware layer. Briefly, the architecture of presented system, in the figure is as follows: In the Perception layer (I), the sensor nodes are deployed with connectivity module; The Transport layer (II) provides connectivity between the Perception and the Middleware layer where the core of the system is placed (III). In the Middleware layer are placed the Gateway devices (2), for accepting and forwarding data to the core, server farms (3), distributed databases and storage systems (4) and web application servers (5).

The analyzed and formatted data is distributed to the end user via web services in the application layer (IV). Business layer uses the analyzed data and reports, for delivering strategic decisions and managing the observed system.

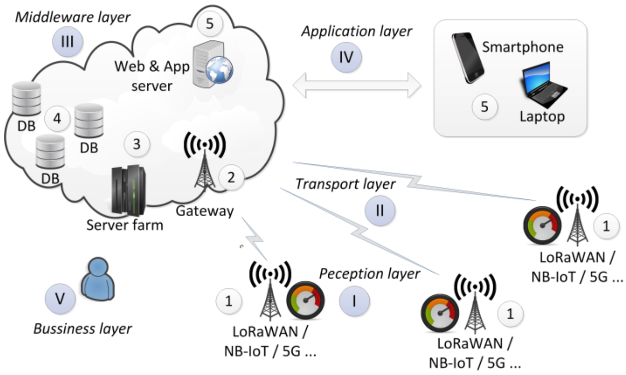


Figure 2 Data store position in typical IoT systems

VI. CAP Theorem and NoSQL

In previous section the authors have touched the CAP theorem and its role in distributed systems. The CAP theorem was introduced in year 2000 by Eric A. Brewer and it states that at any point in time, a distributed system can provide two out of three characteristics [4, 14]. Those are consistency, availability and partition tolerance which are defined as follows:

- Consistency – all users see same data no matter which node of the system they read from.
- Availability – a read request is fulfilled even if a numbers of nodes are down.
- Partition tolerance – system must remain fully functional despite any communication errors causing nodes to be temporarily unreachable.

The CAP theorem is presented in Figure 3. Insisting on consistency means losing on performance, as system has to forward the request to other nodes and wait until it has been acknowledged by all the replicas, thus taking more time. Giving precedence to availability will make it so that client always has access to the system, but there is no guarantee that response won't contain stale data – data which was not updated to new values.

Considering that most NoSQL databases are designed at their core to be distributed onto large number of nodes it is very unlikely to expect them to give up partition tolerance. Doing so would be to expect complete lack of packet losses or disconnections from the network, both of which are unattainable in real life. For these reasons most NoSQL databases are classified as CP or AP, based on whether they focus on consistency or availability along with partition handling, while small number fall into CA category [14].

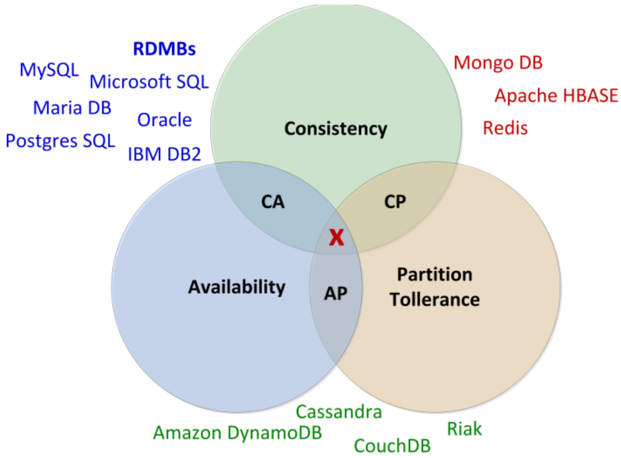


Figure 3 Graphical presentation of CAP theorem

As it is shown in Fig 3 Cassandra, Amazon DynamoDB, Riak, and CouchDB are classified as AP, and Mongo DB, Amazon HBASE and Redis are classified as CP.

A. Using Cassandra in IoT systems

Apache Cassandra is a distributed NoSQL database designed to handle huge amounts of data, providing extensive scalability and data availability options. The potential usage of Cassandra database in IoT systems is discussed in [15].

One of the key features of an IOT based application is the ability to process data in real time. Processing the data in real time requires an ability to process tens of thousands of records per second. To process data at this velocity can be quite a challenging task. The ability to mine data, which helps various institutions to gain vital insights and make decisions, is also very important ability. Cassandra provides a number of features which apply to this scenario.

Cassandra is able to handle this data velocity and provides a good modeling platform for time series data. This is possible because individual rows in Cassandra can have dynamic number of columns. Cassandra allows tuning its consistency thus providing an option to choose consistency or availability according to CAP theorem. Cassandra supports dynamic addition/removal of data nodes and provides features like storage compression, Time-to-Live, etc [15, 16].

Some of the features of Cassandra architecture are:

- Linear performance growth with additional processing capacity
- Flexible schema
- There are no master or slave nodes, full multi-master architecture

- A ring-type architecture (nodes are logically distributed like a ring),
- Automatic workload and data distribution across all the nodes,
- Online cluster resizing (scale out or scale in),
- Replication of data across the nodes for redundancy,
- Tokens (hash based) are used to distribute the data among nodes in the cluster,
- The Cassandra's architecture supports multiple data centers [17] and data can be replicated between data centers (Figure 4).

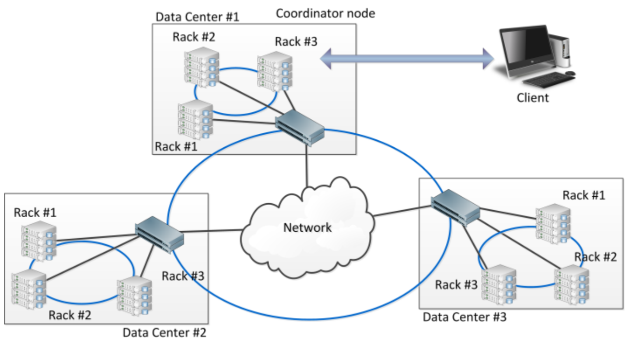


Figure 4 Cassandra ring-type architecture in multiple data centers

Physically, Cassandra servers are placed in racks. A rack is a logical grouping of machines into a physical housing with distinct fault regions (like top of the rack switches, PDU, etc). Similar approach is used for network topology and infrastructure design. Each machine in the rack is computer with its own CPU, memory, and storage. The usage of SSD devices is highly recommended.

Storing data in Cassandra involves several steps. The data is written to a *commitlog* on persistent storage on all the participating members of the write process (e.g. quorum nodes). A write acknowledgement is sent to the client after multiple nodes confirmed a successful write. Choosing which member will participate in this operation is based on the hash value of the partition key (token). Nodes write data to an in-memory table called *memtable*. From the *memtable*, data is written to an *sstable* in memory. The *sstable* stands for Sorted String table. The SST structure represents a pre-formatted structure of the data on the storage. A group of *sstables* on the storage is visible through the Cassandra system s one uniform table.

If a replica partner (node that stores one copy of the data) is not responsible, it is down or temporary unavailable, the data will be written to the coordinator node (the node that facilitates the quorum between the replica members). The coordinator node will hold the data temporarily till the responsible node comes alive or the data expires.

The Cassandra read process is designed to ensure fast reads. Read happens across all replica members in parallel. If a node is down, the data is read from the remaining replicas of the data set. Priority for the replica is assigned on the basis of distance.

Just as the write process has, the Cassandra read process also has several steps. Data on the same node is given first preference and is considered data local. The second preference is given to the data on the same rack and it is considered rack local. Third preference is given to the data on the same data center is and it is considered data center local. The least reference is given to the data in different data center. Data in the *memtable* is checked first, and then the *sstable* to ensure the data can be retrieved faster if it is already in memory. [18]

The example of the Cassandra cluster is given if Fig. 5. It has two data centers: data center 1 and data center 2. Data center 1 has two racks, while data center 2 has three racks. Eighteen nodes are distributed across this cluster.

The first copy (replica) is stored on node 1, the second copy is stored on node 3 and the third copy is stored on node 4 (all nodes are in the data center 1). The fourth copy is stored on node 11 in the data center 2. If a client process is connected to the node 3 and wants to access data stored at node 3, node 3 will be given the highest preference as the data is local here. [18]

The next preference is node 1 where the data is rack local. The next preference is given to node 4 because it is the same data center. The least preference is given to node 11 that is in a different data center. So the read process preference in this example is node 3, node 1, node 4, and node 11 in that order. For additional High Availability (HA) design suggestion check the architecture proposal in Section VI.

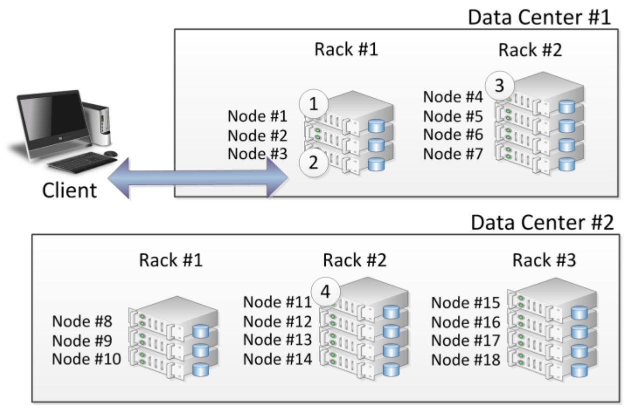


Figure 5 Typical Cassandra deployment in multiple data centers with rack and node organization

B. *Using Cassandra in IoT systems*

In this sub-section the proposal of deployment of a distributed Cassandra DB in typical IoT system is given. The proposal is given according to the author's research findings and working experience.

Cassandra is very applicable in IoT based monitoring systems, with large area coverage (middle sized or large cities) e.g. systems for air pollution, infrastructure health or living condition monitoring (UV index, noise pollution, waste management, etc,) where the larger number of sensors are deployed all around the city and data is frequently sent (in relatively short intervals) to the middleware layer.

In order to achieve the consistency of the data the deployment of the Cassandra system should be considered with the following principles:

- At least 3 copies of the data (number of replicas or RF),
- 3 or more racks deployed per datacenters (number of racks should be equal or greater than number of replicas - #R >= RF),
- Number of nodes per rack depends on the total amount of the data per replica, and can be adjusted as needed,
- Between 500 GB to 1 TB of storage space per node with up to 50 percent utilization (the extra space is needed for efficient compaction),
- In order to achieve better High Availability (HA) the #R >= RF principle should be followed, for example, in Fig. 6 one data center is represented with three racks (three nodes per rack) layout, and the other data center with four racks with three nodes each,
- Equal number of nodes per rack in the same data center is highly recommended,
- If an expansion of a cluster needed equal number of nodes should be added to each rack in a data center.

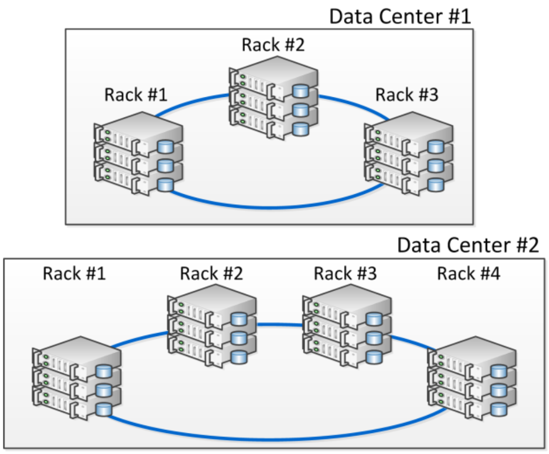


Figure 6. Balanced Cassandra deployment in multiple data centers

In order to achieve good performance with Cassandra, the data model needs to be built according to queries that are planned to be executed. Due to Cassandra's high speed write path one can choose to de-normalize the data model in order to gain read performance. [15]

The key points in using Cassandra in IoT systems are as follows [15]:

- Data modeling should be based around query patterns and following application flows,
- Data duplication is required to ensure data availability,
- Data needs to be distributed evenly across all the nodes in the Cassandra cluster and each node should contain around the same amount of data,
- One of design goals should be to minimize reading from multiple partitions.

VII. Conclusion

This paper gives a brief overview of position and potential deployment of NoSQL database in distributed IoT system. After the general discussion on the most popular database systems, a brief explanation of the features of different NoSQL systems is given. Next, the CAP theorem basics are given, together with a discussion about the relation between CAP theorem and NoSQL databases. Finally, the usability of Cassandra is explored together with a proposal of its deployment in before mentioned IoT systems. This paper and the presented proposal is the result of authors' academic research and practical experience in working with the databases.

References

[1] D. Dobrilovic, "Networking Technologies for Smart Cities: An Overview," Interdisciplinary Description of Complex Systems - scientific journal, Croatian Interdisciplinary Society Provider Homepage: http://indecs.eu, vol. 16(3-A), pp 408-416, 2018.

[2] R. Sahal, J. G. Breslin, M. I. Ali, "Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case", Journal of Manufacturing Systems, Vol. 54, 2020, pp 138-151,, https://doi.org/10.1016/j.jmsy.2019.11.004.

[3] N. M. Kumar, P. K. Mallick, "The Internet of Things: Insights into the building blocks, component interactions, and architecture layers", Procedia Computer Science, Vol. 132, 2018, pp 109-117, https://doi.org/10.1016/j.procs.2018.05.170.

[4] A. Makris, K. Tserpes, V. Andronikou, D. Anagnostopoulos, "A Classification of NoSQL Data Stores Based on Key Design Characteristics", Procedia Computer Science, Vol. 97, 2016, pp 94-103, https://doi.org/10.1016/j.procs.2016.08.284

[5] P. Sethi, Smruti R. Sarangi, "Internet of Things: Architectures, Protocols, and Applications", Journal of Electrical and Computer Engineering, vol. 2017, Article ID 9324035, 25 pages, 2017. https://doi.org/10.1155/2017/9324035

[6] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang and W. Zhao, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications," in IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1125-1142, Oct. 2017, doi: 10.1109/JIOT.2017.2683200.

[7] R. Khan, S. U. Khan, R. Zaheer and S. Khan, "Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges,the " 2012 10th International Conference on Frontiers of Information Technology, Islamabad, 2012, pp. 257-260, doi: 10.1109/FIT.2012.53.

[8] DB-Engines Ranking, https://db-engines.com/en/ranking, Retrieved, October 2020.

[9] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, S. Belfkih, "Comparison and Classification of NoSQL Databases for Big Data", International Conference on Big Data, Cloud and Applications BDCA 2015, Tetuan, Morocco, 2015

[10] N. Neeraj, "Mastering Apache Cassandra", Packt Publishing, 2013.

[11] J. Carpenter, E. Hewitt, "Cassandra: The Definitive Guide", 2nd Ed., O'Reilly Media, 2015

[12] H. Khazaei, M. Fokaefs, S. Zareian, N. Beigi-Mohammadi, B. Ramprasad, M. Shtern, P. Gaikwad, M. Litoiu, "How do I choose the right NoSQL solution? A comprehensive theoretical and experimental survey", Big Data & Information Analytics, 1 (2&3), pp 185-216, 2016, doi: 10.3934/bdia.2016004

[13] S. Rautmare and D. M. Bhalerao, "MySQL and NoSQL database comparison for IoT application," 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, 2016, pp. 235-238, doi: 10.1109/ICACA.2016.7887957.

[14] A. Corbellini, C. Mateos, A. Zunino, D. Godoy, S. Schiaffino, “Persisting big-data: The NoSQL landscape”, Information Systems, Vol. 63, 2017, pp 1-23, https://doi.org/10.1016/j.is.2016.07.009

[15] A. Bane, Supporting IOT Applications with Cassandra - Cassandra Optimisations for IoT Applications, https://blog.thinkitive.com/supporting-iot-applications-with-cassandra/, February 20, 2020.

[16] The Apache Software Foundation. Appache Cassandra, https://cassandra.apache.org/doc/latest/architecture/, Retrieved October 2020.

[17] M. Ben Brahim, W. Drira, F. Filali, et al. “Spatial data extension for Cassandra NoSQL database”, J Big Data 3, 11 (2016). https://doi.org/10.1186/s40537-016-0045-4

[18] Simplilearn Solutions, “Apache Cassandra Architecture Tutorial”, https://www.simplilearn.com/cassandra-architecture-tutorial-video, retrieved October 2020.

The impact of modern technologies on IT projects success

T. Zorić*, V. Makitan* and E. Brtka*
* University of Novi Sad/Technical Faculty “Mihajlo Pupi”, Zrenjanin, Serbia
tamarazoric27@gmail.com, vesna@tfzr.uns.ac.rs, eleonorabrtka@gmail.com

Abstract - The Information Technology industry is widely known as the industry where projects are subject to risk and failure. On the other hand, the development of Industry 4.0 brings revolutionary changes in the field of project management. By exploring the connection between the IT industry and Industry 4.0 technologies, the possibility of applying modern technologies in project management with the aim of achieving project success, is reviewed. Although Industry 4.0 includes a number of different technologies, this paper focuses on the Cloud Computing, Internet of Things and Artificial Intelligence.

I. INTRODUCTION

With the intention of achieving their goals, an increasing number of companies are starting to realize their business through projects. This form of business realization is usual for organizations, companies and institutions in the field of information technology (IT) industry. Nevertheless, achieving the success of IT projects is a challenge that makes difficulties for many organizations in obtaining their goals. As the failure of IT projects has become a recognizable trend in the IT industry, a new challenge that many researchers are dealing with, appeared. There are numerous researches that are identifying the obstacles of project success and ways to overcome them. This is presented in the form of guidelines whose application may enable project success. Yet the IT industry remains an industry marked by projects that are prone to risk and failure.

The revolutionary changes that have taken place in recent decades are the result of the intensive development of new technologies, commonly referred to as Industry 4.0 technologies. The issue of the consequences of these changes is the subject of research in various areas, one of which is the area of project management. When introducing these technologies into the field of project management, it was assumed that the intensity of their impact would be this large that researchers [1] define a new era in the development of project management called Project Management 4.0. Researchers and experts in this field pointed out that IT have an impact on project organizations and on the issue of project management as well, considering that they have a crucial role in increasing project complexity, but at the same time provide new solutions and tools to support project management. The resulting changes require a new perspective and a new approach in project management.

The use of technology is a key aspect of project management, and numerous innovations in technology introduce new threats, but also new opportunities in project management practice. Organizations that do not adapt to the changes are in danger of losing their market position and disappearing. In order to monitor technological changes and eventually achieve the success of projects based on them, organizations are beginning to invest in the drivers of these changes – modern technologies. Although there are researches that provide better understanding of project management tools and techniques, modern technologies and their application are not fully explored and their impact and contribution is not completely understood. Lack of knowledge and understanding of their influence and contribution, leads to their selection based on popularity. However, the only way to have a positive impact on the implementation and project management itself is to choose the right technologies and use them in the right way.

This paper deals with the connection between the IT industry and Industry 4.0 technologies, as the possibility of applying modern technologies in project management with the aim of achieving project success. Although Industry 4.0 includes a number of different technologies, this paper focuses on the Cloud Computing, Internet of Things and Artificial Intelligence.

II. RELATED WORK

Organizations invest significant resources in technology initiatives with the intention of achieving an increase in a number of parameters that describe successful business. However, [2] emphasizes that the technologies applied in an organization must be developed with the aim to achieve certain business or project goals. Considering the IT project management aspect, [3] points out that tools and techniques help project managers and the project team to perform tasks in all 10 knowledge areas defined in the PMBOK Guide. In this paper, the results of a survey [4] conducted in 2006, which included 753 project and program managers and in which it was necessary to assess how certain technologies affect the improvement of project success was presented. The results of that research defined the so-called super tools, as the tools that are most used and have the greatest impact on improving the project success. They include tools for scheduling tasks, determining scope, analyzing requirements, and tools for lessons learned reporting. However, in addition to these tools, there are a number of other tools that contribute to the projects success [3]. The list of all technologies and tools can be too extensive, which is why it is important for project managers to consider and understand which tools and technologies will contribute the most to their project.

Although tools and technologies are often used, it is considered that examination of their contribution to project management has not been sufficiently researched [5]. Lack of knowledge and understanding of the impact of technologies, tools and techniques puts project managers in a situation where they make their selection based on popularity rather than a real contribution to the project success. Industry 4.0 brings many challenges and opportunities, which enable organizations to achieve a competitive advantage. However, it is important to consider how specifically modern Industry 4.0 technologies contribute to projects and how they affect project success.

For IT projects, as for any other projects, communication, coordination and cooperation are considered vital for the project success [3]. For this reason, it is crucial that project managers communicate effectively, which means an effective and complete exchange of information within the team, as well as between project team members and stakeholders. The project result may be completely compromised if communication, coordination and collaboration are not enabled in the right way [6].

A. Cloud and IoT

As researchers pointed out [6-8], the use of the *cloud* in project management increases the efficiency and effectiveness of the project realization as the cloud enables easier communication. On the other hand, the application of *Internet of things* (IoT) technologies also contributes to overcoming various problems in terms of communication. Taking into account the basic concept of IoT, which assumes that all smart devices are integrated into the network, it is clear that a group of people can connect to each other to exchange relevant data at a given time and thus use all the possibilities offered by this technology [9-11]. By connecting people into a group in this way, they get a basis for communication on the project and project management.

Taking into account that the project has a common goal for different stakeholders, coordination and cooperation with all stakeholders are considered to be key aspects for the project success and achieving that common goal, especially in case of projects where the team, clients and suppliers are not in the same location. Unlike traditional project management where cooperation took place at a specific location, modern project management often addresses the issue of how to achieve remote cooperation with stakeholders.

IoT technology supports all the connections needed by project participants to a much greater extent than any other IT system. Specifically, IoT provides support to all stakeholders throughout the project lifecycle by providing a means to achieve collaboration at any time [9-12].

On the other hand, the authors [6-8, 13] address the question of how the cloud enables real-time collaboration from different locations. Collaboration through this technology refers to the support that the cloud provides to business as well as resource management. Specifically, the cloud enables the sharing of documentation, video content, presentations, photos, calendars, and other resources relevant to project management. In addition, the cloud provides remote access to the most up-to-date versions of project support applications. This technology allows project managers to organize meetings through web conferences and generate framework for joint project activities. A special contribution to cooperation with stakeholders is provided by the use of artificial intelligence technologies [14]. It is believed that the use of chatbots, as a form of application of artificial intelligence, can increase the efficiency of human interaction on the project.

Another issue from the aspect of cooperation and coordination on the project is the issue of managing virtual teams. Regardless of location team work enables increasing of productivity and work flexibility [6]. For this reason, the focus of many cloud-based applications is on the availability of resources, business solutions, development platforms, and on-demand infrastructure [15]. The use of the cloud in project management has a major impact on collaboration within the team, which [15] describes as the use of collaboration software that brings together a virtual workforce and tries to make the collaborative experience become natural and productive in the same way as working on the same physical location. In addition, virtual technology, as part of the cloud, influences employees to be able to choose the way of working that best suits their needs [6]. On the other hand, IoT technologies enable better team performance by enabling the inclusion of smart devices in the process of communication and collaboration [12]. Specifically, using smart devices, it is possible to participate in video conferencing, share information and resources related to the project. The use of smart boards reduces the need to transcribe text, as this technology allows text to be transferred to a project specification by adding schemes and images created during brainstorming meetings. These technologies remove all barriers and therefore they are considered to be key technologies to support virtual teams. IoT technologies enable greater team efficiency, allowing them to work faster by reducing the time spent on certain tasks, preventing the execution of tasks that are not necessary and, finaly, increasing balance and reducing tension in the team. Based on the identified possibilities of modern technologies, various researchers [6, 8, 12, 15, 16] conclude that these technologies increase the level of team collaboration on the project, better performance and greater efficiency and effectiveness of team work.

According to the previous considerations, it is clear that numerous problems in the implementation of IT projects can be solved by using these technologies. Researchers [12, 14, 17] conclude that these technologies have a positive impact on solving problems such as the problem of customers input lack, incomplete or variable requirements and specifications, as well as unclear project objectives. Improving these aspects through modern technologies is achieved by involving stakeholders from all levels of the project in project management activities. Stakeholders can gain access depending on the granted access rights. This allows the project team an insight into the project requirements in real time and in all changes that occur during the project realization. This creates transparency between the client and the company as clients can see the project progress status at any time. The same opportunity is provided to project managers by allowing them to review project progress and problems. The possibility of involving clients in the project management process suggests that these technologies can be applied to implement agile methodologies. This contribution of technology is very important for the success of projects since agile methodologies are recognized as methodologies that enable improvement in various aspects of management, such as time savings and budget savings [18].

B. AI

Artificial intelligence technology is recognized as a technology that can significantly improve project performance, while a high degree of performance can create value for clients and thus help organizations better position themselves in the market [14]. First of all, this technology enables the automation of various project activities, through which up to 80% of routine work can be eliminated [14]. In this way, artificial intelligence technology provides an opportunity for project managers to perform other, more complex tasks. However, research results [14, 19] show that artificial intelligence has evolved from simple task automation to predictive data analytics. This technology enables assessments based on information and knowledge gathered from previous projects. Research on the use of technologies in project management [2] emphasizes that the use of technology that has access to and works with historical data and organizational and tacit knowledge creates the possibility of improving project performance and cooperation, building trust and better decision making. Further linking these technologies with technologies that enable open communication and trust building (such as cloud and IoT) can further lead to the exchange of knowledge and experience. In this way, modern technologies have an impact on information gathering and facilitate the development and knowledge exchange.

Artificial intelligence enables predictive analysis of the project based on all collected data [2, 14, 19]. Estimates can refer to simple estimates such as project cost estimates and time required. Using artificial intelligence, software can also predict the trend of project results, the probability of realization and similar parameters. Based on historical data from previous projects, as well as the collected knowledge, technologies allow developing project plans more precisely and accurately, project scope and costs estimation, schedule creating, resource management, and above all to selecting projects using decision making support systems. Another way of applying historical data with the help of artificial intelligence refers to the selection of managers and team members [20]. This refers to the area of people management as resources on the project, and includes the deployment of people to adequate positions, according to the activities they will perform and the technologies they use. Accordingly, artificial intelligence technologies help the decision-making process, which is very important for project management since research [21] indicates that decision-making is one of the most important activities of a project manager. In general, it can be concluded that artificial intelligence also has an impact on decision making, problem solving, project planning and its monitoring.

Modern technologies have an impact on project-oriented companies, which today represent the widest segment of corporations. Their impact is considered to be radical changes since the changes that occur affect all activities in companies, strategic decision-making, investments and above all productivity and work efficiency [22]. The impact of modern technologies on the success of IT projects can be observed through their impact on numerous project success factors. As it is shown in the paper, a review of the literature has identified a number of studies that suggest that modern technologies affect various aspects of project management, which can be identified as factors of project success. Considering ten success factors defined by [23]: project mission, higher management support, project plan/schedule creating, consulting with clients, team members, technical tasks, clients approval, monitoring and feedback, communication and problem solving; and based on presented researches, it may be concluded that observed technologies have influence on these success factors. Connecting people and providing the environment needed for their communication, collaboration and cooperation, it may be directly influenced on customer consultation, communication, monitoring and feedback. In addition, it is clear that the existence of communication, collaboration and cooperation involves stakeholders throughout the project management process which has a major impact on various success factors, such as understanding the project mission, creating a plan/schedule and accepting the project by the client. As modern technologies enable the automation of certain activities, they provide the possibility of achieving greater efficiency in communication, monitoring and providing feedback. The literature review also identified that predictive project analysis enables better performance in project planning and monitoring, based on which it is possible to make better decisions regarding the project and its management. In this way, modern technologies additionally influence the factors of project planning, project monitoring and problem solving. Modern technologies also have an impact on the factors of team members and technical tasks, since they enable the best way of their selection and management. Finally, the question arises as to what extent these technologies influence the defined factors and through them contribute to the success of the project. The answer may be elaborated in the future research.

III. CONCLUSION

As it has been recognized in the literature and practices that technological progress has a great impact on the field of project management, the motivation for conducting this research aimed at examining the impact of modern technologies of Industry 4.0 on the field of IT project management. The reason why the subject of this research is focused on modern technologies lies in the recognized impact that these technologies have on the project success. Given that IT projects are recognized in practice and in the literature as projects that are highly likely to fail, examining the impact of modern technologies on the issue of success can identify new ways to change this practice and contribute to project success. From the aspect of this paper, cloud, IoT and artificial intelligence technologies are considered as modern technologies.

ACKNOWLEDGMENT

This research is financially supported by Ministry of Education and Science of the Republic of Serbia under the project number TR32044 “The development of software tools for business process analysis and improvement”, 2011-2014.

REFERENCES

[1] Maylor, H., 2001. Beyond the Gantt Chart: Project Management Moving on. European Management Journal, 19(1), pp. 92-100.

[2] Anantatnula, S. V., 2008. The Role of Technology in the Project Manager Performance Model. Project Management Journal, 39(1), pp. 34-48.

[3] Schwalbe, K., 2016. Information Technology Project Management, Revised. 8th ur. Boston, MA: Cengage Learning.

[4] Besner, C. & Hobbs, B., 2006. The Perceived Value and Potential Contribution of Project Management Practices to Project Success. Project Management Journal, 37(3), pp. 37-48.

[5] Patanakul, P. I. B. & Milošević, D., 2010. An Empirical Study on the use of Project Management Tools and Techniques across Project Life-Cycle and their Impact on Project Success. Journal of General Management, 35(3), pp. 41-66.

[6] Francke, E. & Weideman, M., 2012. Adopting disruptive technologies by managing projects in the cloud. Durban, South Africa, Cape Peninsula University of Technology.

[7] Mlitwa, N. & Pekane, A., 2015. Technology innovations and cloud computing in project management: an analysis of the academic and industry perspectives in Cape Town, South Africa. Seville, Spain, s.n.

[8] Sloniec, J., 2015. Use of Cloud Computing in Project Management. Applied Mechanics and Materials, Tom 791, pp. 49-55.

[9] Hu, Z., Yuan, Q. & Zhang, X., 2009. Research on Agile Project Management with Scrum Method. Proceedings of 2009 IITA International Conference on Services Science, Management and Engineering. Zhangjiajie, China , IEEE.

[10] Lou, P., Liu, Q., Zhou, Z. & Wang, H., 2011. Agile Supply Chain Management over the Internet of Things. Proceedings of 5th International Conference on Management and Service Science (MASS). Wuhan, China, IEEE.

[11] Lou, P., Lui, Q., Zhou, Z. & Wang, H., 2011. Agile Supply Chain Management over the Internet of Things. Wuhan, China, s.n.

[12] Andrei, G., Ioan, F. & Florin, D., 2018. A new vision over Agile Project Management in the Internet of Things era. Procedia - Social and Behavioral Sciences , Tom 238, pp. 277-285.

[13] Asava, R. & Mzee, H., 2010. Cloud Computing meets Project Management. PM World Today, 12(6).

[14] Wang, Q., 2019. How to apply AI technology in Project Management. PM World Journal, 8(3), pp. 1-12.

[15] Dillon, L. & Scanlon, R., 2011. Project management use of cloud collaboration. Paper presented at PMI® Global Congress 2011—EMEA. Dublin, Leinster, Ireland, Project Management Institute.

[16] Botterman, M., 2009. Internet of Things: An Early Reality of the Future Internet, Prague, Czech Republic: European Commission Information Society and Media.

[17] Khan, A., Shaikh, P., Dhembre, C. & Gawali, S., 2011. Cloud Services for Collaborative Web Based Project Management System. International Journal of Computer Science Issues, 8(6), pp. 180-184.+

[18] Charvat, J., 2003. Project Management Methodologies: Selecting, Implementing, and Supporting Methodologies and Processes for Projects. Canada: John Wiley & Sons, Inc.

[19] Lahmann, M., Keiser, P. & Stierli, A., 2018. AI will transform project management. Are you ready?, Switzerland: PricewaterhouseCoopers International Limited.

[20] Pireto, R., 2019. Impacts of Artificial Intelligence on Management of Large Complex Projects. PM World Journal, 8(5), pp. 1-20.

[21] Chen, S. H. & Lee, H. T., 2007. Performance evaluation model for project managers using managerial practices. International Journal of Project Management, Tom 25, pp. 543-551.

[22] Kortuem, G., Kawsar, F., Fitton, D. & Sundramoorthy, V., 2010. Smart Objects as Building Blocks for the Internet of things. IEEE Computer Society, pp. 44-51.

[23] Slevin, D. P. & Pinto, J. K., 1986. Project implementation profile: new tool for project managers. Project Management Journal, 17(3), pp. 57-70.

A Framework Based on Internet of Things (IoT) Technology for Smart Healthcare Services

D.K. Rizk*, H. M. Hosny*, E. M. El-Horbaty** and A. M. Salem**
*Faculty of Computer Science, The American University in Cairo, Cairo, Egypt
**Faculty of Computer & Info. Sciences, Ain Shams University, Cairo, Egypt
drizk@aucegypt.edu; hhosny@aucegypt.edu; sayed.horbaty@yahoo.com; absalem@cis.asu.edu.eg

Abstract - Smart healthcare is of high interest to researchers and governments due to the development of new smart cities. The integration of digital healthcare services with web technologies and intelligent computing needs a concrete framework. Nowadays, Internet of Things (IoT) technology enables communication among real world objects to interact in an intelligent way. Accordingly, IoT is all about connectivity and automation that require intelligent algorithms to make the smart objects work together. The IoT technology helps the healthcare sector by offering the appropriate aid to patients and specialists. The main objective of this paper is to examine the different existing frameworks that discuss smart healthcare services and offer a smart framework based on the IoT paradigm.

I. Introduction

Healthcare services have been exponentially increasing worldwide [1] as there is a significant volume of data generated on a daily basis by medical and clinical organizations [2]. This data is important and vital for decision-making [3] and the lack of access to medical information may negatively affect the delivery of the best care for patients [4]. Storing the records of patients electronically [5] facilitates the exchange and availability of information for healthcare processes [6] and hence increases the productivity of any patient care system that takes a central position and provides easy accessibility and usage [5]. The introduction of the most recent technological innovations in cloud computing for the healthcare sector [5] is becoming a pressing requirement in order to optimize the resources in terms of computational and storage capabilities [2]. Cloud computing is a cost-effective means for facilitating data collection, data storage and exchange between healthcare communities [3].

This paper is organized as follows: section 2 presents a brief introduction about the smart healthcare market, section 3 presents a comprehensive analysis of the different frameworks that proposed smart healthcare services during the last years, section 4 shows our results and discussion based on the analysis conducted in the previous section, section 5 presents our proposed framework. Finally, section 6 presents conclusions and future work.

II. Smart Healthcare Services

According to Sundaravadivel et al. [7], smart healthcare is becoming more demanding nowadays due to its enhancement of the user's quality and experience. It also helps in maximizing the available resources to their extreme potential. Smart healthcare provides remote observation of patients and aides in diminishing the expense of the treatment. Additionally, it provides a means for doctors to increase their services without any regional barriers. Figure 1 indicates the wide range of smart healthcare elements, which mainly consist of services, medical devices, technologies used, applications, system management and end users.

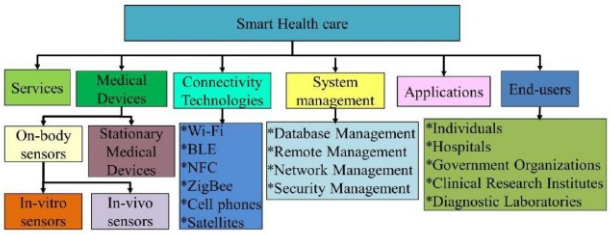


Figure 1: Smart Healthcare Elements (adapted from [7])

Sundaravadivel et al. [7] discussed the features of a smart healthcare architecture. They first talked about the requirements of the architecture and divided it to functional requirements, which should be specific per component, and non-functional requirements that include performance requirements and ethical requirements. The components of the smart healthcare architecture according to Sundaravadivel et al. [7] are sensors and actuators, computing devices, data storage, and networking components. The three categories for classifying the characteristics of smart healthcare are: App- oriented; which is the transmission of smartphone application's data to sensors, Things-oriented; which is application-based adaptation such as real time monitoring, and Semantics-oriented; which is the creation of behavioral patterns dependent on the recently gained data. The framework for the smart healthcare architecture should include the libraries and environments being used in the healthcare platforms.

III. Related Work

A. *Previous Models for Smart Healthcare Frameworks*

In this section, we first discuss some of the previously studied healthcare frameworks where we give a brief description of each one. In section B, we compare the outputs of the discussed frameworks and mention their advantages and disadvantages from our perspective.

1. Reid et al Model

In 2005, Reid et al. [8] introduced a Framework for a Systems Approach to Health Care. The following are the four main levels and Figure 2 illustrate them:

- *Patient:* whose necessities and inclinations ought to be the characterizing factors in a patient-centered health care system. Therefore, the function of the sick person has changed from ordinary patient accepting instructions to a greater energetic participant in care delivery.
- *Care team:* which consists of health physicians, patients' family members, and others who help in offering care to the patients.
- *Organization:* it is the third level of the health care system such as hospitals or clinics. They are responsible for the infrastructure and additional resources needed for the development of care teams and clinical microsystems.
- *Environment:* it consists of the political and economic environment that incorporates administrative, money related, and installment systems. It also consists of entities that affect the structure and execution of health care organizations and all other levels of the system.

According to Reid et al. [8], patients should have access to the same information offered to the physicians and care team in order to be interrelated to the health care system, which leads to speeding up the process of diagnosis and treatment. Moreover, in order to have a professional care team system, physicians in this system should have on-request access to basic clinical and authoritative data, as much as the executives, correspondence, and choice help data, along with having access to the instructive devices to blend, dissect, and utilize that data. They think that in order for care teams or clinical microsystems to become the primary agents of patient-centered clinical care then the rules of engagement between them and the patients should change. For example, care teams have to be receptive to the requirements and inclinations of patients and include them and their families in the plan and execution of care. Moreover, care groups should give patients ceaseless, advantageous, convenient access to quality care. One individual from the care team has to be liable for guaranteeing successful correspondence and coordination between the patient and different individuals from the care group.

Reid et al. [8] placed the organization at the third level of the health care system as it includes the decision-making systems, information systems, operating systems, and procedures to facilitate the exercises of various care groups and supporting units and deal with the flow of human, material, and money resources and information on the side of care groups. The organization is the business level, the level at which most investments are made in information systems and infrastructure, process management systems, and system tools. However, the environmental level, which is the last level of the health care system, has many factors and forces such as some administrative approaches that do not support the goals and objectives of the patient-centered, superior well-being care organizations or the social insurance conveyance framework.

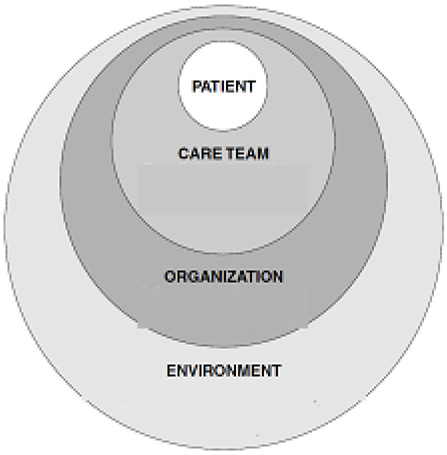


Figure 2: Model of Smart Healthcare System (adapted from [8])

Finally, Reid et al. [8] suggested that the usage of information/communications systems helps in speeding the exchange of information such as:

- Patient-based information: laboratory results, current diagnoses and medications
- Institution-based information: drug-resistance to different antibiotics
- Profession-based information: clinical-practice rules along with best practices for different circumstances
- Real-time decision support: cautions about potential medication cooperation or dosing designs in a patient with an undermined medication digestion system
- Practice-surveillance support: updates about forthcoming screening tests prescribed for a patient
- Population health data: for epidemiological research, disease and biohazard observation, and notice of post presentation adverse medication occasions.

2. Demirkan Model for Smart Healthcare Systems framework

In 2013, Demirkan [9] proposed a "Smart Healthcare Systems framework for conceptualizing data- driven and mobile- and cloud-enabled smart healthcare systems." According to Demirkan [9] "Healthcare spending increases reflect longer life expectancies, advances in developing countries' standard of living, and the corresponding ability to afford high-quality medical treatments, and technological advances that create new possibilities for curing diseases and delivering services."

Cloud plus electronic health records plus medical sensors / wearable devices leads to delivering automated, intelligent, and sustainable healthcare services. A Smart Healthcare Systems Framework (SHSF) offers healthcare organizations to deploy platform-, technology-, and location-independent solutions. SHSF is divided into three major layers: business processes, service-oriented architecture, and service-oriented infrastructure all under a cloud services execution architecture that manages services to each layer's role. SHSF also, includes a conceptual model of an intra- and inter- organizational business process. Moreover, one of the main components of the SHSF's services and resource execution architectures is the inter-organizational supply chain, which is based on using smart docking stations.

Demirkan [9] proved that the adoption of smart healthcare systems through offering four organizations who used parts of the SHSF or a similar framework namely are:

a. The EU-supported Smart Health project focusing on cancer diagnostics, which offered an open integrated architecture, which showed that healthcare quality highly, improved with earlier detection.

b. The University of Pittsburgh Medical Center's Smart Room technology, which used an intelligent system with three different screens for patient, caregiver, and nurses and aides. Each of these screens offers the appropriate patient's data according to the user.

c. The Mayo Clinic's data-driven care, which used an algorithm for brain aneurysms (swelling of blood vessel in the brain) detection and the results, were 95% accurate.

d. Ambient Assisted Living, which is a home-care mobile monitoring solution, resulted in effective services.

3. Basaez et al Model

In 2014, Basaez et al. [10] carried out an exploratory study to evaluate the transition progress of traditional e-healthcare system into smart healthcare services. They reviewed the implementation of different information and communication technologies (ICTs) such as internet of things (IoT), cloud computing, and the usage of smart item technology such as wireless sensor networks (WSN), and radio frequency identification devices (RFID). They showed that the enhancement of e-healthcare system to smart healthcare services, leads to empowering physicians to follow-up the patients remotely while carrying out their ordinary day schedule. Accordingly, the patient receives high-caliber healthcare services along with saving time and eliminating health costs leading to the enhancement of healthcare quality service and the gratification of patients.

Basaez et al. [10] extracted from the many researches they reviewed the essential elements to propose a complete cloud-based framework for smart healthcare services. This framework endeavors to represent the new extent of e-healthcare by using WSN-RFID smart items and representing a smart healthcare scenario. The scenario is based on extracting information from patients using smart items then transferring them to the cloud. Healthcare institutions can then use the patient's personal health information (PHI) in the cloud such as PHR, EMR, EHR to screen patients' crucial parameters and take the appropriate action such as diagnostics, treatment, and rehabilitation. Their framework is based on two main groups of clients: i. Healthcare organizations such as physicians and care staff, and ii. Family unit individuals such as patients and family members. Both groups can access the data at any time through the EMR and Patient care portal, respectively.

Based on Basaez el al.'s [10] readings, their proposed framework is divided into five levels coming from different researches: (a) Patient input level, (b) Smart item level, (c) Data aggregation level, (d) Cloud computing level, and (e) Monitoring output level.

4. AL-Azzam et al Model

In 2019, Al-Azzam et al. [11] proposed how to mingle mobile health with smart cities to get a smart health framework. They explained how the healthcare services have developed from doctor's visits to patients at home up until the amplified m-health which results in s-health. The example used for the extended m-health was how the traffic lights can be adjusted if a cyclist did an accident and the smart city is notified through a wearable band at the hand of the cyclist in order for the ambulance to reach the cyclist in minimum time. Moreover, they made a comparison between m-health and s-health from the source and flow of information point of views. M-health source of information comes from the patient while in s-health it comes also from the smart city sensing infrastructure. Regarding the flow of information, m-health is user-focused or personalized approach, while s-health has the city-centric as an additional approach. Al-Azzam et al. [11] presented two of the challenges that can face s-health which are multidisciplinary research and interaction, and security and privacy. On the other hand, they presented 7 opportunities for s-health namely; (a) Data collection, presentation, and analysis, (b) Prevention as well as administration of critical incidents, (c) Effectiveness and environmental assessment, (d) Engaging patients and families in managing their health, (e) Improving policy decisions, (f) Epidemic control, and (g) Cost saving.

5. AlAbdulatif et al Model

In 2019, the importance of analyzing health data is not only to increase the quality of patient care, but also to decrease healthcare costs. Therefore, the arrangement of smarter and more money saving healthcare services are becoming highly required. Moreover, since the Internet of Things (IoT) devices are producing a huge amount of data and transferring them to the cloud computing for storage and analysis, then this will delay the development of healthcare services that relies on time.

Accordingly, the authors [15] presented a framework that collected health data from different wired/wireless sensors and transmitted them to a secure Edge-of-Things (EoT) layer. Its importance does not only lie in transferring data, but it also has the capabilities of real-time analytic services. This layer does a partial real-time analysis on the biosignal data to determine abnormal ones and accordingly send smart decision-making to clients. This layer also, sends the input health data along with the analyzed ones to the cloud computing for further analysis and to be saved due to its huge quantity. Therefore, the EoT acts as a middle computing layer between IoT actuators and the cloud computing as it has the ability of real-time computing and storage properties, but on a smaller scale than cloud computing. Moreover, the EoT can diminish latency and bandwidth consumption by placing the computing power adjacent to the source of data. Then appointing a portion of the time-critical calculation tasks to the edge devices and then transferring the computation-intensive jobs to the cloud computing.

The architecture of their framework is designed to accomplish specific tasks, which are data aggregation, safe storage, and finally analysis of data. The data aggregation is represented in the architecture as the "Community Members (CM)". These are the people whose health data are being collected whether from inside the hospitals or within smart areas. The "Cloud-enabled Database (CD)" is the safe storage area for the encrypted data. Finally, the "Abnormality Detection Model (ADM)" is used to represent the analysis engine in the architecture. They all work together to detect the abnormality of the bio signal data. The results of the ADM can be decrypted safely at the CM side if needed.

The proposed EoT framework uses a Fully Homomorphic Encryption (FHE) to protect the sensitivity of the patient's data and to analyze these data in an encrypted domain. FHE is important because it makes sure that the patients' data are stored securely at the cloud and that it has precise computation capability on the encrypted data.

In order to analyze the large and heterogeneous amount of data, the authors [15] developed a distributed method for clustering-based techniques, which are considered an appropriate tool for providing clinical decision-making. They used two clustering-based techniques are: "K-Means Clustering (KMC) and Fuzzy C-Means Clustering (FCMC)". By using these techniques, the authors [15] were able to detect the normal bio signal patterns and any deviation in the patterns means anomaly.

The main advantage of this manuscript is that the authors [15] did an experimental evaluation to their proposed framework using the "Google Cloud Platform (GCP)" and "a real heart disease dataset from the University of California Irvine's (UCI)". They showed how their framework performs from an accuracy and execution time points of view.

B. Results and Discussion

TABLE I. ADVANTAGE AND DRAWBACKS OF MODELS

Model	Advantages	Drawbacks
Reid et al Model [8]	Combination of four main levels; patients, care team, organization, and environment along with their needs and requirements.	Prioritization of the metrics to optimize the performance of the framework.
Demirkan Model– SHSF [9]	It offers healthcare organizations to deploy platform-, technology-, and location-independent solutions.	It is not a referenced framework where it is based on components that can be misused by the different healthcare organizations.
Basaez et al Model [10]	It represents the extension of earlier e-healthcare by introducing the usage of WSN- RFID smart items in order to reach a smart healthcare scenario.	The difficulties of implementation and expensiveness of the usage of the WSN-RFID smart items across different categories of people such as: gender, age, health condition, and country's health insurance support.
Al-Azzam et al Model [11]	It is how to mingle mobile health with smart cities to get a smart health framework.	The same concept was proposed earlier in 2014 by Solanas et al [12].

IV. PROPOSED REFERENCE FRAMEWORK MODEL

The concept of IoT allows physical objects or things around us to interact with each other in an intelligent way with the aid of key elements like Radio-Frequency Identification (RFID) tags, processors, sensors, actuators, WAN, W-LAN etc., with minimal human intervention. Through IoT, anyone, from anytime and anywhere can have connectivity for anything and it is expected that these connections will extend and create an entirely advanced dynamic network. IoT resides on three basic technology components for the everyday 'things' to be equipped with tracking and sensing. When the 'things' are properly equipped it allows to gather, analyze, communicate and distribute data among themselves. The components include hardware, software, and presentation tools [13, 14].

On this aspect, our proposed model aims to discuss the application of IoT technology in healthcare services. We are now in the process of building a reference framework model to the healthcare organizations in a SMART form and with the exact metrics to achieve the best performance. The metrics used in our model are divided into functional (accuracy of health data, usage of equipment) and non-functional (confidentiality, punctuality).

The framework model will use the Internet of Things (IoT) technology due to its diverse proficiencies in connecting human, machines/healthcare equipment, and smart devices. The IoT gives solutions based on the mix of data provided through the hardware and the software to be either stored, recovered, or processed. In the field of healthcare services, IoT can help in image administration, drug management, and patient flow analysis.

The proposed framework will act as a reference because our ultimate goal is to satisfy the requirements of the different stakeholders in the medical community. We will also take into consideration the importance of the quickness to alert healthcare givers with any variation in the patient's health in order to evaluate the severity of the case leading to save the lives of many patients.

V. Conclusion

This survey showed that there is currently no standard smart framework for healthcare. Our research's goal is to provide the medical community with the accurate and timely information about the patients to take the right decision at the right time. It acknowledges the importance of time and security for critical cases and hence, the data offered to physicians should satisfy the main non-functional requirements of accuracy, punctuality, and confidentiality. This survey should help us identify the basic healthcare service components to pave the way for our research aim, which is to create a cloud computing reference architecture for smart healthcare services framework that captures the best practices and that introduces innovative features to suit the target users.

References

[1] N. Chondamrongkul and P. Chondamrongkul, "Secure Mobile Cloud Architecture for Healthcare Application," International Journal of Future Computer and Communication, vol. 6, no. 3, 2017.

[2] C. Pino and R. Di Salvo, "A Survey of Cloud Computing Architecture and Applications in Health," Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE), 2013.

[3] H. Aziz and A. Guled, "Cloud Computing and Healthcare Services," Journal of Biosensors & Bioelectronics, vol. 7, issue 3, 2016.

[4] M. Kyazze, J. Wesson, and K. Naude, "The Design and Implementation of a Ubiquitous Personal Health Record System for South Africa," Global telehealth, 2014.

[5] G. Reddy and G. Reddy, "Study of Cloud Computing in HealthCare Industry," International Journal of Scientific & Engineering Research, vol. 4, issue 9, 2013.

[6] "IT Reference Architecture for Healthcare," ATOS, 2011. [Online]. Available: https://atos.net/wp-content/uploads/2016/06/atos-itah-architecture-for-healthcare-whitepaper.pdf

[7] P. Sundaravadivel, E. Kougianos, S. Mohanty, and M. Ganapathiraju, "Everything You Wanted to Know about Smart Health Care: Evaluating the Different Technologies and Components of the Internet of Things for Better Health," IEEE Consumer Electronics Magazine, vol. 7, no. 1, pp. 18–28, 2018.

[8] P. P. Reid, W. D. Compton, J. H. Grossman, and G. Fanjiang, "A Framework for a Systems Approach to Health Care Delivery," Building a Better Delivery System: a New Engineering/health Care Partnership – The National Academies Press – 2005.

[9] H. Demirkan, "A Smart Healthcare Systems Framework," IT Professional, vol. 15, no. 5, pp. 38- 45, 2013.

[10] M. O. Basaez, S. G. Muhammad, D. A. Aranda, and V. Stantchev, "Smart Healthcare Services: A patient Oriented Cloud Computing Solution," The 3rd International Business Servitization Conference, 2014.

[11] M. Al-Azzam, and M. Alazzam, "Smart City and Smart–Health Framework, Challenges and Opportunities," International Journal of Advanced Computer Science and Applications, vol. 10, no. 2, pp. 171–176, 2019.

[12] A. Solanas, C. Patsakis, M. Conti, I. S. Vlachos, V. Ramos, F. Falcone, O. Postolache, P. A. Pérez-Martínez, R. Di Pietro, D. N. Perrea, and A. Martínez-Ballesté, "Smart health: A context-aware health paradigm within smart cities," IEEE Communications Magazine, vol. 52, no. 8, pp. 74-81, 2014.

[13] Andrew Whitmore & Anurag Agarwal & Li Da Xu, "The Internet of Things—A survey of topics and trends", Springer Science, New York, 2014.

[14] Somayya Madakam, R. Ramaswamy, Siddharth Tripathi, "Internet of Things (IoT): A Literature Review", Journal of Computer and Communications, vol. 3, pp: 164-173, 2015.

[15] A. Alabdulatif, I. Khalil, X. Yi, and M. Guizani, "Secure Edge of Things for Smart Healthcare Surveillance Framework", IEEE Access, vol. 7, pp.31010-31021, 2019.

Examples of Implementation of Fixed and Mobile Air Quality Monitoring Systems in Urban Areas

Milica Mazalica*, Dalibor Dobrilović*
* University of Novi Sad/Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia
milica.mazalica@tfzr.rs; dalibor.dobrilovic@tfzr.rs

Abstract – Considering the importance of the implementation of air quality monitoring systems, especially in urban areas, this paper gives the presentation of typical systems. Two different approaches in building such systems are used. One approach is focused on fixed and the other approach is focused on mobile sensor networks for air pollution monitoring. In this research, the focus is on analyses of the most common communication technologies and hardware platforms used for the systems, as well as on the systems deployment scenarios. This research and its findings should be used as a first step in collecting the necessary data to help in designing the urban pollution monitoring system and its particular deployment in particular scenarios, e.g. for smart city environments and cities of up to 100.000 inhabitants.

I. INTRODUCTION

In this paper, the focus is on using emerging wireless technologies for monitoring air pollution and air quality. According to many studies, exposure to air pollutants is seriously increasing respiratory and cardiovascular illness and mortality (causing approx. 2.8 million deaths). This problem is even more present in certain parts of the world because the developing countries experiencing the worst air pollution [1]. Thus, urban air pollution remains an important (maybe the biggest) health risk in the world, especially for millions of people in low-income countries. The Global Burden of Disease project estimated the number of 3.2 million deaths and 3.1% of disability-adjusted life-years is caused by outdoor air pollution annually, making exposure to outdoor air pollution the crucial and leading environmental risk factor and one of the most important modifiable risk factors for the global disease burden [2]. Speaking about the pollutants, it can be pointed out that approximately 89% of the world's population is currently exposed to PM2.5 (particulate matter with an aerodynamic diameter < 2.5 μm) concentrations which are above the World Health Organization (WHO) air quality guidelines [3, 4, 5].

On the other hand, it can be said that traditional approaches for measuring air quality based on fixed measurements are not completely adequate for personal exposure monitoring. To target this issue, usage of small, portable gas-sensing air pollution monitoring technologies is increasing. In recent years an increasing number of researchers and individuals employ portable and mobile methods to obtain more spatially and temporally representative air pollution data [6].

To establish an efficient system for air pollution monitoring, data of high spatial and temporal resolution is required. It is also required to improve models of air quality estimation, especially for monitoring of individual exposure to potentially harmful atmospheric gases and particulate matter (PM). Considering the act that atmospheric gases and particulate matter are constantly chemically reacting and evolving in the atmosphere, the deployment of hundreds or thousands of sensors dispersed through an urban environment to provide data simultaneously could be used to build improved models of a pollutant release, formation, transport, and deposition within the urban environment. The combination of smart devices with sensors will allow improved estimates of personal exposure to pollutants which, in turn, will revolutionize epidemiological studies that relate pollutant exposure to human health. Currently, regional monitoring sites are often used to estimate an individuals' exposure to pollutants such as ozone, NOx, or particulate matter. In numerous cases, regional monitoring data may not accurately reflect an individual's true exposure. So, personal based smart sensing platforms can revolutionize our understanding of how exposure to atmospheric pollutants affects human health [7].

This paper gives a short review of different approaches in building both fixed and mobile sensor networks for air pollution monitoring. Those approaches are tailored to the different challenges that are facing solution developers. The examples of some of those approaches are given in this paper. In this research, the focus is on analyses of the most common communication technologies and hardware platforms used for the systems, as well as on the systems deployment scenarios. This paper is structured as follows: in the introduction section, the importance of air quality for pollution monitoring is given, after the short presentation and discussion on key enabling technologies for that systems, the description on example approaches o both portable and mobile system is given. At the end of the paper, the discussion and concluding remarks are given.

II. CONNECTIVITY

There are a lot of technologies that can be classified as short-range. Those technologies are Bluetooth, Blue Low Energy (BLE), Z-Wave, and ZigBee with the range from 10m to 50m in the indoor and in some cases up to approx. 100m in outdoor environments.

ZigBee is a standard that defines a set of communication protocols for low-data-rate short-range wireless networking. ZigBee-based wireless devices operate in 868 MHz, 915 MHz, and 2.4 GHz frequency bands. The maximum data rate is 250 Kbits per second.

ZigBee is targeted mainly for battery-powered applications where low data rate, low cost, and long battery life are the main requirements. The ZigBee basic purpose, as well as its applicability in the industrial environment, is described in [8]. In short, it is designed for: home and building automation, smart energy, personal health care, telecom services, industrial asset management, etc.

Medium range technologies have extended coverage and usually have range from 100 to 300m. This group includes WiFi, ULP (802.15.4q), WI-SUN (802.15.4g), 6LoWPAN, and Wireless M-BUS. In this group, the most used is WiFi technology. This technology is based on IEEE 802.11 standards with numerous sub-standards (IEEE 801.11b, IEEE 801.11a, IEEE 801.11, IEEE 801.11n, etc.). It operates at 2.4 Hz and 5.8 GHz using ISM bands. These bands are subdivided into multiple channels (the number of the channel varies from the region) and each channel can be time-shared by multiple networks. The technology works best in line-of-sight. Some versions of Wi-Fi can achieve speeds of over 1 Gbps in close range.

The third group includes longer range technologies with a range from 1 km to 15, 30, 50, or more kilometers. The group includes/consists of: WiFi Low Power (802.11ah), Weightless–W, Weightless–N, Weightless–P, LTE-M Release 12/13, DASH7, LoRa, LoRaWAN, SigFox, LTE, LTE-A, WiMAX, and Ingenu. It should be noted that: Weightless, LoRa, LoRaWAN, SigFox, and NB-IoT can be called Low-Power Wide-Area Network (LPWAN) [9, 10, 11, 12].

LoRa is one of the varieties of technologies that made an appearance as a Low-Power Wide Area Network (LPWAN) group of technologies [13]. LPWAN initially has the application in wide-area environments such as the Internet of Things covering wide areas and designed for systems such as smart cities and smart transportation. Those environments are designed for outdoor space, and LoRa™ two main advantages such as long-range and small energy consumption are perfect for those environments.

The detailed overview of applicable wireless technologies in IoT systems and their comparison is given in [14]

III. FIXED AND PORTABLE AIR POLLUTION MONITORING SYSTEMS

In this section, the examples of the fixed and mobile system of air quality monitoring will be described.

A. *Fixed air quality sensor networks*

The fixed air quality sensor networks are planned to be located at a number of locations in fixed sites and to monitor the environment over a while. There are no standardized protocols for defining the number of nodes to be placed on a certain area and generally, they do not need to be a part of the network unless they are already interconnected and transmit data to a central location. If the monitoring nodes are not connected they can log the data to the local data stores, and data can be later transferred to a processing center for further analyses.

The study [15] describes the monitoring system used during the Hong Kong marathon. The system monitored gases like nitric oxide (NO), carbon monoxide (CO), ozone (O_3), etc. The purpose of air quality monitoring networks is to produce pollution maps.

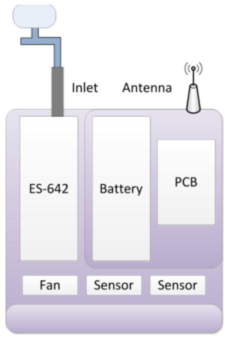


Figure 1. Schematic of mini air station (MAS) system with different components [15]

China has a developed widespread sensor network for mapping urban and regional air pollution. According to rough estimates, there are currently over 30.000 sensors operating to monitor the concentration of air pollutants in China, with more than 10,000 sensors installed in north China. [16]

The PUWP monitor uses a low-cost PM sensor, which measured particle count based on light scattering. This node also has a microprocessor, real-time clock, data logger, temperature and humidity sensor, and small LED display. This platform is used in Xi'an's in China for air pollution measured and concentration PM2.5. [17]

B. *Mobile air quality sensor networks*

Mobile air pollution systems use a different approach. These systems try to cover a large area with a relatively small number of sensors. The sensors are mounted on moving objects such as cars, bicycles, UAV (Unmanned Arial Vehicles), pedestrians, etc. thus achieving larger coverage with the mobility and providing data of high spatial resolution.

Such systems can be designed for decision making in a variety of fields, like the fight against climate change, the protection of health and environment, the optimization of industrial and food processes, or traffic management. In contrast to using high cost and large instruments to measure air contamination the combination of non-specific gas sensors with enhanced features (low cost, low power, low size) combined with pattern recognition methods could develop largely portable intelligent systems [18].

The study in China has been utilized city taxis as mobile platforms for low-cost sensors. In this study, 1,000 taxis were equipped with PM sensors and they crossed the distance of more than 23,000 km which cover 95% of the roads in the city and provided 1.2 million PM data points per day. The results of this study assisted the city Government to develop a more effective air pollution control strategy.

In 2017 Koval and Irigoyen designed and tested UAV-based air pollution monitoring systems using a catalytic sensor (TGS6812-D00) to measure and detect concentrations of hydrogen, methane, and petroleum gas. The data processing was done at the ground station which incorporated a robot operation system coupled whit a drone autonomy package. The results of this study suggested that future improvements will be shaped with the need for low-cost sensors/monitors to be used on UAV platforms. [19]

Sensors/monitors worn or carried by individuals are used for assessment of personal exposure to various types of pollution. The study by Jerrett was conducted in Barcelona during the period from September 2013 to February 2014 with 56 participants. Personal sensing monitors for measuring personal exposure to pollution were used in this study. The monitors collected data every 10 seconds and used Alphasense CO, NO, and NO_2 sensors and also temperature sensors, GPS, and GRPS transmitter. The results show that the system can detect concentrations of the pollutants in different microenvironments. [16]

The AirSense platform for personal exposure monitoring was designed and tested by Zhuang 2015. This platform used sensors for GPS, dust, temperature, humidity, and accelerometer in New York. The AirSense reacted to changes in microenvironments, such as changes in commuting models, activity levels, during activities at home, were tested with collected data in short periods by one participant. The results of this study show suitability of AirSense for personal exposure monitoring as well as for complementing routine ambient monitoring. [20]

Mobile air quality sensor networks use low-cost sensors such as DN7C3CA006 which was built by SHARP. It continually samples the air every 10ms and provides relative consistent readings. The sensor Lighthouse 3016IAQ is used for sensor calibration and system evaluation. These platforms have an advanced portable sensor with 0.1 μg/m3 estimated error. The urban taxis are used to mount sensors and to collect the mobile data in real-time. These sensors are equipped with low-cost PM2.5 sensors, GPS, control, transmission modules, and power interface, and they can be shared by vehicle igniter.[21]

The small unmanned aerial vehicles (UAV) with air quality sensors, allow precise characterization of blasting plumes in near-real-time. For this approach was used UAVs whit fixed-wing and multi-rotor. These platforms require sensors with high temporal resolution, weight restriction, and light-emitting diode (LED-based optical sensors).[22]

Figure 1 shows architecture of the fixed-wing UAV with a dust sensor. This telemetry dust sensor is carried by multi-rotor drone. For interpretation of air quality measurements, the meteorological data are used. The meteorological data were collected also with a UAV platform. The integration of dust sensor data with gas sensors data allows a variety of applications [22]. Other UAV architecture is shown in Figure 2.

In the following example, bicycles are used for mobile measurements. The Aeroflex is a measurement bicycle, which has been developed as a mobile platform for local air quality monitoring systems in urban environments. The Aeroflex used additional measurement devices such as GlobalSet BU-335 GPS with SiRFstarlll chipset, a Center 322 sound level meter, Microsoft LifeCam Cinema 720 p HD camera, temperature, and relative sensor.[23]

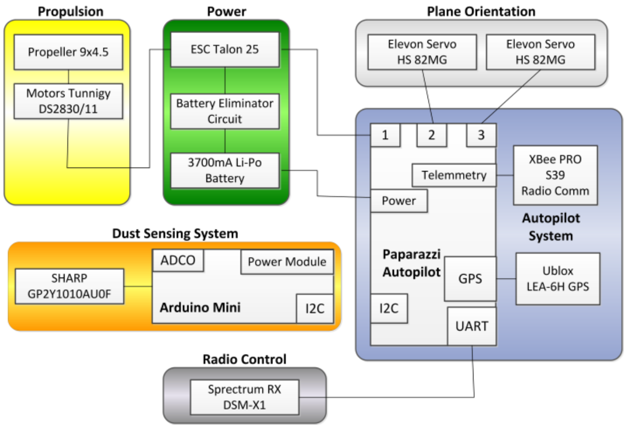


Figure 1. The systems for the fixed-wing UAV with dust sensor [22]

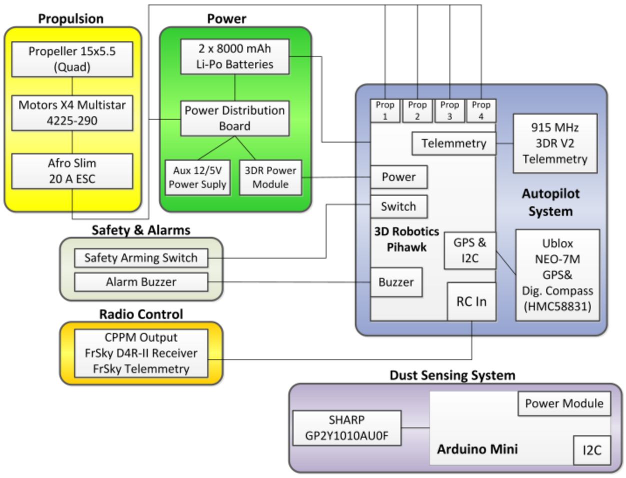


Figure 2. System architecture for quadcopter UAV with the independent gas-sensing system. [22]

IV. CONCLUSION

In the era of human population growth, followed by the growth in industry and growth in the number of vehicles, the question of air quality monitoring becomes considerably important. To keep the environment healthy as much as possible, the air quality monitoring systems rely on recent advances in the area of wireless sensor networks and wireless communication technologies.

This paper presents the selected projects of air quality monitoring systems. The presented projects are divided into two groups. One group is a group of projects with sensors deployed on fixed positions and the other group is based on mobile sensors. In the presented research the focus is given on the analyses of projects that can be used as examples of deploying the air quality monitoring systems. The results of the analyses and the experience of the presented project implementation will be used for future work in designing and planning air quality monitoring systems in smart city environments.

ACKNOWLEDGEMENT

Ministry of Education, Science and Technological Development of the Republic of Serbia, supports this research under the project ”The development of software tools for business process analysis and improvement”, project number TR32044.

REFERENCES

[1] Y. A. Aliyu, J. O. Botai, “Appraising city-scale pollution monitoring capabilities of multi-satellite datasets using portable pollutant monitors, Atmospheric Environment, Volume 179, 2018, pp 239-249, https://doi.org/10.1016/j.atmosenv.2018.02.034

[2] S.S. Lim, T. Vos, A.D. Flaxman, G. Danaei, K. Shibuya, H. Adair-Rohani, et al., A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet 380, 2224–2260, 2012. http://dx.doi.org/10.1016/S0140-6736(12)61766-8.

[3] WHO (World Health Organization), 2016. Global Urban Ambient Air Pollution Database Update, 2016. http://www.who.int/phe/health_topics/outdoorair/databases/WHO_AAP_database_May2016_v3web.xlsx?ua=1 (Accessed in September 2017).

[4] M. Brauer, M. Amann, R.T. Burnett, A. Cohen, F. Dentener, M. Ezzati, et al., Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. Environ. Sci. Technol. 46, pp 652–660, 2012.

[5] N. S. Ngo, S. V. J. Asseko, M. O. Ebanega, S. M. Allo'o Allo'o, P. Hystad, The relationship among PM2.5, traffic emissions, and socioeconomic status: Evidence from Gabon using low-cost, portable air quality monitors, Transportation Research Part D: Transport and Environment, 2018, https://doi.org/10.1016/j.trd.2018.01.029.

[6] G. R. McKercher, J. A. Salmond, J. K. Vanos, Characteristics and applications of small, portable gaseous air pollution monitors, Environmental Pollution, Volume 223, 2017, pp 102-110, ISSN https://doi.org/10.1016/j.envpol.2016.12.045

[7] J. E. Thompson, Crowd-sourced air quality studies: A review of the literature & portable sensors, Trends in Environmental Analytical Chemistry, Volume 11, 2016, pp 23-34, https://doi.org/10.1016/j.teac.2016.06.001

[8] S. Farahani, ZigBee Wireless Networks and Transceivers, Elsevier Inc., 2008.

[9] R. S. Sinha, Y. Wei, S. Hwang, A survey on LPWA technology: LoRa and NB-IoT, ICT Express, Vol. 3, Issue 1, pp 14-21, https://doi.org/10.1016/j.icte.2017.03.004, 2017.

[10] K. Mekki, E. Bajic, F. Chaxel, F. Meyer, “A comparative study of LPWAN technologies for large-scale IoT deployment”, ICT Express, 2018, https://doi.org/10.1016/j.icte.2017.12.005.

[11] R. Sanchez-Iborra, M.-D. Cano, “State of the Art in LP-WAN Solutions for Industrial IoT Services, Sensors”, Vol. 16, Issue 708; doi:10.3390/s16050708, 2016

[12] U. Raza, P. Kulkarni, M. Sooriyabandara, "Low Power Wide Area Networks: An Overview," in IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 855-873, Secondquarter 2017.doi: 10.1109/COMST.2017.2652320

[13] B. M. Wilamowski, J. D. Irwin, Industrial Communication Systems, CRC Press, 2011.

[14] D. Dobrilović, "Networking Technologies for Smart Cities: An Overview." Interdisciplinary Description of Complex Systems: INDECS 16, no. 3-A, pp 408-416 , 2018

[15] S. Li et al. “Development and Application of a Next Generation Air Sensor Network for the Hong Kong Marathon 2015 Air Quality Monitoring.” Sensors (Basel, Switzerland) vol. 16,2 211. 5 Feb. 2016, doi:10.3390/s16020211 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4801587/

[16] L. Morawska, P. K. Thai, X. Liu, A. Asumadu-Sakyi, G. Ayoko, A. Bartonova, A. Bedini, F. Chai, B. Christensen, M. Dunbabin, J. Gao, G. Hagler, R. Jayaratne, P. Kumar, A. Lau, P. Louie, M. Mazaheri, Z. Ning, N. Motta, B. Mullins, … R. Williams, Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?. Environment international, 116, pp 286–299, 2018.. https://doi.org/10.1016/j.envint.2018.04.018

[17] M. Gao, J. Cao, E. Seto, A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China, Environmental Pollution, Volume 199, 2015, pp 56-65, https://doi.org/10.1016

[18] J. I. Suárez, P. Arroyo, J. Lozano, J. L. Herrero, M. Padilla, Bluetooth gas sensing module combined with smartphones for air quality monitoring, Chemosphere, Volume 205, 2018, pp 618-626, https://doi.org/10.1016/j.chemosphere.2018.04.154.

[19] A. Koval, E. Irigoyen, Mobile Wireless System for Outdoor Air Quality Monitoring. pp. 345–354, 2017

[20] Y. Zhuang, F. Lin, E. Yoo, Eun-Hye, W. Xu. AirSense: A Portable Context-sensing Device for Personal Air Quality Monitoring, 2015. 10.1145/2757290.2757293.

[21] H. Guo, G. Dai, J. Fan, Y. Wu, F. Shen, and Y. Hu, A Mobile Sensing System for Urban PM2.5 Monitoring with Adaptive Resolution, Journal of Sensors, Vol. 2016, Article ID 7901245, 15 pages, http://dx.doi.org/10.1155/2016/7901245

[22] M. Alvarado, et al. “Towards the Development of a Low Cost Airborne Sensing System to Monitor Dust Particles after Blasting at Open-Pit Mine Sites.” Sensors (Basel, Switzerland) vol. 15,8 19667-87. 12 Aug. 2015, doi:10.3390/s150819667 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574675/

[23] E. Bart et al. “The Aeroflex: a bicycle for mobile air quality measurements.” Sensors (Basel, Switzerland) vol. 13,1 221-40. 24 Dec. 2012, doi:10.3390/s130100221 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574675/

[24] V. Angelakis, E. Tragos, H.C. Pöhls, A. Kapovits, A. Bassi, Designing, Developing, and Facilitating Smart Cities Urban Design to IoT Solutions, Springer International Publishing, Switzerland, 2017.

[25] C. Xiaojun, L. Xianpeng and X. Peng, "IOT-based air pollution monitoring and forecasting system," 2015 International Conference on Computer and Computational Sciences (ICCCS), Noida, 2015, pp. 257-260. doi: 10.1109/ICCACS.2015.7361361

[26] M. Simić, G. M. Stojanović, L. Manjakkal and K. Zaraska, "Multi-sensor system for remote environmental (air and water) quality monitoring," 2016 24th Telecommunications Forum (TELFOR), Belgrade, 2016, pp. 1-4. doi: 10.1109/TELFOR.2016.7818711

[27] N. Dam, A. Ricketts, B. Catlett and J. Henriques, "Wearable sensors for analyzing personal exposure to air pollution," 2017 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, 2017, pp. 1-4. doi: 10.1109/SIEDS.2017.7937695

[28] K. Zheng, S. Zhao, Z. Yang, X. Xiong and W. Xiang, "Design and Implementation of LPWA-Based Air Quality Monitoring System," in IEEE Access, vol. 4, pp. 3238-3245, 2016. doi: 10.1109/ACCESS.2016.2582153

[29] O. A. Postolache, J. M. Dias Pereira and P. M. B. Silva Girao, "Smart Sensors Network for Air Quality Monitoring Applications," in IEEE Transactions on Instrumentation and Measurement, vol. 58, no. 9, pp. 3253-3262, Sept. 2009.doi: 10.1109/TIM.2009.2022372

[30] M. B. Marinov Alvarado, Miguel et al. “Towards the Development of a Low Cost Airborne Sensing System to Monitor Dust Particles after Blasting at Open-Pit Mine Sites.” Sensors (Basel, Switzerland) vol. 15,8 19667-87. 12 Aug. 2015, doi:10.3390/s150819667

[31] M. Alvarado, et al. “Towards the Development of a Low Cost Airborne Sensing System to Monitor Dust Particles after Blasting at Open-Pit Mine Sites.” Sensors (Basel, Switzerland) vol. 15,8 19667-87. 12 Aug. 2015, doi:10.3390/s150819667 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574675/

[32] E. Bart et al. “The Aeroflex: a bicycle for mobile air quality measurements.” Sensors (Basel, Switzerland) vol. 13,1 221-40. 24 Dec. 2012, doi:10.3390/s130100221 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574675/

The Importance of Data Security and Corporate Social Responsibility

Mihalj Bakator, Dragica Radosav
University of Novi Sad, Technical Faculty "Mihajlo Pupin", Đure Đakovića bb, 23000, Zrenjanin, Republic of Serbia
mihalj.bakator@uns.ac.rs; dradosav@tfzr.rs

Abstract - The globalization of markets has put pressure on enterprises when it comes to achieving and maintaining a competitive position on competition-ridden markets. Globalization brought intense competitive relations, and small and medium-sized enterprises (SMEs) are becoming competitors to big corporations and vice-versa. In such conditions, customer satisfaction is an imperative. This further requires an effective customer relationship management (CRM) system within the enterprise. As modern CRM systems rely on customer data, besides the benefits, the issues of data security rises along with the rise of cyber-attacks. Here, another important concept comes into light - corporate social responsibility (CSR). Enterprises have to ethically, and morally conduct business activities. When the sensitivity of customer data is introduced, CSR becomes an imperative. Responsibly managing customer data, requires adequate data security systems. In this paper the importance of data security and CSR is analyzed. The main goal is to determine factors which affect business performance from the aspect of data security. The paper also discusses guidelines and suggestions when it comes to CRM systems, data security and CSR.

I. Introduction

Conducting business amidst the globalization of markets carries challenges and risks for enterprises who want to achieve and maintain a stable competitive position on the market. Globalization brought the necessity for the digitalization of businesses [1]. The digitalization process is not only affecting businesses and their activities, but also every other social aspect of human behavior. As digitalization brought e-commerce, collaboration software and other digital platforms and applications, small and medium-sized enterprises (SMEs) became competitors to big global corporations. Now, with the fast advancement and spread of information-communication technologies (ICTs) enterprises have to overcome the challenges and barriers, which are brought by the changes and to adapt as effectively as possible [2].

In order to achieve and maintain a relatively strong competitive position on the market, enterprises have to implement and apply modern ICTs [3]. This means not only randomly implementing a new technology or random digitalization of a department in the enterprise, but rather, a process of identifying, evaluating possibilities on the market and filling gaps in the market by introducing advanced methods of creating value for customers. Besides the challenges of globalization, enterprises have to adapt to the challenges of conducting business within the framework of fourth industrial revolution - Industry 4.0 [4]. This requires strategic implementation of adequate ICT in order to increase productivity, quality and value which is created for the enterprise and the customer as well. Enterprises have to adapt their business model for the dynamic market, and to adapt adequate ICTs in order to effectively and efficiently distribute valuable information [5].

Enterprises aim at transferring their business activities online (partially or fully). This process requires adequate consumer targeting and effective customer relationship management (CRM). Modern CRM systems rely on customer data and they are one of the cornerstones of adequate competitive advantage on the market. CRM utilizes customer and market data in order to develop user-friendly and specific metrics for customers' interests [6]. Managing a CRM systems requires strict additional protocols if customer data is involved. Namely, customer data or better say, sensitive customer data which can be targeted by malicious individuals or groups, has to be secured with an effective and efficient data security system. These data security systems are an imperative for socially responsible enterprises.

In this paper, the importance of cyber security, or more precisely, data security and the importance of corporate social responsibility (CSR) are addressed. The main goal is to identify key factors of conducting business in the digital age and to identify and outline a framework for future research in the domain of CRM and data security. The paper includes three main sections (excluding the Introduction and Conclusion sections). The first section addresses the challenges of conducting business in the globalized and digitally transformed market and the necessity for CRM and data security within CRM. Further, in the second section the role and importance of CSR regarding customer data and its security is highlighted. Finally, guidelines and suggestions for improving data security and for improving business performance of enterprises who are "battling" on the international market. Next, conclusions are drawn and guidelines for future research are proposed.

II. Conducting Business in the Global Market and CRM

The modern, globalized market presents a challenge for enterprises when it comes to achieving and maintaining an adequate competitive position. Competitive position on heavily saturated markets is important for the survival of the enterprise. Due to the constant changes on the market, enterprises have to constantly adapt from various business aspects. This may include the necessity for innovation, increased quality, or advanced CRM systems for increasing customer satisfaction. In addition, long-term flexible strategies should be put in place in order to effectively tackle "barriers" on the international market.

In such a turbulent market environment, SMEs are becoming competitors to big global enterprises which results in intense competitive relations. From here, even the concept of competitive ability is changed, and long-term competitive positions with status-quo strategies are not achievable. Domestic enterprises have to improve product quality, productivity, implement modern ICTs and to practice modern management tools and techniques [7].

Besides the pressure from the changes which are brought by globalized markets, enterprises have to adapt to the framework of the fourth industrial revolution - Industry 4.0. This revolution is characterized by advanced ICTs and the integration of manufacturing processes and other sector in the enterprise [8]. This further implies that enterprises have to focus on the customer and to satisfy their needs. For an improved customer experience, enterprises have to apply an effective CRM system. Through such CRM systems, the enterprise possesses the ability to adequately manage the relations with its customers. Modern CRM systems heavily rely on customer and market data. Their goal is to improve customer satisfaction, in order to improve the enterprise's position on the market [9].

CRM systems integrate various management, marketing and quality functions as they are based primarily on ICT. CRM systems collect, analyze and process customer data and stores it in specifically allocated databases [10]. The data which is stored may contain less sensitive information such as anonymous customer reviews, competition market share and market trends. However, besides these, there are sensitive information such customers' contact details or even bank account details which if targeted by malicious individuals or groups can severely harm the enterprise's reputation and even end it as customer data leaks may lead to law suits which destroy the enterprise.

CRM systems have to serve for obtaining an objective overview of the market and trends regarding customer behavior and satisfaction. The main goal systems should be increasing customer satisfaction which further positively affects business performance. In the next section CRM and customer data importance is further discussed and analyzed.

In order for a CRM system to be effective amidst the globalized markets, customer data has to be collected, analyzed and stored. Based on the analyzed data, strategic and operational business actions should be taken. SMEs retain customer data as reference points for future data analyses. The main goal of such customer data is customer retaining and developing a competitive edge and improved financial performance. [11]

As modern ICTs made it possible for almost any enterprise to process, and analyze data from customers, there is an increased risk of mishandling large batches of this valuable data and the number cyber attacks increases every year. The potential benefits often outweigh the risks, as an effective CRM system is a necessity for effective managing customer relations on the market. Enterprises can apply CRM systems along with the gathered and analyzed customer and market data for developing short-term and long-term strategies which would attract and retain customers. This further implies, that modern ICT-based CRM systems are an important business factor from the aspect of sustainability and sustainable development in the domain of customer loyalty, brand loyalty, and other business metrics. [12].

Some of the information that a CRM system collects includes survey data on customer satisfaction, reviews from various websites, perceived product and service quality, brand loyalty, brand trust, customer retention rate, revenue per customer, and other business and market performance metrics. This data is grouped, and analyzed with the goal to obtain information, which is necessary for conducting adequate actions for improving customer retention rates, and customer satisfaction levels. CRM practices certainly have a positive effect on the relationship quality between customers and enterprises. If good relations are maintained over a longer period of time, there is a higher chance for customer loyalty development. Here, good relations are mainly based on creating and communicating value to the customer. [13].

III. CUSTOMER DATA SECURITY AND CORPORATE SOCIAL RESPONSIBILITY

Each year, there is an increase of cyber attacks aimed at enterprises and sensitive customer data. Therefore, enterprises have to develop effective data security systems for securing not only customer data, but also market data, financial data, and employee data as well. Constant optimization and backups are an imperative for such data security systems. Further, data security systems have to integrate prevention procedures and protocols which include storage of cyber-attack records (specifically designed databases); reporting system for potentially malicious e-mails (phishing, malware, Trojan attachment or other suspicious attachments); database aggregation and thorough analysis (data integrity checks, data backup checks); setting up filters for inbound and outbound connections and network access; and finally, make a public list of malicious e-mails [14]. Enterprises have to implement prevention techniques, detection software and to define protocols for cyber attack response. Some of the technological techniques and tools may include firewalls (for inbound and outbound connections); intrusion detection systems; antivirus software on all levels of the IT infrastructure; and finally, an effective protocol for responding to incidents is advised. [15].

Modern data security systems consist of several main components. These are internal and external risk analysis; internal and external defense from attacks; detection of internal and external threats and risks; revival, restoration and recovering of affected data by the attacks; and evaluation and monitoring of employees, data routes, network connections, business activities and protocols; conducting timely backups, updates and overall developing awareness among employees about the necessity for secure databases. [16].

Overall, in order to safely manage customer and other data, enterprises have to address five main security factors. These factors are confidentiality, availability, authentication, integrity and non-repudiation. These are essential for a secure network which is the cornerstone of a data security system and a modern CRM system.

Now, why is data security an imperative? First, database breaches can lead to lawsuits from affected customers. Second, it can completely "obliterate" the reputation of the enterprise. Therefore, CRM systems and data security systems are complementary with the concept of CSR. Namely, CSR represents the integration of voluntary activities of an enterprise with the goal to enhance the narrower or wider community. Some of the activities of CSR include the promotion of social goals, marketing that is connected to social goals, corporate social marketing, voluntary community work, and social responsibility as a business practice [17].

It is evident that data security systems are not directly integrated into the main approach of CSR. However, CSR has several aspects: economic (positive outcomes for shareholders, employees, and the overall community); social aspect (responsibility towards customers, quality products, employee skills improvement, employee motivation and responsibility towards the community); and ecological aspect (the enterprise has to conduct business in a way which won't have direct or indirect negative effects on the environment) [17]. Beside these aspects, CSR involves the top management of enterprises and their decisions which should be in accordance with ethical and moral standards. It was also noted that the main goal of CSR is to prevent the potential negative effects of conducting business and to maximize the positive aspects of conducting business of an enterprise [18].

From here, it is evident that data security systems and secure handling of customer data is within the CSR domain. More precisely, it is the social aspect of CSR and the prevention of harm domain which has to be taken seriously in order to successfully run a business long-term. In the modern business environment, CRM systems are necessary for improving customer satisfaction. Data security systems have to be in place to secure data and comply with some of the aspects of CSR.

This way, even if the customer knows that its data is stored, through data security systems and effective CSR where this information is communicated to the customer, customer trust will annul the "bad feeling" of leaving personal information online for an enterprise. Knowing that their data is safe, customers will comply; build trust and loyalty over time. Therefore, an effective CRM system surely involves the CSR approach as well. Data security is the connective "link" between these two concepts and represents a crucial element of building trust and loyalty with customers.

IV. GUIDELINES AND PROPOSITIONS

Based on the analyzed literature in the domain of CRM systems, CSR and data security importance, the following guidelines for and propositions for improving these systems are proposed:

- Enterprises have to apply modern ICTs in their business activities in order to effectively "tackle" the challenges of the fourth industrial revolution [19].
- CRM systems have to safely manage customer data, and to communicate with the customers that their data is safely stored [20].
- Data security systems are an imperative regardless of the type and amount of data is stored.
- Enterprises can opt for database providers where maintenance costs and security risks are generally lower. However, for highly secretive enterprise, it more adequate to use in-house databases.
- Corporate social responsibility has to integrate modern ICTs in their domain of action. A large part of conducting business and communication with customer is realized online. Therefore, corporate social responsibility has to be practiced globally, and not only locally [21].
- Through CRM and CSR systems, as well data security systems, the main goal is to provide value and trust to the customer. This way the chances of customer retention are higher and customer satisfaction is improved [22].

Overall, in the modern business environment, enterprises have to adapt to changes brought by the globalization of markets. As customer data became the ultimate tool for modern CRM systems, new challenges arise as enterprises have to secure this data from malicious attacks. The security of this data should be part of the enterprise's CSR. Responsibly conducting business is an imperative for achieving a long-term competitive position on the international market.

V. CONCLUSION

Continuously satisfying customers is the to loyalty development. As customers are often aware that their data is collected and stored by the enterprise, it is necessary to implement adequate data security systems, and also to communicate data security to the customer, in order to build customer trust. In this paper, CRM, CSR and data security were interconnected through the modern approach to conducting business where customer and market data is collected, analyzed, evaluated and stored by enterprises. It can be concluded that amidst the globalization of markets and within the frameworks of Industry 4.0, enterprises, and especially domestic enterprises which are in a transitional environment, have to implement modern ICTs within various sectors. In this paper, the importance of CRM systems, CSR and data security were reviewed.

The main limitation of this paper is the lack of surveyed data on the use of data security systems in domestic enterprises. However, this paper is an adequate starting point for further research in this domain. For future research, domestic enterprise managers could be interviewed on the application of modern ICTs and the competitive potential that it brings. This current paper contributes to the existing body of literature in the form of a concise overview and in the form of a future research "checkpoint".

REFERENCES

[1] M. Bakator, D. Đorđević, and D. Ćoćkalo, "Developing a model for improving business and competitiveness of domestic enterprises," Jou. of Eng. Manag. and Comp. (JEMC), vol. 9, no. 2, pp. 87-96, 2019. UDC: 658.562(497.11) 339.13(497.11)

[2] D. Ćoćkalo, D. Đorđević, S. Bogetić, M. Bakator, and C. Bešić, "Competitiveness of Domestic Enterprises in Changing Markets and Industry 4.0," In L. Monostori, V. D. Majstorovic, S. J. Hu & D. Djurdjanovic (Eds.), Proceedings of the 4th International Conference on the Industry 4.0, 5-7th June, 2019, Springer Nature, Switzerland, pp. 113-127, 2019. ISSN 2195-4356, doi: 10.1007/978-3-030-18180-2_9

[3] M. Pomffyová, and L. Bartková, "Take Advantage of Information Systems to Increase Competitiveness in SMEs," Procedia - Social and Beh. Sci., no. 220, pp. 346–354, 2016. doi:10.1016/j.sbspro.2016.05.508

[4] L. Ardito, A. M. Petruzzelli, U. Panniello, and A. C. Garavelli, "Towards Industry 4.0," Bus. Process Manag. Jou., 2018. doi:10.1108/bpmj-04-2017-0088

[5] M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh, "Customer relationship management and big data enabled: Personalization & customization of services," App. Comp. and Inform., 2018. doi:10.1016/j.aci.2018.05.004

[6] M. Holmlund, Y. Van Vaerenbergh, R. Ciuchita, A. Ravald, P. Sarantopoulos, F. V. Ordenes, and M. Zaki, "Customer experience management in the age of big data analytics: A strategic framework," Jou. of Buss Res., 2020. doi:10.1016/j.jbusres.2020.01.022

[7] S. Vukotic, T. Milivojevic, and N. Zakic, "Theory, Practice and Characteristics of Management in Serbia," Int. Jou. of Adv. in Manag. and Econ., vol. 2, no. 2, pp. 63-75., 2018.

[8] M. Rüßmann, M. Lorenz, P. Gerbert, M. Waldner, J. Justus, P. Engel, and M. Harnisch, "Industry 4.0: The future of productivity and growth in manufacturing industries," Boston Consult. Gr., vol. 9, no. 1, pp. 54-89, 2015.

[9] P., Kotler, and G. Armstrong, "Principles of Marketing (17 ed.)," Pearson Education: Essex, England. ISBN-13: 978-0134492513. pp. 38-41., 2017.

[10] S. Pinnington, and M. Jones, "Performance Measurement for Customer Relationship Management (CRM)," Proceedings of the 2007 Academy of Marketing Science (AMS) Annual Conference, pp. 112–117., 2007.

[11] W. C. Li, M. Nirei, and K. Yamana, "Value of data: there's no such thing as a free lunch in the digital economy," US Bureau of Economic Analysis Working Paper, Washington, DC, 2019. (Available on: https://www.rieti.go.jp/jp/publications/dp/19e022.pdf)

[12] M. Vesal, V. Siahtiri, and A. O'Cass, "Strengthening B2B brands by signalling environmental sustainability and managing customer relationships," Indus. Mark. Manag., 2020. doi:10.1016/j.indmarman.2020.02.024

[13] S. Das, and M. Mishra, "The Impact of Customer Relationship Management (CRM) Practices on Customer Satisfaction," Bus. Gov. and Soc., pp. 43–54, 2018. doi:10.1007/978-3-319-94613-9_4

[14] S. Kabanda, M. Tanner, and C. Kent, "Exploring SME cybersecurity practices in developing countries," Jou. of Org. Comp. and Elect. Comm., vol. 28, no. 3, pp. 269–282, 2018. doi:10.1080/10919392.2018.1484598

[15] S. Venkatraman, and R. Venkatraman, "Big data security challenges and strategies," Aims mathematics, vol. 4, no. 3, pp. 860-879, 2019.

[16] S. F. Aboelfotoh, and N. A. Hikal, "A Review of Cyber-security Measuring and Assessment Methods for Modern Enterprises," JOIV: Int. Jou. on Inform. Visual., vol. 3, no. 2, pp. 157-176, 2019. doi:10.30630/joiv.3.2.239

[17] R. Piasecki, and J. Gudowski, "Corporate Social Responsibility: the Challenges and Constraints," Comp. Econ. Res., vol. 21, no. 4, pp. 143-157, 2017. doi: 10.1515/cer-2017-0032.

[18] C. Bešić, and D. Đorđević, "Korporativna društvena odgovornost," Fakultet tehničkih nauka: Čačak, p. 82, 2015.. ISBN: 978-86-7776-178-3.

[19] A. Schumacher, T. Nemeth, and W. Sihn, "Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises" Procedia CIRP, vol. 79, pp. 409–414, 2019 doi:10.1016/j.procir.2019.02.110

[20] A. Idzikowski, P. Kuryło, J. Cyganiuk, and M. Ryczko, "Customer Relationship Management (CRM) - Philosophy and its Significance for the Enterprise" Sys. Safe.: Hum. - Tech. Fac. - Environ., vol. 1, no. 1, pp. 1004–1011, 2019. doi:10.2478/czoto-2019-0127

[21] C. Chedrawi, and S. Osta, S. "ICT AND CSR in the Lebanese banking sector, towards a regain of stakeholders' trust: the case of Bank Audi" Int. Jou. of Serv. and Stand., vol. 12, no. 2, pp. 205-219, 2018. 10.1504/IJSS.2018.091848

[22] I.M. O'Brien, R. Ouschan, W. Jarvis, and G.N. Soutar, "Drivers and relationship benefits of customer willingness to engage in CSR initiatives", Jou. of Ser. Theo, and Prac., vol. 30 no. 1, pp. 5-29, 2020 https://doi.org/10.1108/JSTP-08-2018-01

Two Recommender Systems: Technical Decisions and Lesson Learned

Evgeny Cherkashin[1,2], Viktoria Kopylova[2], Boris Shevchenko[2], Nikita Lukyanov[2]
[1] *Laboratory of Complex informational systems*,
Institute for system dynamics and control theory, SB RAS, Irkutsk, Russia
eugeneai@irnok.net
[2] *Institute for informational technologies and data analysis*,
National research Irkutsk state technical university, Irkutsk, Russia
kopylovika@mail.ru, lukyanovnd@istu.edu

***Abstract*—The research of the paper is devoted to expose experience gained in development recommender systems for two domain areas: real estate of a region and choosing the specialty of a high school by entrants. These systems implement different approaches to obtain data and the technique of figuring out the user interest. The common in the techniques are the usage of preliminary data analysis for user and object classification that reduces the computations and data stored.**

Technique of overcoming “cold start” problems in these domains are proposed. A general system design and implementation are considered. The future development is indicated.

I. Introduction

Recommender systems (RS), being considered as an example of decision support systems, are useful to support users with additional information and decision variants. In this paper we consider our experience with developing RS in two domains: real estate of a region and choosing the specialty of a high school. Both systems were developed within master degree at Institute for informational technologies and data analysis, National research Irkutsk state technical university. That’s why they have status of pilot projects showing the basic technique implementations.

The first system deals with the problem of selling real estate: a flat, a room in a flat, a house. Most of the objects on the market have similar characteristics in a class, the user interest is expressed with obvious characteristics like price, location, level, number of owners. But the most interesting aspect of the problem is the special cases of realty like private family hotels, countryside houses, shops. These are usually being sold for a long time, and the real estate firms utilize too many efforts for that. The time spent to find new owner of the property is critical for the business: sellers have losses due paying the taxes for the estate being not utilized, buyers have no possibilities to invest available capital. RS focusing user interest on the previously filtered subsets of the property, preliminary sorting by relevance, could speed up the sales.

Each school student sooner or later faces the problem of choice of a future specialty, university to enroll, and the set of courses to study. Decisions made at this time significantly affect their future life, their job and career perspectives, families’ life level. Therefore, any help at this stage seems to be a necessary. In comparison to the real estate, where there is a whole industry supporting decision-making – the services of realtors –, in this field the educational environment supports only an information provision service: students are given only career guidance, advertisement for the high school capabilities. The student is on his/her own in the fundamental life decision. Of course, parents and favorite mentors can supply the additional life experience examples helping to focus the analysis of the present conditions, but the actuality of their advices are generally of questions due to specifics of their experience, subjectivity, absence of the topical information of the present state of the labor market, scientific progress, *etc*.

In a period of intense competition, multidisciplinary educational institutions with similar areas of study are interested in a tool that allows them to determine the target audience more accurately, as well as entrants are interested to find a relevant university. The work with the audience would be focused and directed. This will help to attract applicants interested in a particular direction, which is being implemented in the educational institution. This tool could improve the effectiveness of career guidance activities for high school students. Recommender systems are the tools of such kind.

Recommender systems (RS) [1] are decision support information systems designed to assess the user’s level of interest in a particular product or service (object) based on available information about user and object. The RS development industry began to actively develop with the emergence of online sales services, and now it is one of the active areas of development of decision support systems, a direction of artificial intelligence, focused primarily on commercial use, as well as on solving problems of increasing the productivity of searching for relevant information.

Our experience gained when producing a prototype RS for the above mentioned domains is expressed as already built and published at Github.com [2] RS, and in the second RS, being now on the stage of data source import routines implementation.

II. Data Source for RS

A. Real estate RS

In Russian sector of Internet, real estate market is presented with various *classified advertisement* websites for selling goods, jobs positions, real estate, cars, services, like `Avito.ru`. In Irkutsk region and Moscow, every second flat is sold via such services, with publishing a number of offers for the same flat. This usually results in reduction of the price to the level below the average on the market. In Ekaterinburg, real estate is generally sold by realtor firms, in this case the process is more predictable for the seller. The number of deals in Irkutsk region was raising by 20% from 2010 to 2016 on the secondary housing market. This increase is due to the raising of the activity of the websites.

The source of the real estate data is `atlcom.ru` that provide an actual database for real estate market of almost any region or city of Russia. One can register and supply filter setup by a form for selection of the part of the database to download. The downloaded data are represented in XML format. The format is developed by `Yandex.com` within the R&D of describing selling commodities. A special variant is for real estate.

The file comprises a list of `offer` records (subtrees) describing a flat, house or a room. The description include metadata (where published, sale agent), location address, price, area in square meters, a photo, number of rooms, floor level, type of building and its general characteristics, as well as the conditions for obtaining additional information. Each record is identified by a global identifier.

The RS users (agents) have roles: unknown, seller, buyer, owner, realtor, expert, and invalid user. Before registering users' have role "unknown" and as each object in the system is assigned an GUID (global 128 bit identifier) related to the web browser with cookies. With this cookie and GUID user is being tracked: the RS server collects data on his/her interest in acquiring offers identifiers being visited on the RS website. An offer is graded as positive if user has spent more than predefined seconds viewing its webpage. According to the review [3] forcing user to explicitly grade an offer is useless.

Users having role "expert" are for construction of the taxonomy of realty objects, naming of the obtained taxons (classes of realty), and storing the results in the database. Expert's service is needed at the stage of development of the general principles of RS functioning at the cold start, the main problem we are to solve within the R&D. Realtor role is connected with data load activities (select a XML file, import it in the database, control the results). In Fig. 1 we showed the Use case diagram of the system functions.

Therefore, the constructed RS was to be based on *collaborative filtration* technique, where there is no expert data available for the object of user interest. To support our case, note, that the set of offers is dynamically changed day-by-day, and it is very hard to supply the data with expert decisions reflecting all user classes opinions. Having expert constructed the taxonomy, which seems to be made once, the new "flats" can be classified easily by measuring the distances to the representative instances of the classes. The taxonomy allows us to consider the realty class as the object of user interest, a set of similar realty is a property characterizing the class.

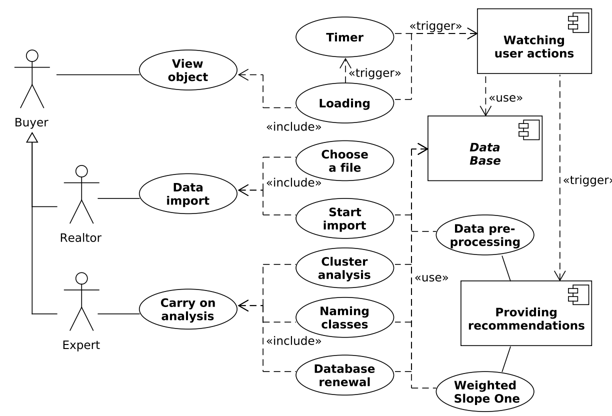


Fig. 1. Use case diagram for real estate RS

B. Entrants' RS

Functioning of an RS could be based on analysis of entrants' personal properties, matching them with already enrolled students, analyzing existing individual direction of training and even a particular educational program. We have been developing and implementing a prototype RS for helping to choose a direction of training at a university for entrants on the base of algorithms for predicting their possible choice based on digital footprints in a social network. As a source data, *e.g.*, subscriptions of Russian social network "In Contact" (IC) (in Russian – "В контакте") are used as sources of personal characteristics, and enrollment orders of our university.

Nowadays social networks are one of the main places of person's life, and even more so for a teenager. This is clear from the data from the Glas Runet online poll service. Among the 2000 participants of the survey, the majority (86%) of them know about the existence of Internet social networks and use their services. Among those who know about the existence of social networks, only 10% do not use them [4]. The most popular social networks according to the Romir holding are Odnoklassniki, VKontakte and Instagram. The youngest audience, from 14 to 30 years old, visits VKontakte, 45% does it daily, and 70% does more often than once a day [5].

Often, it is in social networks that one can track users' self-expression, interests and attitude to the world. The analysis of groups, posts on the main blog, reposts, statuses, *etc.* can be the sources of the valuable information [6].

Of course, if a person is an adult, then it is unlikely that all subscriptions to groups will be relevant ("entered and forgot"). But school students, according to research [7], update subscriptions every six months or a year: irrelevant content in the feed annoys them. Accordingly, subscriptions can become the original dataset that can be analyzed.

In 2016, Tomsk State University researchers proved that the VKontakte profile reflects the cognitive personality traits, *i.e.*, the educational interests of schoolchildren are closely related to their behavior in the social network [8].

C. Basis of entrants' RS

There are three basic approaches to development of an RS: *content-based filtering*, *collaborative filtering* and implementing as a *knowledge-based system*. Collaborative filtering is efficiently used when there is no initial data on object features describing user interests to the items, as well as definitive set of items features, which used in representing user's interest aspects. This approach is based on continuous data accumulation of user's behavior (the decision made) within a system (*e.g.*, internet shop). In our case we cannot collect user's decision for each individual user continuously as user makes only one decision on faculty to study at.

This can be done for user classes; each enrollment entrants are classified and their decisions update their classes collected data. As this is done rarely, usually once a year, and can be processed off-line, then content-based filtering is more suitable approach for this function mode. The third approach could be usable later after acquiring more data for data mining.

For the ease of further consideration of the RS basic aspects of development, we took "Content-based filtering" section of Wikipedia webpage (`https://en.wikipedia.org/wiki/Recommender_system`) and filled it with our data.

Content-based filtering methods are based on a description of the item (course in our case) and a profile of the user's (student and entrant in our case) preferences [9], [10]. These methods are best suited to situations where *there is known data on an item* (name, location, description, *etc.*), but *not on the user*. Content-based RS treat recommendation as a *user-specific classification problem* and learn a classifier for the user's preferences (likes and dislikes) based on an item's features.

The item's source data are documents, enrollment orders, where all the formal data presented for each enrolled student. In contradiction, we have all data about user's features from their IC profiles. Thus, we have to design machine learning structure relating entrant's features to the classes of courses of a university.

In this system, *keywords* are used to *describe the items* and a *user profile* is built to indicate the *type of item* this user likes. In other words, these algorithms try to recommend items that are similar to those that a *user liked in the past*, or is examining in the present. It does not rely on a user sign-in mechanism to generate this often temporary profile. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research.

In order to obtain the keywords describing courses, we are to create a taxonomy of courses and relate it to the known characteristics, for example, competence names or keywords obtained from course descriptions. User profile is built out of the analysis of the user's profile and subscriptions at IC. We cannot obtain information on the previous behavior for the individual entrants, but we can consider as a user a class of similar users, *i.e.*, join users in classes and treat a class as a RS user. New entrants are classified and their IC profiles represent new data on user's past behavior.

To create a user profile, the system mostly focuses on two types of information:

1) A *model of the user's preference*.
2) A history of the user's interaction with the recommender system.

Model of user's preference implemented as relation of user class profile features to a study course features expressed by keywords.

Basically, these methods use an item profile (*i.e.*, a set of discrete attributes and features) characterizing the item within the system. To abstract the features of the items in the system, an item presentation algorithm is applied. A widely used algorithm is the tf–idf representation (also called vector space representation) [11]. The system creates a content-based profile of users based on a weighted vector of item features. The weights denote the importance of each feature to the user and can be computed from individually rated content vectors using a variety of techniques. Simple approaches use the average values of the rated item vector while other sophisticated methods use machine learning techniques such as Bayesian Classifiers, cluster analysis, decision trees, and artificial neural networks in order to estimate the probability that the user is going to like the item [12].

We decided to use neural networks for the estimation, as nowadays there appeared a bunch of good programming libraries, which are easily applied.

In general, the research consists of our variant of the standard scenario that includes the following steps:

1) Import and analyze enrollment orders existing at the university administration;
2) Crawl the social network for personal characteristics and subscriptions for all student included in the enrollment orders;
3) Import curriculum data for all courses we interested in, gathering related data such as competence;
4) Creating taxonomies for students and study courses, name resulting classes, juxtapose them with related data;
5) Devising classification procedures for relating entrants to the classes;
6) Importing and carrying on data analysis, resulting organizing a rational database having wrapped in an ORM (Object-relational mapping) to ease of usage;
7) Implementation the content-based filtering for students that are entrants with data in the database.
8) Create web application for entrants and university staff.
9) Testing the system and interviewing students.

There are regular problems of feature selection describing users and study courses, as well as naming their classes after stage of taxonomy inference. The subscription group (channel) names can be used as features too. They also could be processed with clustering to generalize subsets of groups, giving additional data on the similarities and control correspondence to their set of subject tags. As for names of courses classes competence names can be used also possibly after a generalization. Classification of new students is to be implemented with two-layer neural network for easy implementation. The data will be processed in context of the university and its subdivision departments, institutions. There is other data that could be used in the preliminary data analysis: faculty name, course relation data, competence definition text, relation of competences to their courses, course data like study direction and other describing features, and, in principle, results of feed text analysis.

III. Technologies Used

In both projects we used C#-based platform [13] for implementation as it is a compromise of expressiveness and computational performance, it has rich set of libraries and frameworks. The RS are web ASP.NET and MONO applications with database represented with Entity Framework Object-relational mappers (ORM) [14]. One of the advantages of the C# language is partial classes allowing composition of a class dividing its definition between files, which can be both generated and programmed manually. This technology supports great flexibility of Entity Framework. The language also supports properties and decorators, making it a convenient application programming language.

Database back-end of the realty RS is open-source RDF (Resource Description Framework) triple store `BrightstarDB`, allowing in future research to use data in logical inference engines, *e.g.*, for constructing knowledge-based RS and connect the system with other servers via SPARQL. The second RS back-end database is SQLEXPRESS, the default setup for Microsoft Visual Development Studio.

The input data for entrants' RS is provided by the input forms and automatic processing of students enrollment orders in DOCX format, using our syntactic analyzer. Orders contain header, which is parsed to determine the date of the enrollment, specialty and other parameters. The only table part lists the enrolled students. Additionally, the list records are checked by regular expressions determining the sum of points gained. The reading of the document data is implemented by means of C# library `Microsoft.Office.Interop.Word` with its `Paragraphs` collection and `Range` objects.

The list of students is traversed, and for each student his/her IC profile is queried. The data of interest are the profile and the subscriptions. The query is done via IC API as HTTP queries [15]. According to IC API policy, the queries can be issued not more than three times per second, otherwise error 6 meaning “Too many requests per second” is obtained. Additional constraints are stated to the set of issued remote methods. The policy for the set is not published, its violation results to the Captcha test issue or rejection of a concrete method usage with no restriction for others.

One student can have more than one IC accounts. In this case a form appeared to choose the right one manually. The decision is stored in a database table attribute.

TABLE I
Comparing features and their weight coefficients

Attribute	Calculation technique	v_k, %	Formula
`string Name`	not used		
`ILocation Location`	equality	10	(1)
`string Address`	–”–	10	
`float Price`	relative difference	25	(2)
`CurrencyEnum …`	not used		
`float Area`	relative difference	35	(2)
`AreaUnits AreaUnit`	not used		
`string ImageURL`	–”–		
`string URL`	–”–		
`int Rooms`	relative difference	100	(2)
`int RoomsOffered`	–”–	100	(2)
`int Floor`	–”–	30	(2)
`int FloorTotal`	–”–	10	(2)
`BuildingEnum …`	equality	30	(1)
`IBuilding …`	–”–	30	(1)
`PropertyEnum …`	–”–	30	(1)
`CategoryEnum …`	–”–	100	(1)
`string Description`	not used		
`string GUID`	–”–		

IV. Preliminary Data Processing

In realty RS, two program agents are implemented, which are executed on an event occurrence. The first one starts by timer of web browser and implements issuing user positive evaluation, if user spends some time viewing a page of a realty object. The second one, located at the server side, is activated if the first one raised event of a positive evaluation. Server agent receives the evaluation and may start recalculation of estimates for recommendations if there are enough computational resources.

All data are stored in a database, which structure is represented as objects in Fig. 2. The central entity is `IOffer`, representing the offer of a realty object. The information on an object, which is independent of an offer, is stored as `IObject`-object. The object structure corresponds to standard Yandex and Google form of real estate representation.

The real estate domain specifics is the total availability of data for realty objects, so the import function provides the whole body of information on objects by importing it from other sites.

A. Classification of realty objects

Comparison of the object is done in a vector space according to the features presented in Table I.

In the table, with (1) and (2) we denote the following formulas:

$$d_k(i,j) = \begin{cases} 0, & \text{if } a_i = a_j, \\ 1, & \text{if } a_i \neq a_j, \end{cases} \tag{1}$$

$$d_k(i,j) = \frac{|a_i^k - a_j^k|}{a_i^k + a_j^k}, \quad a_i^k + a_j^k > 0. \tag{2}$$

The aggregate assessment is made with formula:

$$d(i,j) = \frac{\sum_{k=1}^{m} |v_k \cdot d_k(i,j)|}{\sum_{k=1}^{m} v_k}, \qquad 0 \leqslant d_{i,j} \leqslant 1,$$

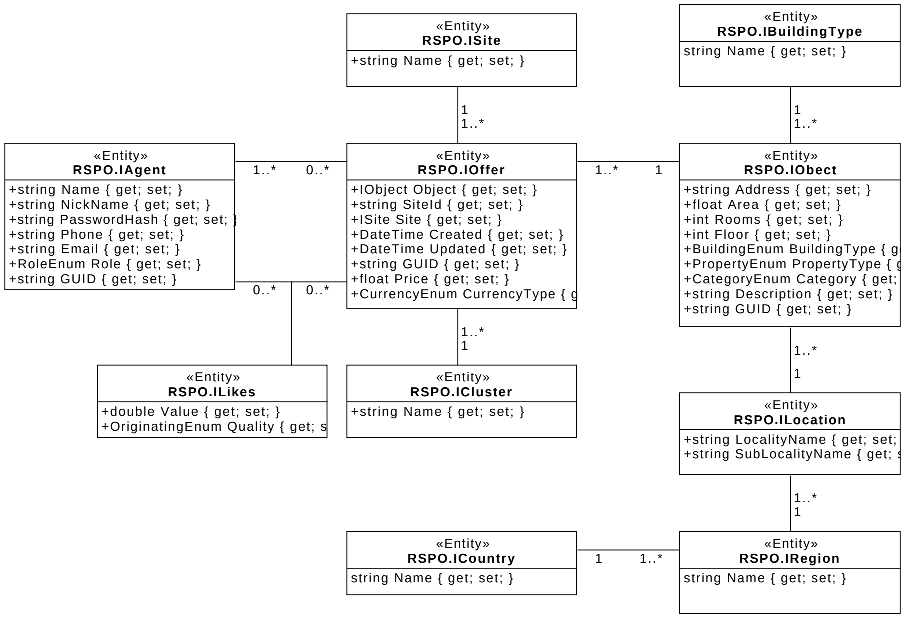


Fig. 2. Class diagram for real estate RS

where v_k is a weight of k-th attribute value in the table.

In order to create a taxonomy, expert user must run the hierarchical clustering function supplying the number N of general clusters to figure out. After the function finishes, expert will see the numbers of top N clusters with number of their objects. Next step is naming the clusters. Expert enters the subsets and reviews the objects, forming an image of the common properties of the objects. After naming the image in the first interface, expert changes name of cluster from a number to the name. This is to be done for each $\overline{0, N-1}$ cluster. If expert cannot form a concrete image of the class, them we recommend repeat the procedure with larger number of clusters.

Some clusters can gain the same name, and after renaming, clusters with the same name are joined. Hierarchical clustering allows us to make experiments with data without recalculation of the taxonomies. Before running clustering function, expert can restrict the set of objects to be examined, *e.g.* by 200 objects, so we can use even agglomerative hierarchical clustering.

The following classes are presented on Irkutsk region real estate market: two-room flats, one-or-two-room flats (comfort class realty), one-room flats, dachas, houses, rooms, commercial space, a garage, and elite realty (flats with three and more rooms).

User interests are acquired as follows. New user is suggested a set of objects from all the classes of realty. Watching users activities, which objects he/she is viewing, the systems gains initial interest, and then specifies the class of objects to view at following steps. Each object observed by user during long time considered to be of interest, the positive evaluation is added to database. Recommended objects are shown in a list located at the bottom of the interface window. The set of suggested objects is restricted by 20 items.

V. Recommendations Generation

The RS for real estate market is implemented Slope One based collaborative filtration [16]. We assume that buyer purchases a flat not frequently and then each buyer is a new user, and no data about his/her preferences are available, as well as one flat cannot be bought several times as compared to a regular internet shop of commodities. New users are assigned a GUID and tracked by cookies. Tracking data forms preference data. After a predefined number of viewed pages, RS suggests registering, it is an engine preventing losing the data due to moving user to another workstation, and starts to issue recommendations. The user is not required to register, as it was specially stated before us as a requirement, but we explain the advantages of the registration action.

RS tracks user movement around various classes and fixed for each class the estimation of the object explicitly viewed by user. For each visited class, RS performs filtering the set of objects and range them according to all collected user data in the context of the class. The relevance is calculated by Slope One technique. The average values of differences of objects' evaluations are calculated by request and cashed in the user session data. After calculation of the evaluations for each object, which was not seen by user, in the class, objects are sorted by relevance and a predefined number of them are displayed.

The Slope One algorithm can generate recommendations for an object (a class) if there is at least one positive evaluation of an object of the class. At the cold start period if there were no such evaluations, RS shows random 20 objects of the same class of objects user is viewing.

VI. Web Applications

Both RS use MVC [17], [18] architecture in construction of user interface. Realty RS is based on `Nancy` ASP.NET framework, where for each URL template a lambda function is related, having supported parameters of the Request objects. At the beginning of each function we restore `session` object, and at the end, the Response object is created by applying a context model and a view object to a template. As template engine SharpTAL is used, an implementation of Zope page template (ZPT) with Template attribute language (TAL). This implementation follows traditions of C#, namely, compiling its content into objects at the phase of project compilation.

As for each web template engine, ZPT uses `model`, `view`, which implements *controller* functions as well, and `template` to generate HTML content. Model is referred as `context` of a view, which function is simplification of the access to context by template, and transforming user actions into model change upon `POST` query. For using global application and other data, as context we pass in a general case `application`, `appview`, `user` and `message` objects. User defines the currently authenticated user, and message contains text to show in a view, announcing an event happening in the RS, *e.g.* the fact that a record is successfully stored in the database.

ZPT and TAL technologies were chosen as they give a very powerful tool for HTML tree structure manipulation including usage of named and parametrized macros, attributes subtrees, defined in one-page template, and multiply used in other ones, as well as the technology tries not to extend XHTML syntax with new entities for implementation of element replacement functions.

VII. Related Works

There are many research groups and commercial firms dealing with developing RS technologies for various domains. We will consider related domains and other interesting cases.

In [19] a case-based reasoning RS for real estate market is presented. Its goal is to find a similar instance in the database of existing cases, describing the cases for the user. The obtained relations are used in followed collaborative filtration and content filtering. In theoretical research [20] the users are divided on sellers and buyers. Sellers “advertise” their realty by highlighting the important features by their opinion. So the sellers play expert roles for forming data for content filtration. Research [21] showed that usage of Internet does not influence significantly on the time efficiency, flexibility and satisfaction criteria of the search for a flat to buy. There are no statistical difference between 2009 and 2011, since in 2011 88% of buyers used Internet just as the main source of information. Authors created RS with case based reasoning utilizing an ontology model.

In [21] authors highlight three key features which significantly affects the decision: location, housing unit properties and price. Final decision is made after evaluation of the environment, such as distances to shops, kindergartens, schools. This can be accounted reducing the additional data to the value of the feature of housing unit properties using ontology. The environment constraint is entered by user: user draws a circle around the preferable location, which should contain interested services and realty.

Paper [22] considers a complex logistic problem of organization of a process of real estate control, involving various people group in changing business environment. The main aim is to develop a model, where the groups will be maximally satisfied in a rational micro and macro environment. The efficiency of the realty utilization is evaluated with criteria of market, ownership, renewal prices, capacity, number of operations to be done while ownership transferring, safety, comfort, the time of physical and technical exploitations, *etc*. The software functions shift from “the most economically efficient control of the property” to multicriterial choice, raising computation efficiency.

For recent RS dealing with student’s problem of choice the following papers were observed. In [23] an RS is constructed for helping entrants to choose an education direction (expressed as sets of corresponding high schools) according to his/her grades in GCE (General certificate of education), gained competing prizes, physical activity, and hobby. Three methods of content filtration were implemented: by distances between vector characteristics, axiomatic method of Pareto-set contraction, and analysis of hierarchies. Authors figured out coefficients relating study directions to the features of user using series of experimental assessments controlled by experts.

In [24], a problem of learning outcomes assessment of higher education students is considered. A course RS on the base students’ graduate attributes, which describe their developing values, is proposed. RS rates improvement after each course and suggests new courses by a collaborative filtering algorithm in order to improve student’s average competence profile. Students are presented as long vectors of courses they already taken with the corresponding grades. Similarity is calculated with a variation of angle cosine of the vectors. The recommendations are the prediction of the grades for the new courses. Similar by the problem statement paper [25], where RS predicts course learning trajectory patterns, the students is suggested elective curses. Source data for collaborative filtering are existing learning trajectories acquired from senior students in the image and likeliness one student advices younger one. Students and courses are divided on clusters. For each student cluster (a group), mean values are calculated. RS algorithms are based on combination of collaborative filtering and fuzzy-like rule based system, which defines process of production of a decision. The article has a good reviews of related works and three classical approaches basics.

Work [26] has the similar aim to develop RS for advising students new courses, but is focusing on description of courses, taking advantage of natural language processing over course documents to acquire descriptions. Feature descriptions contain formal course data (name, structure, lecturer) and a placement in a keyword appearance frequency space. Existing student data, represented as grades of the courses he/she already got, related to the course data. The resulting recommendations are represented as top-5 new courses, where student will gain the best grades. This is a simplistic direct approach and, by authors opinion, gives good results; some measurement supporting the opinion were carried on.

Interesting research has been done in [27], where the problem of overcoming the cold start problem in recommending genres and music compositions, as well as movies. The authors developed the RS "EZSurf" automating the process of web surfing and content filtering using a user profile on the social network "VKontakte", as well as API services `last.fm` and TheMovieDB, to obtain information about similarity of the objects. This approach greatly simplifies storage of RS data, since it does not require the designing own system of classifications. In [28], a problem of choosing top–N objects, with higher evaluations of user interest, in content filtration. Authors proposed RS model based on fuzzy sets, RS quality assessment technique, and a corresponding algorithm. The models have been tested on `last.fm` data.

In a comprehensive RS review dealing with text (scientific) document relevance grading [3], authors noted that more than half (55%, 34 of 62) RS were based on content filtration. Collaborative filtration was used only in 18% (11 of 62) cases. Stereotype based and hybrid methods are also presented. Authors concluded that 81% in of cases modeling user does not produce significant results in comparison to explicit indication the set of keywords. The papers were described by keywords and, less frequently, N–grams included in the text, as well as metadata such as authors and references. Most popular recommending model is based on vector space approach. User interest modeling is implemented with graphs, where vertices are the papers, and arcs are their relationships, and topic lists assigned to user by machine learning. The topics are organized in hierarchical catalogs using ACM classifiers. In the RS which used collaborative filtration, the explicit ratings was not collected at all as users were too lazy to supply a rating to a paper he/she looked through. Implicit ratings were obtained by measuring the number of pages user read, document user interaction (loading, editing), co–loadings, co-view and co-citing by users of one group.

The collaborative filtration is currently more popular as it reflects practical experience: most of RS R&D have to solve "cold start" problem of initial lack of information on objects and users, as well as have adaptation to user group interest transformations due to modern trends.

VIII. Discussion

The realty RS was tested by a number of human being fictitious buyers, none of them mentioned misbehavior of the system like spontaneous transitions from one class to another, or supplying empty sets of recommendations.

The overview of the presented experience shows that C# .NET and MONO technologies and used techniques are well suitable for construction RS of various kind, fulfilling the following typical requirements:

1) working in "cold start" conditions;
2) do not require user to register and authorize while implicit gathering information needed for recommendation generation;
3) testing systems are to be done in a local conditions to be able to check the results comparing with existing "natural" data;
4) using open-source technologies, modules and libraries.

The obtained real estate RS implementation as well as the second project do not take advantage of the nowadays methods of R&D development, which corresponds to current traditions. Most attention is paid to preliminary data processing and obtaining minimal valuable product (MVP).

IX. Conclusion

We presented results of two master degree projects in the field of recommender system (RS) development in real estate and university entrants' study direction decision-making. The first one is finished, and the second one is on a half way. Both systems have similar design, but rely on different RS technique: the realty RS is based on collaborative filtration, the entrants' one is based on content filtering. Both systems use object-oriented representation of application entities, which are persistent as well.

Solutions of "cold start" problem for both domains has been considered. They are based on creating taxonomies of objects by cluster analysis and partial transition to consider classes as objects of users' interests. In the entrant RS we also created taxonomies of users according their subscriptions in a social network.

The RS are implemented as MVC web applications for users, the administrative and expert user interfaces are implemented as window applications in entrant RS. This simplifies user role model: web users must register, administrator runs form application at workstation and is not required to register. Testing the finalized RS showed stable work when used by human users.

The described technique could be developed further by adaptation the standard directions such as:

- development or adaptation of ontology of domain to extend search capabilities especially in students' courses domain;
- case based reasoning for the same domain;
- predictive modeling of attribute values for realty;
- providing geospacial data of infrastructural service organizations in the environment, *e.g.* schools, kindergartens and shops.

Acknowledgements

Authors thank Evgeny Vasil'evich Brezheshov, director of a realtor office, for helping with problem domain description and problem statement for real estate RS.

References

[1] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich. Recommender Systems: An Introduction. Cambridge University Press, 2010.

[2] E. Charkashin, B. Shevchenko. Source code of a prototype real estate market recommender system. URL:https://github.com/eugeneai/RSPO-CSharp. (access date: 21-09-2020)

[3] J. Beel, B. Gripp, S. Langer, C. Breitinger. Research-paper recommender systems: a literature survey. International Journal on Digital Libraries (2016) 17: 305. DOI:10.1007/s00799-015-0156-0

[4] S.P. Balandina, M.V. Bautin, A.V. Maiyorov, E.R. Smirnova. Social networks as means of education and educational participants interactions. Pedagogicheskie i informacionniye tehnologii v obraxovanii. No. 15 (2016). (in Russian) URL:https://journals.susu.ru/pit-edu/issue/view/31

[5] K.S. Veber, A.A. Pimenova. Comparative analysis of social networks. Vestnik Tambovskogo universiteta. Series: Natural and technical sciences. Vol. 19, No. 2, 2014. p. 636–643. (in Russian)

[6] A.V. Rozhkova. Self-expression (self-presentation) in social Internet networks as a phenomenon of human cyber-socialization. Homo Cyberus. 2017. No. 2(3). (in Russian) URL:http://journal.homocyberus.ru/Samovyrazhenie_v_socialnyh_internet-setjah_kak_fenomen_kibersocializacii

[7] A.V. Fescshenko. Social networks in education: experience analysis and development perspectives. Gumanitarnaya informatika, Tomsk, Issue 6. 2012. p. 124–134. (in Russian)

[8] V.V. Matsuta, P.B. Kiselyev, A.V. Fescshenko, V.L. Goyko. Investigation of social networks potential for revealing talented students. Psychologiya i psihotehnika. No. 4. 2017. p. 104–121. (in Russian)

[9] Ch.C. Aggarwal Recommender Systems: The Textbook. Springer. 2016. ISBN 9783319296579.

[10] P. Brusilovsky. The Adaptive Web. 2007. p. 325. ISBN 978-3-540-72078-2.

[11] D.H. Wang, Y.C. Liang, D.Xu, X.Y. Feng, R.C. Guan. ”A content-based recommender system for computer science publications”, Knowledge-Based Systems. 2018. 157: 1-9

[12] S. Blanda. Online Recommender Systems – How Does a Website Know What I Want?. American Mathematical Society. 2015.

[13] J. Albahari, B. Albahari. C# 6.0 in a Nutshell: The Definitive Reference 6th Edition, 2015. 1138 p. ISBN-13: 978-1491927069

[14] Entity Framework Tutorial. URL:https://www.entityframeworktutorial.net/ (access date: 05.07.2020).

[15] VKontakte API documentation. URL: https://vk.com/dev/manuals (access date: 05.07.2020).

[23] E. A. Belotskiy, A. V. Suetin, Construction of a recommendersystem for choosing higher education institutions for entrants, Vestnik S.-Petersburg Univ. Ser. 10. Prikl. Mat. Inform. Prots.Upr., 2016, Issue 1, 66–77. URL:http://www.mathnet.ru/links/1c01f91d57c8dc6f050c7e6d3fed3604/vspui277.pdf (in Russian)

[16] D. Lemire, A. Maclachlan. Slope One Predictors for Online Rating-Based Collaborative. URL: http://cogprints.org/4031/1/lemiremaclachlan_sdm05.pdf (access date: 10.05.2018)

[17] A. Freeman. Pro ASP.NET MVC 5. 5th ed. Apress. 2013. 832 p. DOI:10.1007/978-1-4302-6530-6

[18] J. Chadwick, T. Snyder, H. Panda. Programming ASP.NET MVC 4: Developing Real-World Web Applications with ASP.NET MVC. O'Reilly Media; 1st Edition. 2012. 492 p. ISBN-13: 978-1449320317

[19] E. M. Alrawhani, H. Basirona, Z. Sa'ayaa. Real estate recommender system using case-based reasoning approach. Journal of Telecommunication, Electronic and Computer Engineering (JTEC). Vol. 8 No. 2. p. 177-182.

[20] E. V. Britina. Segmenting recommender system with client–server connection based on programmable configured networks and usage protocol with fast-jumping IP-address. // Sovremennye problemy nauki i obrazovaniya. No 6. 2015. URL:https://www.science-education.ru/ru/article/view?id=16875 (in Russian)

[21] X. Yuan, J.-H. Lee, S.-J. Kim, Y.-H. Kim. Toward a user-oriented recommendation system for real estate websites. Information Systems, 38 (2013). p. 231–243.

[22] T. Ginevičius, A. Kaklauskas, P. Kazokaitis, J. Alchimovienė. Recommender system for real estate management. Verslas: Teorija ir praktika (Business: Theory and Practice). 2011 12(3). p. 258–267 DOI:10.3846/btp.2011.26

[24] B. Bakhshinategh, G. Spanakis, O. Zaiane, S. ElAtia. A Course Recommender System based on Graduating Attributes. In Proceedings of the 9th International Conference on Computer Supported Education, Volume 1: CSEDU, ISBN 978-989-758-239-4, 2017, pp. 347-354. DOI:10.5220/0006318803470354

[25] A. Al-Badarenah, J. Alsakran. An Automated Recommender System for Course Selection” International Journal of Advanced Computer Science and Applications(ijacsa), 7(3), 2016. DOI:10.14569/IJACSA.2016.070323

[26] J. Naren, M. Z. Banu, S. Lohavani. Recommendation System for Students'Course Selection. A. K. Somani et al. (eds.), Smart Systems and IoT: Innovationsin Computing, Smart Innovation, Systems and Technologies 141. 2020. pp. 825–833. DOI:10.1007/978-981-13-8406-6_77

[27] B. R. Avhadeev, L. I. Voronova, E. P. Ohapkina. Development of recommender system on the base of profile in social network “VKontakte”. Vestnik Nizhnevartovskogo gosudarstvennogo universiteta. Issue No 3. 2014. (in Russian)

[28] A. S. Amel'kin, D. M. Ponizovkin. Mathematical model of top-N problem for content recommender systems. Izvestiya MGTU “MAMI” No 3(17), 2013, Vol.2. p. 26-31. (in Russian)

Improving the Design of Convolution Neural Network Architecture: Enhancement Methods, Applications and Challenges

Zainab Fouad*, Marco Alfonse*, Mohamed Roushdy** and Abdel-Badeeh M. Salem*
* Computer Science Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt
** Faculty of Computers and Information Technology, Future University in Egypt, New Cairo, Egypt
zainab.mfouad@cis.asu.edu.eg; marco@cis.asu.edu.eg; Mohamed.Roushdy@fue.edu.eg; abSalem@cis.asu.edu.eg

Abstract - Recently, Convolution Neural Network (CNN) has accomplished great success in numerous issues of machine learning. Many machine learning methods have been developed for such objective, for example: Artificial Neural Network (ANN), logistic regression, Support Vector Machine (SVM), deep learning, etc. Deep learning (specifically CNN) is one of the strategies by which we are able to delude the challenges of the feature extraction process. Usually deep learning models are capable of extracting the proper features by themselves. Also, deep CNN models are usually designed manually and the key parameters of it are decided by experience and repeated tests which incredibly limit the applications of deep CNN. Therefore, it is a great challenge to design the proper deep CNN model and reduce the dependence on manual involvement and expertise. So this review will focus on improving the convolution neural network design from different aspects with various methods, including how to automatically design CNN model without operator intervention, change on the convolution or pooling layers, adding some features to save the computational resources and how to use adaptive CNN parameters, which change according to the characteristics of the input signal. The recognition accuracy reaches 98.97% in some represented systems.

I. Introduction

Many researches have taken place in the area of pattern recognition over the last years. The field of pattern recognition is concerned with the use of computer algorithms to automatically discover the regularities in data and use these regularities to take action, for example, classifying the data into several categories [1].Many techniques of machine learning have been developed for this objective. For example: logistic regression, artificial neural network, deep learning, support vector machine, etc. Deep learning and especially CNN are getting to be increasingly well known in tackling different applications of computer vision that deal with how computers can be made to choose up deep understanding from videos or pictures. It looks to automate tasks that the human visual system can do [2]. One of the huge challenges with conventional machine learning models could be a process called feature extraction. This stage is exceptionally critical and influences the in general execution of the learning framework. But usually feature engineering is a manual stage in the machine learning system and it is taking a long time. The software engineer must tell the computer what kind of things it ought to be searching for that will be instructive in making a decision. Feeding the algorithm crude information rarely ever works, so feature extraction may be a critical portion of the conventional machine learning system. This place a tremendous burden on the software engineer and the algorithm's effectiveness depends intensely on how insightful the software engineer is. For complex issues such as object recognition or handwriting recognition, this can be a tremendous challenge. On the other hand, Deep learning is one of the only strategies by which we will delude the challenges of feature extraction. This is often, since deep learning models are able of learning to focus on the proper features by themselves, requiring small guidance from the software engineer. This makes deep learning an amazingly effective tool for advanced machine learning [3]. Deep, CNN models are ordinary designed manually. And usually the key parameters of it are decided by repeated tests and experience which incredibly constrain the application of deep CNN. So, this review will examine how to improve the convolutional neural networks from different aspects to design a proper model. The organization of this review is as follows: Firstly, the CNN architecture and applications are presented in the section II. Then, the methods that automatically design of the CNN model in the section III. Then model the details of the methods that improve the CNN model are documented in the section IV. Then, conclusion and future work are outlined in the section V.

II. CNN Architecture and Applications

The blocks of CNN are the convolutional layers, pooling layers, and fully connected layer [4]. These blocks are shown in Fig. 1.

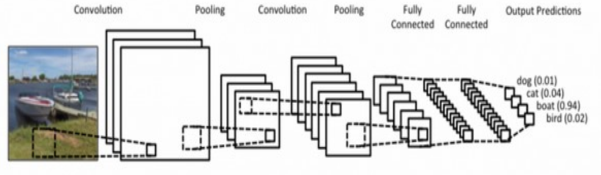


Figure 1. The building blocks of CNNs [5]

The basic block of the CNN is the convolutional layer. The layer's parameters have a set of filters (to detect the features) with random size. In the forward pass, each filter is convolved over the width and height of the image, the filter slides with a given step horizontally, then moves with another step vertically, until the full image has been checked, Fig. 2 [5].

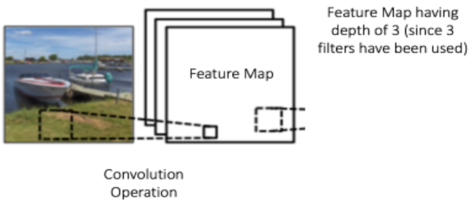


Figure 2. The convolution layer [5]

More deep features can recognize by adding more convolution steps. For example, in the task of classification, CNN detects edges from pixels in the 1st stage, then detect simple shapes in the 2nd stage, and then detect higher-level features, as shown in Fig. 3.

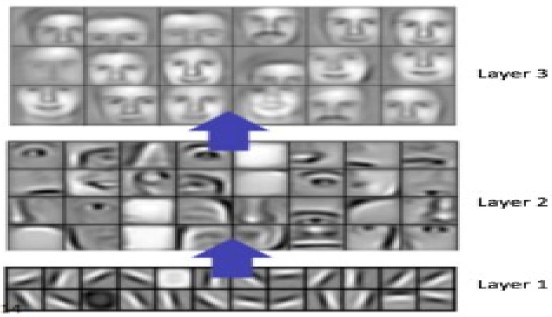


Figure 3. The more convolution layers will get more complex features [5].

Decreasing the dimensionality of the feature map and keeping the critical data are the task of the pooling layer as shown in Fig. 4.

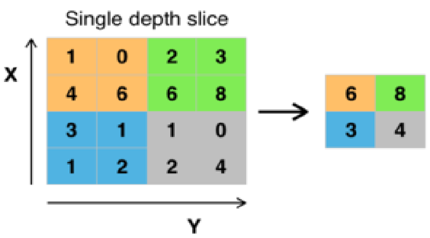


Figure 4. The Pooling Layer [5].

This layer has a filter component (called the kernel), the yield of a kernel is the (mean or max.) value of the region. If the maximum value is chosen, it could be a max pooling layer. Instead of taking the maximum element we could also take the average (average pooling) [5]. After that fully connected layer can be any kind of classifiers such as multi-layer perceptron.

The CNN model can be applied in various applications including image classification, object tracking, pose estimation, text detection, action recognition, scene labeling, speech and natural language processing.

III. Automatic Design of CNN Model

This section will introduce some methods that design the CNN architecture automatically.

In [6] Yanan Sun et al, proposed an automatic architecture design method for CNNs by using Genetic Algorithms (GA), which is able of finding a promising design of a CNN on dealing with image classification tasks. The proposed algorithm does not require any preprocessing before it works, nor any post- processing on the CNN, which implies that it is totally automatic. This objective has been effectively accomplished by using encoding methodology for the genetic algorithm to encode different depths of CNN, joining the skip connections to advance more deep CNNs and creating a parallel and a cache component to essentially quicken the fitness evaluation with restricted computational resource [7]. The results demonstrate that the algorithm accomplishes the best accuracy of classification among other CNNs models. Also, in [8] Haiman Tian et al, proposed to choose the most excellent CNN and domain based on GA. A different genetic encoding strategy is displayed that shows distinctive pre-trained models. Amid the evolutionary stage, the ideal genetic code that displays the most excellent model is chosen, or new competitive individuals are delivered utilizing the genetic operations.

In [9] Masanori Suganuma et. al, try to build CNN architectures automatically based on Cartesian Genetic Programming (CGP) [10,11]. The CNN structure and connectivity represented by the cartesian genetic programming encoding method are optimized to maximize the validation accuracy. Instead of designing CNN architecture with trial and error. This method finds automatically competitive CNN design by the experimental results. Fig. 5 illustrates an overview of automatic design of CNN model.

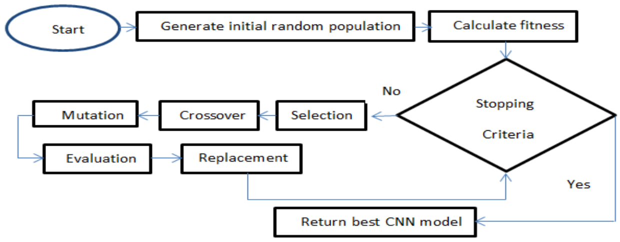


Figure 5. Automatic design of CNN model.

The summarization of different methods of automatic design of CNN model is presented in Table I.

TABLE I. AUTOMATIC DESIGN OF CNN MODEL

Ref.	Main contribution	Application	Dataset	Evaluation
Yanan Sun et al, 2019, [6].	Automatic architecture design method for CNNs using genetic algorithm.	Image classification.	CIFAR10 CIFAR100	Classification accuracy 95.22% 77.97%
Haiman Tian et al, 2018, [8].	Select the best CNN pretrained model for each dataset and domain based on genetic algorithm.	Image classification.	Disaster Network Camera 10K CIFAR-10	Precision 0.38 0.7 0.651
Masanori Suganuma et. al, 2017, [9].	Automatically construct CNN architectures based on CGP	Image classification.	CIFAR10	Error rate CGP-CNN (ConvSet) 5.80% CGP-CNN (ResSet) 5.66%

The methods mentioned in the table I have accomplished competitive automatic CNN design for the image classification task, but the method based on CGP requires high computational cost.

IV. CNN Enhancement Methods

There are several researches that address the improvement of the CNN model with various methods. This section will introduce these methods.

In [12] Lu Zhang et. al, presented a Scale adaptive CNN (SaCNN) model with a fixed receptive field for crowd counting. They extricate feature maps from numerous layers and adjust them to yield the same size of the output and combine them to create the density map. The count of individuals is calculated by joining the density map. Moreover, they present a relative number loss together with the density map loss to enhance the generalization of the network on swarm scenes with a few individuals, where a lot of models work ineffectively on. The proposed strategy can effectively adjust to individuals of distinctive scales and points of view. The tests were done on the UCF CC 50, Shang-haiTech and WorldExpo datasets, that they collect form swarm scenes with few individuals. They illustrate that the scale-adaptive CNN make an enhancement in the state of art [13].

In [14] Yuanyuan Zhang1 et. al, proposed an Adaptive Convolutional Neural Network (ACNN) for face recognition, whose structure can be decided by expanding it automatically concurring to the requirements, so it is simpler and saves the training time. Since there is no hypothesis around the network development of CNN (number of layers, number of feature maps, and so on), researchers develop a lot of CNN structures on their experience and performance comparison. Other researchers focus on hardware acceleration, that speed up the performance comparison [15,16]. The ACNN algorithm is as the following:

Step 1 Network Initialization: Just one branch in the initial network.

The structure of 1st branch incorporates 2 convolutional and 2 pooling layers with only one feature map at each layer and using back propagation to update the weights. Then decide if the initial network is convergent or not with (1).

$$\mathrm{err\ (initial)} - \mathrm{err\ (present)} \geq T. \qquad (1)\ [14].$$

where err (initial) is system average error of the former training, err (present) is average error of the current training, and T is the threshold value of convergence rate, which is set to 0.1. The system average error (err) is given in (2).

$$\mathrm{err} = 0.5\ \sum_{j=1}^{a} \sum_{i=1}^{m} (y - y_{\mathrm{label}})^2 \,/\, a \quad . \qquad (2)\ [14].$$

(a: number of training samples, m: number of output units).

In case (1) is fulfilled, the initial network is converging, and it is trained till meets the requirements. Else, the global expansion is required.

Step 2 Global Expansion: The global expansion of a new branch based on the previous one.

Before expansion, QA (output form branch A) is saved, and QA is combined with the output of branch B. The output of branch B is defined in (3).

$$y = f\,(QA + QB) \quad . \qquad (3)[14].$$

The weights branch B only are updated by BP. Then, also check if this branch is convergent or not. If it is not convergent, so another global branch is required. Else, training this branch till the average error of the system reaches the target.

Step 3 Local Network Expansion.

We require local expansion in case that the recognition rate of training samples does not reach the target.

A Fusion of local and global branches is given by (4).

$$C1local = f\,(S1A * kA + S1B * kB). \qquad (4)\ [14].$$

(C1local feature map at 1st convolutional layer of local network), (S1A, S1B: feature maps of branch A and B),

(kA, kB: convolutional kernels).

Qglobal at the output layer of the global network is saved and combined the local branch. The output of local network is given by (5) and (6).

$$Qglobal = QA + QB. \qquad (5)\ [14].$$

$$y = f\,(Qglobal + Qlocal). \qquad (6)[14].$$

If the global network as one branch there is no fusion done.

Incremental learning is needed in real applications to save memory of previous training samples, Fig. 6.

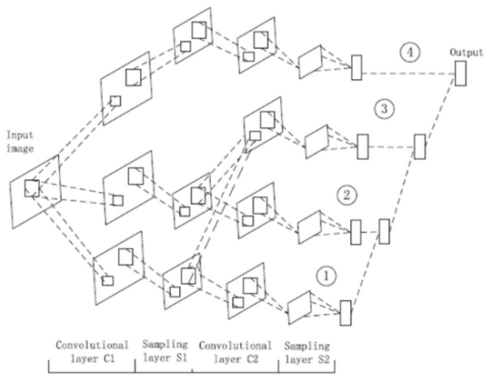


Figure 6. The model of incremental learning [14].

The output of whole network is defined in (7).

$$y = f\,(Q1 + Q2 + Q3 + Q4). \qquad (7)\ [14].$$

The experiments based on ORL face recognition show that ACNN finds the nearly ideal architecture of CNN.

In [17] Min Lin et. al attempt to improve the model discriminability for local patches within the receptive field in a structure called "Network in Network" (NIN). Fig. 7 shows the overall NIN structure.

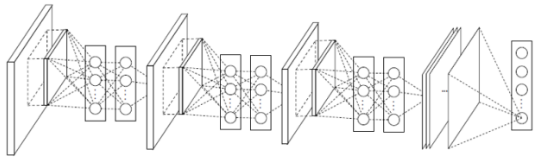
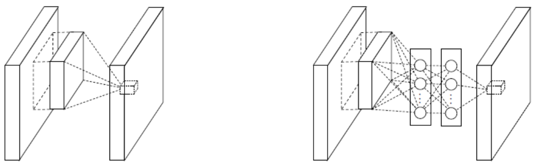


Figure 7. The NIN model [17].

Their structure replaces the linear filters in the convolutional layer with Mlpconv that use a multilayer perceptron to convolve the input and replace the fully connected layers with a global average pooling layers [18]. They model the local patches better with Mlpconv layers, and global average pooling acts as a basic regularizer that avoids overfitting. Fig. 8 outlines the distinction between straight convolutional layer and mlpconv layer.



(a) Linear convolution layer (b) Mlpconv layer

Figure 8. The distinction between straight convolutional layer and mlpconv layer [17].

They show that the NIN model proves a state-of-art performance on CIFAR-10, CIFAR-100 and SVHN datasets.

In [19] Bert De Brabandere et. al proposed that the traditional convolutional layer, the learned filters stay fixed after training. In contrast, so they introduce the Dynamic Filter Network, that is dynamically produce filters based on the input as shown in Fig. 9. They prove that this structure is a powerful one, without an excessive increase in the number of model parameters. They demonstrate the effectiveness of the dynamic filter network on the tasks of video and stereo prediction and show improvements in performance compared with other models.

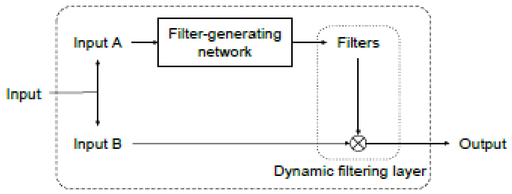


Figure 9. The dynamic filter module [19].

There is tow component of the dynamic filter module: first network to generate a filter based on the input, the second dynamic filter layer that use the generated filter to apply it on the input. The two inputs of the module can be either identical or different, depending on the task.

In [20] Dingjun Yu et. proposed a mixed pooling strategy that replaces the max, sum, average, etc. pooling (that has some downsides Fig. 10) with the combination of max pooling and average pooling.

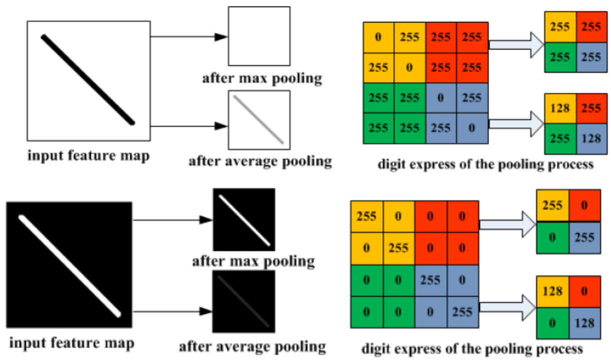


Figure 10. Example of downside of max and average pooling [20].

The mixed pooling function can be formulated in (8)

$$y_{i,j,k} = \lambda \max_{(m,n)\in R_{ij}} a_{m,n,k} + (1-\lambda)\,(1/|R_{ij}|) \sum_{(m,n)\in R_{ij}} a_{m,n,k} \quad (8)\ [20].$$

where λ is a random value 0 or 1 (the choice of using average or max pooling). The λ is recorded in the forward process and will be used for the backpropagation operation. The overfitting problem encountered by CNN is solved by this method. An experiment was done on image classification datasets show that the proposed method is better than conventional pooling.

In [21] Xudong Li et. al proposed an Adaptive Deep CNN for Scene Specific Object Detection (ADCNN). Numerous previous works show generic detectors whose performance will drop quickly when they are applied to a surveillance [22]. In their work, they proposed an efficient method to construct a scene-specific regression model based on a generic CNN based classifier. Their model can predict object locations within the surveillance scene. First, they exchange the generic CNN-based classifier on the surveillance scene by selecting valuable kernels. Second, they learn the context information about the surveillance scene in their regression model for accurate location prediction. The schematic view of their model is shown in Fig 11.

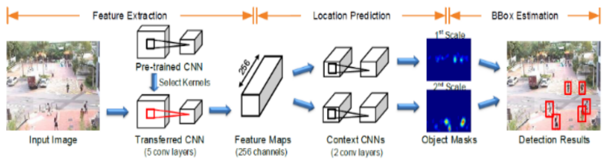


Figure 11. The ADCNN model [21].

This model has two parts: The 1st one is initialized by the pretrained CNN, that extract discriminative features. They transfer this pre-trained CNN to several target domains for improving the scene adaptability and call it Transferred CNN (TCNN). The 2nd part contains S CNNs to produce S object masks to predict object locations for a certain object scale. They introduce the context information into these CNNs to improve the precision of locating objects and call them Context CNNs (C-CNNs). They compare the ADCNN model with some state of art models and found that ADCNN achieves the best performance on surveillance datasets.

In [23] Pierre Sermanet and Yann LeCun proposed multi-scale convolutional networks for traffic sign recognition. Features are learned in ConvNets in each level from data that are tuned to the task at hand. In conventional ConvNets, the output of the final level is fed to a classifier. In their work the outputs of all the stages are fed to the classifier, Fig. 12.

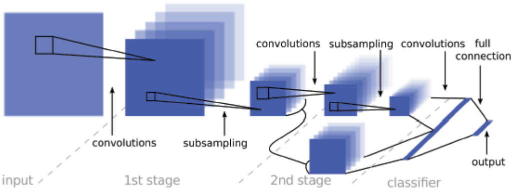


Figure 12. The Multi-Scale Convolutional Networks Architecture [23].

This permits the classier to use low level features instead of using deep features only. The system yielded an accuracy of 98.97%.

In [24] Max Jaderberg et. al proposed spatial transformer network that has a Spatial Transformer part, which allows the network to actively spatially transform feature maps conditioned on themselves without explicit supervision. A spatial transformation characterizes a geometric relationship between each point within the input and output images. Convolutional Neural Networks characterize an exceptionally effective class of models, but are still restricted by the need of ability to be spatially invariant to the input data. The transformation is then performed on the whole feature map (non-locally) and can include rotations, cropping and scaling. Their work introduces a spatial transformer module, that can be included into a standard CNN architecture to provide spatial transformation capabilities. The spatial transformer part can effectively spatially transform an image by creating a suitable transformation for each input sample, Fig. 13.

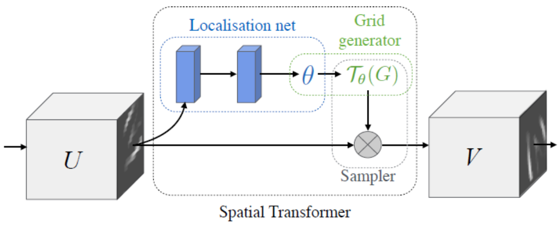


Figure 13. The Spatial transformer module [24].

The feature map is taken in the 1st network and generate the spatial transformation. Then use this transformation to form a sampling grid. At last, the feature map and the sampling grid are taken as inputs to the sampler, creating the output map sampled from the combination of the localisation network, grid generator, and sampler form a spatial transformer. This is a self-contained module which can be dropped into a CNN design, giving rise to spatial transformer networks. They show that the CNN model that is uses the spatial transformers resulting in a state of the art for several classes of transformations and performance on several benchmarks.

In [25] Xueyun Chen et. al proposed Hybrid Deep CNN (HDNN) for vehicle detection to extract multi-scale features. The deep convolutional neural networks extract only the features of the same scale, and hence is insufficient to tolerate large-scale variance of the object. They present a hybrid Deep Convolutional Neural Networks by decomposing the maps of the last convolutional layer into multiple blocks of max pooling receptive field sizes or variable field sizes, to extract variable scale features. Experiments on vehicle databases show that HDNN outperforms DNN.

In [26] Qiuyu Zhu and Ruixin Zhang proposed a Highly Efficient Convolutional Neural Networks (HENet) optimized for speed, storage and accuracy in order to improve the real-time performance of convolutional neural networks. The new design uses an unusual way to combine group convolution and channel shuffle which was mentioned in ShuffleNet.. Inspired by ResNet and DenseNet, they also propose a new way to use element-wise addition and concatenation connection with each block. In order to make better use of feature maps, pooling operations are removed from HENet. The utilization rate of feature map can increase with HENet. Also, it can effectively reduce the amount of calculations. The experiments show that their model's efficiency is more than 1 times higher than ShuffleNet on many open source datasets, such as CIFAR-10/100 and SVHN.

The summarization of all methods that make an improvement in the convolutional neural network architecture in Table II. Mentioned on it the CNN enhancement aspect, the application that the authors of the work applied in, the dataset they used and the performance of their application.

TABLE II. CNN ENHANCEMENT METHODS

Ref.	**Main contribution**	**Application**	**Dataset**	**Evaluation**
Lu Zhang et. al, 2017, [12].	(SaCNN) architecture. It concatenates multiple feature maps of different scales to produce a strong model.	Crowd counting.	UCF CC 50, Shang-haiTech WorldExpo and SmartCity	MAE* MSE* 314.9 424.8 16.2 25.8 8.5 - 8.6 11.6
Yuanyuan Zhang1 et. al, 2015, [14].	ACNN whose structure can be determined by automatic expansion according to performance requirements.	Face recognition.	ORL	Best recognition rate with Local + global expansion + incremental ACNN) = 0.9500
Min Lin et. al, 2014, [17].	Develop deep network structure called “Network in Network” to enhance model discriminability for local patches within the receptive	Image classification.	CIFAR10 CIFAR100 SVHN MNIST	Error rate 8.81% 35.68% 2.35% 0.47%

	field.			
Bert De Brabandere et. al, 2016, [19].	Introduce a new framework, the Dynamic Filter Network, where filters are generated dynamically conditioned on an input.	Video and stereo prediction.	Highway Driving dataset	Video prediction: Euclidean loss of 13.54 Stereo prediction: loss of 0.52
Dingjun Yu et. al, 2014, [20].	Mixed pooling method for convolution neural network to solve the overfitting problem.	Image classification.	CIFAR10 CIFAR100 SVHN	Classification accuracy 89.20% 61.93% 96.90%
Xudong Li et. al, 2017, [21].	Utilize a generic CNN-based classifier by Adaptive Deep Convolutional Neural Networks.	Predicting object locations in the surveillance scene (pedestrian detection and vehicle detection).	CUHK MIT PETS MIT	Detection rate for pedestrian detection 69.2% 66.8% 75.8% Detection rate for vehicle detection 64.4%
Pierre Sermanet and Yann LeCun, 2011, [23].	Multi-Scale Convolutional Networks that allows the classifier to use, not just high-level features, but also pooled low level features, which tend to be more local, less invariant, and more accurately encode local motifs.	Traffic sign recognition.	GTSRB	recognition accuracy of 98.97%
Max Jaderberg et. al, 2016, [24].	Spatial transformer networks that provide CNN with spatial transformation capabilities.	Number recognition and bird classification.	Street View House Numbers CUB-200-2011	Recognition error of 3.6 (64px) and 3.9 (128px) Classification accuracy of 84.1%
Xueyun Chen et. al, 2014, [25].	Hybrid Deep Convolutional Neural Networks (HDNN) to Extract multiscale features.	Vehicle detection in satellite images.	the City of San Francisco	Precision rate = 0.98
Qiuyu Zhu and Ruixin Zhang, 2018, [26].	(HENet) optimized for speed, storage, and accuracy in order to improve the real-time performance of convolutional neural networks.	Image classification and object recognition.	CIFAR-10/100 and SVHN	The speed has been increased nearly twice times than ShuffleNet

*MAE: Mean Absolute Error.

*MSE: Mean Square Error.

All methods mentioned have accomplished an enhancement form many aspects including change on convolution or pooling layers, adding some features to save the computational resources and how to use adaptive CNN parameters, which change according to the characteristics of the input signal. Also, these methods applied to various applications.

V. Conclusion and Future Work

Deep convolution neural networks have made great progress in a lot of tasks (especially in vision tasks) such as video, image, text, and speech processing. Many researches have been carried out to improve the CNN's architecture and achieved state of the art results on several tasks. In this review, we have discussed the improvements of convolution neural networks in different aspects including optimization of parameters, layer design, pooling layer, automation of the design, extract invariance features and fast computations. In spite of the fact that CNNs have accomplished incredible success in experimental assessments, there are still a lot of challenges issues that deserve advanced investigation. First, the CNNs depend on the amount of data you provide to them (If you provide them with less data, expect the CNNs to perform poorly). And as manually collecting labelled dataset requires tremendous amounts of human efforts, it is desired to investigate unsupervised learning of CNNs. Second, although there are some methods to optimize the speed and storage use of the CNN, it is still worthy to create effective algorithms because the CNN models are highly memory requesting and time consuming. This review may not only give a better understanding of CNNs but moreover encourages future research and improvements within the area of CNNs.

References

[1] Vinita Dutt ,Vikas Chaudhry, Imran Khan, “Pattern Recognition: an Overview”, American Journal of Intelligent Systems, 2012, vol.2, pp. 23-27.

[2] Katrine Spirina, “What Computer Vision Is and What It Gives Technology-Led Industries”, InData Labs, 2019

[3] Y. LeCun, Y. Bengio, and G. Hinton, Nature,“Deep learning Nature” , 2015, vol.521, no. 7553, pp. 436–444.

[4] Denny Britz, “Understanding Convolutional Neural Networks for NLP”, WILDML site, November 2015.

[5] Uijwalkarn, "An Intuitive Explanation of Convolutional Neural Networks", Data science blog, August 2016.

[6] Yanan Sun, Bing Xue, Mengjie Zhang, and Gary G. Yen, “Automatically Designing CNN Architectures Using Genetic Algorithm for Image Classification”, arXiv:1808.03818v2 , 2018, pp. 1-12.

[7] Anita Thengade and Rucha Dondal, “Genetic Algorithm – Survey Paper”,International Journal of Computer Applications, 2012, pp. 975-8887.

[8] Haiman Tian, Samira Pouyanfar, Jonathan Chen, Shu-Ching Chen, and Sitharama S. Iyengar, “Automatic convolutional neural network selection for image classification using genetic algorithms”, Proceedings - IEEE 19th International Conference on Information Reuse and Integration for Data Science, 2018.

[9] Masanori Suganuma, Shinichi Shirakawa, and Tomoharu Nagao, “A genetic programming approach to designing convolutional neural network architectures”, IJCAI International Joint Conference on Artificial Intelligence, 2018, vol. 2018-July, pp. 5369-5373.

[10] Julian F. Miller and Stephen L. Smith, IEEE Transactions on Evolutionary Computation 10, 2006, vol.10, pp. 167 -174.

[11] Julian F. Miller and Peter Thomson, “Cartesian genetic programming”, Proceedings of the Third European Conference on Genetic Programming (EuroGP-2000), 2000, vol.1802 of LNCS, pp.121–1 32

[12] Lu Zhang, Miaojing Shi and Qiaobo Chen, “Crowd counting via scale-adaptive convolutional neural network”, Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018, 2017, vol. 2018-Janua, pp.1113-1121.

[13] L. Boominathan, S. S. Kruthiventi, and R. V. Babu. Crowdnet, “a deep convolutional network for dense crowd counting”, MM 2016 - Proceedings of the 2016 ACM Multimedia Conference, 2016, pp. 640-644.

[14] Yuanyuan Zhang, Dong Zhao, Jiande Sun, Guofeng Zou and Wentao Li2, “Adaptive Convolutional Neural Network and Its Application in Face Recognition”, Neural Processing Letters, 2015, vol.43, pp. 389-399.

[15] Clément Farabet, Berin Martini, Polina Akselrod , Selçuk Talay, Yann LeCun and Eugenio Culurciello, “Hardware accelerated convolutional neural networks for synthetic vision systems”, IEEE international symposium on circuits and systems (ISCAS), 2010.

[16] Daniel Strigl, Klaus Kofler and Stefan Podlipnig, “Performance and scalability of GPU-based convolutional neural networks”,Proceedings of the 18th Euromicro Conference on Parallel, Distributed and Network-Based Processing, PDP 2010, 2010, pp. 317-324.

[17] Min Lin, Qiang Chen and Shuicheng Yan, “Network In Network”,2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, 2014, pp.1-10.

[18] E. Wilson , D.W. Tufts, “Multilayer perceptron design algorithm”, IEEE Workshop on Neural Networks for Signal Processing , 2002.

[19] Bert De Brabandere, Xu Jia and Tinne Tuytelaars, “Dynamic Filter Networks”, Advances in Neural Information Processing Systems, 2016, pp. 667-675.

[20] Dingjun Yu, Hanli Wang, Peiqiu Chen and Zhihua Wei, “Mixed pooling for convolutional neural networks”, Rough Sets and Knowledge Technology: 9th International Conference, 2014, vol. 8818.

[21] Xudong Li, Mao Ye, Yiguang Liu and Ce Zhu, ,“Adaptive Deep Convolutional Neural Networks for Scene-Specific Object Detection”, IEEE Transactions on Circuits and Systems for Video Technology2017, vol.29, pp. 2538-2551.

[22] M. Wang, W. Li, and X. Wang, “Transferring a generic pedestrian detector towards specific scenes”, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012, pp. 3274–3281.

[23] Sermanet P, LeCun Y, “Traffic sign recognition with multi-scale convolutional networks”, In: Proceedings of the International Joint Conference on Neural Networks, 2011, pp 2809–2813.

[24] Max Jaderberg, Karen Simonyan, Andrew Zisserman and Koray Kavukcuoglu, “Spatial Transformer Networks”, Advances in Neural Information Processing Systems, 2016, vol. 2015-Janua, pp. 2017-2025.

[25] Xueyun Chen, Shiming Xiang, Cheng-Lin Liu, and Chun-Hong Pan, “Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks”, IEEE Geoscience and Remote Sensing Letters, Oct 2014, Vol. 11, No. 10, pp. 1797-1801.

[26] Qiuyu Zhu and Ruixin Zhang, “ HENet: A Highly EfficientConvolutional Neural Networks Optimized for Accuracy, Speed and Storage”, arXiv:1803.02742v2, 2018.

Online self-assesment as preparation for final exam in primary schools – experience from COVID19 crisis

B.M. Randjelovic*, K.Aleksic ** and D. Stanojevic**
* University of Nis, Faculty of Electrical Engineering, Nis, Serbia
** Institute for Educaton Quality and Evaluation, Belgrade, Serbia
branislav.randjelovic@elfak.ni.ac.rs

Abstract - In this paper we discuss online self-assessment of students in final grade of primary school, implemented during COVID-19 crisis in Serbia. Results and statistics of online self-assessment are explained and discussed.

I. Introduction.

As the number of COVID-19 cases in the world increased and virus began to spread, there were announcements that it will be difficult to finish current school year and teaching process. It became obvious that it will be not possible to carry out all activities and finish the school year in regular conditions.

Ministry of Education, Science and Technological Development of the Republic of Serbia, and thus the entire education system in the country, had to react very fast, especially after the *Decision of the Government of the Republic of Serbia on the suspension of teaching in higher education institutions, secondary and primary schools and the regular work of preschool education institutions ("Official Gazette of RS*", No. 30/2020*)*, which was adopted in accordance with the *Decision on declaring a state of emergency ("Official Gazette of RS*", No. 29/2020*)* and the *Decree on measures during a state of emergency ("Official Gazette of RS*", No. 31/2020*)*. Due to the epidemiological situation in the country, caused by COVID-19 virus, and aiming to ensure uniform acts and treatment in educational institutions, distance learning (Imel, 1998) was organized. Direct work with children in preschool institutions, teaching in all primary and secondary schools, as well as institutions of higher education, has been temporarily suspended. A complex set of activities, that consists of a large number of different programs and digital ways of teaching and learning in preschool, primary and secondary schools was adopted. The focus was on organized distance learning (Moore, et al. 2011), which should contribute to the implementation of the program contents in general education subjects and professional subjects with largest number of classes (Mandic, et al. 2013).

Already on beginning of crisis, a important decision regarding recording of classes and teaching contents has been announced. By this decision, recordings of classes, which would be broadcasted on the national television channels, has been made. Quick reaction of educational system was distinct, but also courage of some teachers, who dared to appear on television, talking to the audience of million, and started very important work for the future of our children and future of our nation. Although distance learning is not our invention and it was known in the world for last several decades, some footage from state television will go down in history and stay in the annals of our educational system[1].

In Section II. we present and discuss online test for self-assessment, that is used in Serbia, during COVID19 crisis.In Section III. we discuss some results and facts, obtained from self-assessment process. Conclusions are given in Secton IV.

II. Online test for self-assesment

During COVID19 crisis, Ministry offered [2] to eighth graders (about 68,000 students in generation[3]) one opportunity to test their knowledge, in order to prepare better for final exam at the end of

This work is supported by Ministry od education, science and technological development of Republic of Serbia, under grants III-43007 and TR-32012

[1] https://www.oecd.org/education/Supporting-the-continuation-of-teaching-and-learning-during-the-COVID-19-pandemic.pdf
[2] https://www.oecd.org/south-east-europe/COVID-19-Crisis-in-Serbia.pdf
[3] https://www.stat.gov.rs/sr-Latn/oblasti/obrazovanje/osnovno-obrazovanje

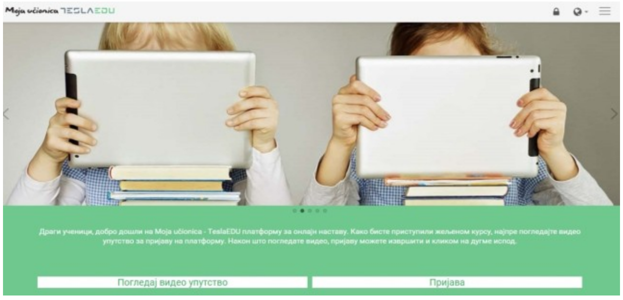


Figure 1. Portal “My Classroom”[9]

primary school. In accordance with situation, but also towards tendencies for digitalization and use of new technologies in education, an online test was only suitable solution. Online self-assessment of knowledge, so called “trial final exam”, was organized in April. Ministry of Education, Science and Technological Development, Institute for Evaluation of Quality of Education, Governmental office for IT and Electronic Administration and Comtrade Company joined efforts, created standardized tests and made it available to eight graders in suitable online environment “MyClassroom” (Fig.1), established and managed by Comtrade, Belgrade.[4]

Online self-assesment, as a tool for teting the sutdnets knowledge is already well-known and used in previous period. There are a lot of webpages, webportals, designed for his purpose[5,6] and a lot of scientific papers and other texts on internet[7,8].

This self-assessment of knowledge was the first and only online test in Republic of Serbia that has been implemented on whole generation, with more than 68000 students. [10] For the first time in the history of Serbian education, all students in one grade had opportunity to test their knowledge in completely changed conditions, in online environment, and with same items. We can expect that this form of support could be usual educational practice in near future.

Tests for self-assessment of knowledge are designed so that students have opportunity to test their knowledge (Barak A. 2011). Two months before official final exam at the end of primary education. They tried to do three tests (Fig.2), that include a total of seven subjects (mother language, mathematics, physics, geography, history, chemistry and biology). So, we decided to develop self-regulated learning and self-assessment of current student’s knowledge level. The self-assessment of knowledge in digital environment was designed with the intention of giving eighth grade students the opportunity to test their own knowledge and, based on those results, to plan further learning.

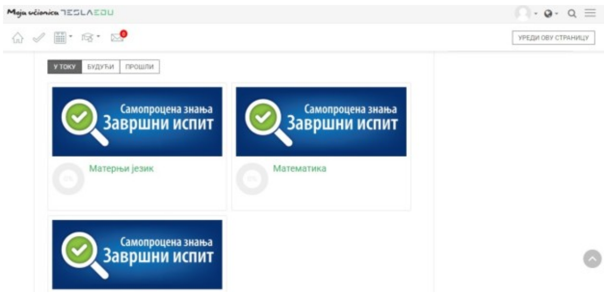


Figure 2. Online enviroment with three tests

Additionally, students had the opportunity to develop and improve their digital competencies.

Institute for Evaluation of Quality of Education prepared items and tasks (mother language, mathematics, chemistry, physics, history, geography and biology) and adapted them for self-assessment in digital environment.

Internet platform “MyClassroom” is based on moodle.[11] Digital environment itself is localized in Serbian language, cyrillic and latin alphabets, as well as in 8 languages of national minorities (Albanian, Bosnian, Bulgarian, Hungarian, Romanian, Ruthenian, Slovak and Croatian).

Eighth grade students took the tests during three days, in period from 8h to 20h. They had a chance to solve each test once, lasting 60 minutes (mother language and combined test) or 90 minutes (mathematics).

---

4 https://www.rasporednastave.gov.rs/lat/alati-uputstva.php
5 https://www.questionmark.com
6 https://www.testportal.net
7 https://elearningindustry.com/qualitative-elearning-assessment-methods-track-online-learners-progress
8 https://www.onlineassessmenttool.com/knowledge-center/assessment-knowledge-center/assessment-methods-and-strategies/item10643
9 https://www.mojaucionica.gov.rs
10 https://www.srbija.gov.rs/vest/en/154782/online-testing-for-eighth-grade-students.php
11 https://www.mojaucionica.gov.rs/mod/page/view.php?id=1

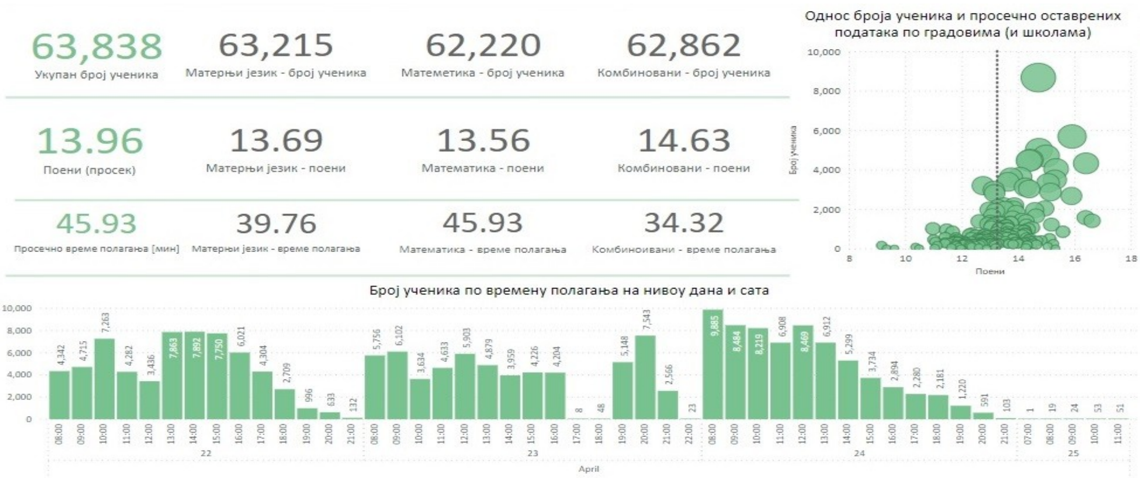


Figure 3. Numbers and facts - after online self-assesment (see [6])

III. RESULTS

Out of a total of 68504 eighth grade students in Serbia, 63215 solved mother language test, 62220 mathematics and 62863 combined test (five sciences). Based on these data, we can be sure that each test was done by more than 91 percent (!!!) of students. So, most of the students approached self-assessment with big responsibility, and there is a reason to believe that they solved tasks independently.

Average number of points was: 13.68 points (mother language), 13.56 points (mathematics) and 14.63 points (combined test)[12]. Everything is shown on following chart, that is obtained from *Comtrade*'s portal (Fig.3, see reference [7])

The results of this online test cannot be compared with the results of the final exams in previous period, having in mind the differences between tests, as well as their purpose. There are good reasons to believe that large percent of students solved tests without someone's help and they used this test properly, for self-assessment of their own knowledge. However, it is interesting how new technologies and portals of this type provide various statistical views and a lot of interesting data, after testing. The following figures shows some statistics, for each subject, for each test separately, for each individual item, all in relation to the language in which the candidate took the test.

After this online self-assessment, all items and results from the all previous final exams, from period 2011 to 2019, were uploaded on portal *www.mojaucionica.gov.rs* sand they are still available to all candidates. Completely in spirit of this distance learning period, eight graders are additionally supported by broadcasting detailed analysis of all items from self-assessment, over national television.

IV. CONCLUSIONS

In this paper, we considered online self-assessment of knowledge of final grade students of primary schools in Serbia. Results and statistics of online self-assessment, are discussed. Based on this consideration, we have several conclusions.

Online self-assessment of eighth grade students was first online testing in our education system, which was conducted comprehensively, on the whole generation of students, at the same time, with the same items and under same conditions - for these reasons, this is a large step forward for the educational system of Republic of Serbia. This model needs to be refined and explored and used.

Using of portal *https://www.mojaucionica.gov.rs* (see [6]) for online self-assessment is a pioneer attempt in our Serbian education and also wider, in region. Despite the fact that existing hardware resources and existing professional support were not

[12] http://www.mpn.gov.rs/analiza-onlajn-testova-za-samoprocenu-znanja-za-zavrsni-ispit-i-postavljanje-testova-i-rezultata-sa-prethodnih-zavrsnih-ispita/

enough to lead to absolutely successful implementation, despite the problems during this process, entire online assessment was successfully completed. Any other subsequent testing, aimed to be performed in similar manner, should be planned more carefully, with additional hardware resources and with slightly better planning.

Expectations about digital abilities of eight graders were significantly lower and such a high percentage of responses to online self-assessment was not expected. As over 91% of eighth graders[13] managed to access the portal and to take online tests, means that most primary school students are digitally literate, that they have some kind of access to internet and that they have appropriate devices (computers, tablets, smartphones) to do test online. However, the goal of our education system for future should be 100% response of students to any digital challenge.

Test results and the number of points that students scored on the online self-assessment, are very similar with earlier final exams or even better. However, they are not relevant for comparison, due to a different environment, different testing conditions as well as a lack of any control. Further work on raising the awareness of students and parents about the importance of online testing without any "cheating" would lead to online testing results that are relevant and usable.

All results are obtained from official webpage of Minsitry of Education or directly from http://tesla.edu.rs and company Comtrade, that is managing portal "My Classroom". There is a very powerfull analysing tool, https://app.powerbi.com that produces various usefull reports, including the one on Fig.3.

Statistics after online self-assessment, as well as all other relevant data on testing, were obtained shortly after completion of tests. They offer us a whole new approach in research work in education, because they go much beyond methods of data processing, that we are using now.

References

[1] A. Barak "Internet-based psychological testing and assessment", In *Online Counseling* . Academic Press 2011, pp. 225-255.

[2] S. Imel, "Myths and realities of distance learning", Columbus, Ohio: ERIC Clearinghouse on Adult, Career, and Vocational Education, Ohio State University. (Eric Document Reproduction Service No. ED 414 446), 1998.

[3] D. Mandic, A. Mandic, "Innovations in Modern Education and Teachers Competences", Asian Journal of Education and e-Learning ,Vol.1,Issue 1, 2013, 70-88.

[4] J.L. Moore, C. Dicskon-Deane, K. Galyen, "e-Learning, online learning, and distance learning environments: Are they the same?", The Internet and Higher Education, Vol.14, Issue 2, 2011, 129-135.

[5] B. Randjelovic, D.Stanojevic, S.Minic, "Analysis of Framework of Digital Competences for Teachers in Serbia and Teachers in European Union. Inv.Lecture on SciConference "Inovative approach to education: situation, dilemas and perspectives, May 31, 2019, Leposavic, Proc.of Papers 35-43.

[6] Ministry od Educaton, Science and Technological Development of Republic of Serbia, Report about distance education in Republic of Serbian emergency state. (May, 2020) (in Serbian).

[7] https://www.mojaucionica.gov.rs

[8] http://www.mpn.gov.rs/analiza-onlajn-testova-za-samoprocenu-znanja-za-zavrsni-ispit-i-postavljanje-testova-i-rezultata-sa-prethodnih-zavrsnih-ispita/

[9] https://www.srbija.gov.rs/vest/en/154782/online-testing-for-eighth-grade-students.php

[10] https://www.rasporednastave.gov.rs/lat/alati-uputstva.php

[11] https://www.paragraf.rs/koronavirus/strucni-komentari/prvi-probni-zavrsni-online-test-za-ucenike-osmi-razred-vanredno-stanje.html

[12] https://www.oecd.org/education/Supporting-the-continuation-of-teaching-and-learning-during-the-COVID-19-pandemic.pdf

[13] https://elearningindustry.com/qualitative-elearning-assessment-methods-track-online-learners-progress

[14] https://www.onlineassessmenttool.com/knowledge-center/assessment-knowledge-center/assessment-methods-and-strategies/item10643

[15] https://sites.psu.edu/onlineassessment/gather-evidence/

[16] https://app.powerbi.com/groups/me/reports/64dd72c3-df55-4c70-8b08-368f554f0c09/ReportSectiond15ca51bc9ecd0381a45?ctid=2b598347-7f8d-4087-9c25-304a5a638626

[13] http://www.mpn.gov.rs/analiza-onlajn-testova-za-samoprocenu-znanja-za-zavrsni-ispit-i-postavljanje-testova-i-rezultata-sa-prethodnih-zavrsnih-ispita/

Usability of Public Web Services

M.Blažić, Lj. Kazi and A. Stojkov
University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
markoblazic93@gmail.com, ljubica.kazi@gmail.com, sahskica.s@gmail.com

Abstract – Contemporary information integration between institutions and companies is implemented using web services. An institution or company enables other business entities to use their data by exposing part of their internal data with the web services, usually hosted at institutional websites. Use of such data is usually regulated with contracts and registration of users. There are other categories of public web services that are hosted for general purposes and do not require any registration and contracting. This paper presents research results related to determining availability and usability of public web services. The results of the paper are presented with the problems of web services usability - the most of free and publicly available web services are socially oriented, but business oriented web services (B2B) require special keys (regulated with contracts and payment) for using.

I. Introduction

Interoperability has been a basic requirement of the modern information systems for over two decades. Data and information interoperability gained increasing attention for several reasons. Excellent progress in communication between users or companies by virtue of the internet, which enabled easy access to the information [1]. The organizations and firms are using distributed information systems to communicate with each other and to enable sharing data to be faster and easier.

Earlier institutions and companies were communicating and sharing information directly or through the internet via e-mail. One of the main problems of that approach was accessibility, credibility and safety of data. Using distributed information systems with use of web services, the possibility of errors was minimized. Aim of this paper is to explore the areas of use of public available web services.

Purpose of this research was to determine which types of public web services exist and what is their area of use. Also the goal is to research existing professional and scientific solutions in integration web services, with web application and integration of information system organization by means of web services.

This paper is organized as follows: first section is related to theoretical background, then related work and professional example is presented, followed by research methodology and results and finally conclusion section.

II. Theoretical Background

Web services could be described as independent, modular, distributed and dynamic applications using XML, SOAP, RESTful, WSDL and UDDI internet open standard backbone protocols. XML is used for tagging content while SOAP or REST are used for data transfer. WSDL are used for showing available services and UDDI are used for listing all services that are provided with WSDL.Web services are also using standard web protocols such as HTTP/ HTTPS.[2]

It is also crucial to mention that web services are providing users with API (application programming interface), set of commands or options, that are allowing access to data. The most famous API accessible to all people are Google maps, API that provides access to geographical data and guidelines from Google Maps database.

The base of every web service is messages, in the next segment, SOAP and REST will be presented.

SOAP is a protocol for access to web services, based on XML instead of HTTP. SOAP aka(Simple object access protocol) is designed as a language of the middle level, which allows programmers from different program languages to communicate. The communication between different language is very complex, and method that was shown as solution was XML, medium that every programming language can understand.[3]

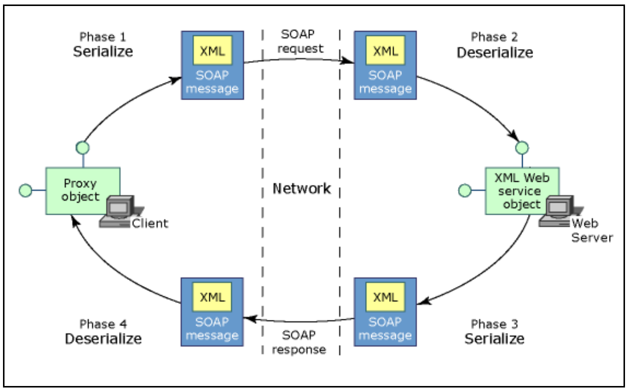


Figure 1. SOAP data transfer[20]

Advantages and disadvantages of SOAP are presented at figure 2.

Advantages	Disadvantages
High level of Interoperability	"SOAP is not Simple"; Complex to implement
HTTP/SOAP can travel through firewalls	Securing SOAP Web services is complicated, "smart" firewalls are needed
Human legibility of SOAP messages; Easy debugging	Higher network load compared to other protocols due to XML format
XML validators may prevent errors in the Web service	Buggy SOAP frameworks may lead to malfunctioning Web services
Comprehensive protocol documentation through technologies such as WSDL	High hardware requirements (CPU/Memory).

Table 2. Advantages and disadvantages of SOAP[21]

REST (Representational State Transfer) is used for developing web services that are "light" by nature, maintaining and scalable. is based on HTTP, which is basic web protocol. [4]

REST architecture key characteristics include:

1. REST client/server – the basic characteristic of REST arhitecture, client is sending request to the server, server is declining or accepting request and sending appropriate response to the client.
2. Stateless – concept is based on the fact that responsability in providing information totally depends on client. Stateless state is meaning that server doesn't remember last request, and each request looks separatly.
3. Cache – concept is developed to prevent flow in statless concept. In some ocations client is sending multiple times the same request to the severs, and to prevent "slowing down" network traffic, the request is saved on client.
4. Multilayer system – this concept provides possibility of adding new layer between server in client, which would host REST web service.

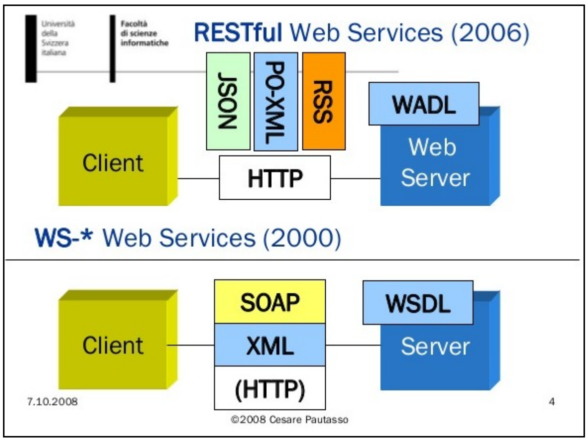


Figure 3. Differences between SOAP and REST [22]

For common users the most important type of web services are e–commerce (electronic commerce), services that can be presented as transactions of goods buying or selling. These transactions are usually business to business (B2B), customer to business (C2B), business to customer (C2B) and customer to customer (C2C). [5]

There are some advantages and disadvantages using e–commerce. Advantages are [6]:

1. Availability - services are available 24/7, approachable to user at any moment,
2. Access speed - with only few click clients can buy or sell goods or data,
3. International access – clients can buy or sell their good in any country,
4. Less cost – e commerce doesn't have cost for rent, supplies, shipment etc.

There are also few disadvantages:

1. The most important one is lack of thrust, it is general thought that paying online is not safe enough,
2. Products are not "touchable" - meaning that clients can't touch the product and see it before buying it.
3. Results are not coming easy – there has to be some time and some buyers for clients until success comes.

Mobile e – commerce is new and growing type of e – commerce, which is based on online transaction with mobile devices. Mobile e – commerce includes mobile shopping, mobile banking and mobile payment.[7]

Mobile e-commerce shopping apps with the most monthly active users in South-East Asia

MALAYSIA	THAILAND	VIETNAM	SINGAPORE	PHILIPPINES
1. Lazada	1. Lazada	1. Shopee	1. Lazada	1. Lazada
2. Shopee	2. Shopee	2. Lazada	2. Qoo10 Singapore	2. Shopee
3. Taobao	3. AliExpress	3. Tiki.vn	3. Shopee	3. Zalora
4. 11street	4. JD Central	4. Sendo	4. Taobao	4. Amazon
5. AliExpress	5. Amazon	5. Adayroi	5. ezbuy	5. AliExpress
6. Zalora	6. eBay	6. AliExpress	6. Zalora	6. BeautyMNL
7. Lelong.my	7. Alibaba.com	7. Amazon	7. AliExpress	7. eBay
8. eBay	8. Chilindo	8. eBay	8. Amazon	8. Sephora
9. Amazon	9. Zilingo Shopping	9. Alibaba.com	9. Amazon Primenow	9. Alibaba.com
10. Go Shop	10. Joom	10. Lotte.vn	10. Asos	10. Althea

Source: App Annie Intelligence

Figure 4. The most used mobile e–commerce applications in South-Asian countries [23]

III. Related Work

Addressing this topic, professor from Faculty of Sciences (Vrije Universiteit Amsterdam), Dieter Fensel and Christoph Bussler from Oracle corporation, stated that web services will transform web from gathering information into distributed system available to anyone.[11]

DAML is providing a description of Web services on the level of the application layer, describing what service can do, not just how service does it [8].

WFMS (workflow management systems) are used for many types of business process for over a decade.

Clients and suppliers are in obligation to sign contract for quality of the product.[12]

Structure of web services allows programmers to create new web services based on the description provided by DAML. This application is easy to implement and allows programmers reusing them, while customers have access to the wide spectrum of complex services [9].

In research [10], potential for achieving dynamic, scalable market as a solution for e market, which is recently focused on semantic web services, web services that are complemented with machine semantic.[10]

SOC (service – oriented computing) allows companies to collaborate in unprecedented way, supporting dynamic composition of services that has to be able to locate and connect to constantly changing numbers in providers based on their QoS.[13]

Discovering right service for use is the key problem, because number of services is expected to increase drastically. In present time, searching for web service is based primally on name of the web service or interface, but in this paper new idea is shown, searching web services based on their description provided by OWL.[14]

In paper [15], it is presented as a set of tools and techniques for analyses of interaction web services which are communicating through XML- messages. Interaction is modelling between complex web services as communication, a global set of messages which is exchanged by web services.

Business-to-business technologies are as old as the web. B2B application are one of the first application to take advantage of advance in computer networking.[16]

As a result of interoperability researchers in life-science are currently limited to the small scale experiment and analyses. As a result of this problem, limitation can be addressed with descriptions and interfaces with semantic information, so this problem can be resolver programmatically.[17]

In the next research Colombo framework is presented, where services are described as atomic processes; their effect on the real world, their behaviour based on tranzition and message through transaction. Through Colombo, composition of automatic services can be analyzed, then use that data to form complex algorithm for complex services.[18]

Two web services compatibility doesn't depend only on correct typing message parameters, but it also depends on their dynamic behavior. Giving a simple behavior description, based on complex algorithms, could be helpful in discovering incompatibility in interaction between two web services [19].

IV. Research Methodology

This paper presents research results from master thesis [24], defended at University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Serbia.

Hypothesis presented in [24] and this paper is: Web services are publicly available on the internet and they are available for integration with web applications.

Methods that are used in this research include:

- Methodological planning,
- Analysis of theoretical concepts,
- Analysis of existing research results,
- Analysis of professional results and examples.

Results are presented analytically and statistically.

In the analysis of professional results, i.e. publicly available web services, analytical approach is performed with one business-oriented example, while statistically processed data are initially collected from various web services hosting web sites.

In order to determine which area is the most popular for using generally available free web API, 150 publicly available web APIs were explored.

V. Research Results

A. *Professional example*

One example that was used to analyze a professional solution was a list of web services offered by the National bank of Serbia. The web services that are available at this website could be used – some could be used free, but others require previous registration. Some of free web services include:

1. Service for access to public codes,
2. Exchange rate list access service,
3. Service for access to the current exchange list,
4. Service for access to exchange rates for foreign exchange transactions,
5. Value service of investment units of voluntary pension funds,
6. Service for access to effective overnight rates,
7. Service for access to register of account holders.

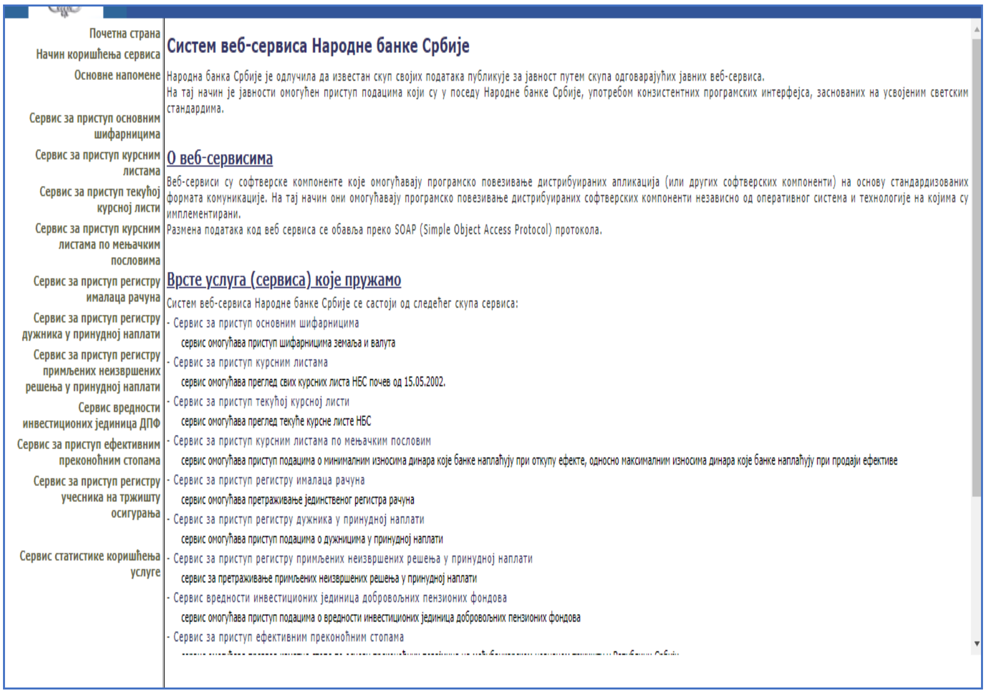


Figure 5. List of publicly available web services of National bank of Serbia [25]

B. *Statistical results*

Results of research show that web APIs are becoming a popular solution for developing web applications. With market development, exigency for web APIs are growing every day with purpose for providing better quality of service for customers. Figure 6. shows results of analysis of usage areas of public available web services.

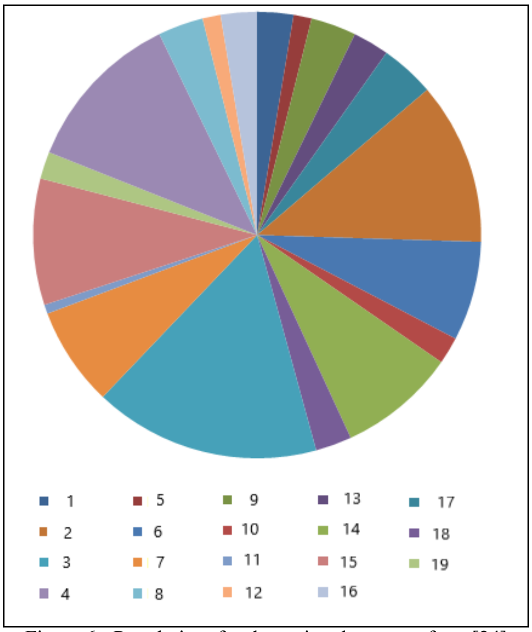


Figure 6.- Popularity of web services by areas of use [24]

Areas of use include: 1- Commercials, 2- Development tools, 3- Google, 4- Public social services, 5- Classifications, 6- Nutrition, 7 -Location, 8 - Social networking tools, 9 -Communications, 10 - Fun Making, 11- API store, 12 - sport, 13 - Criminalistics, 14 – Fun sources, 15 - Misc, 16 - Weather Forecast, 17 - Data, 18 - Gaming, 19 -News.

Statistical analysis presented at Figure 6. shows that the most popular areas of publicly available web services are related to: Nutrition, Fun, Location, Public social services, Social networking tools etc.

Second statistical analysis is presented at figure 7 and it is related to most popular B2B web services.

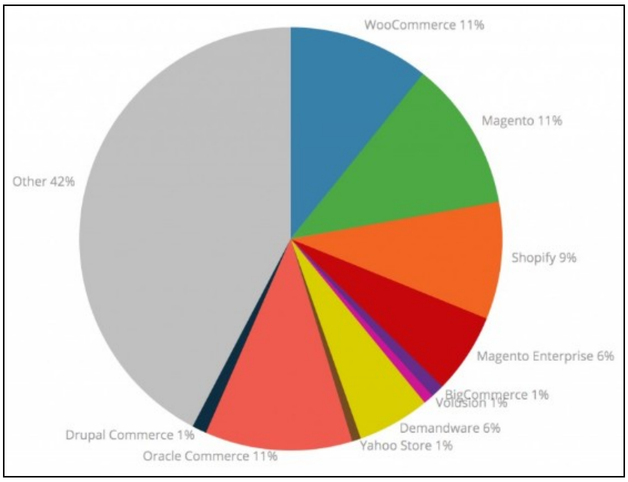


Figure 7.-Statistic of e–commerce solutions usage [26]

The most popular B2B web services are: OpenCart B2B, Shopify Plus, Magento Commerce, Pepperi, BigCommerce, PrestaShop, InsiteCommerce, NetSuite, SAP Hybris, GoEcart.

	Price	Usability	Support	Open Source	Integrations
OpenCart B2B	★★★★★	★★★★★	★★★	YES	★★★★★
Shopify Plus	★★★	★★★★★	★★★★★	NO	★★★
Magento Commerce	★★★	★★★★★	★★★★★	NO	★★★
Pepperi	★★★	★★★★★	★★★	NO	★★★★★
BigCommerce	★★★	★★★★★	★★★	NO	★★★★★
PrestaShop	★★★★★	★★★	★★★	YES	★★★★★
InsiteCommerce	★★★	★★★	★★★★★	NO	★★★★★
NetSuite	★★★	★★★	★★★★★	NO	★★★★★
SAP Hybris	★★★	★★★★★	★★★	NO	★★★★★
GoEcart	★★★	★★★	★★★★★	NO	★★★★★

Figure 8.-Comparation of 10 most popular B2B e – commerce web services[26]

VI. CONCLUSION

Contemporary web application development includes using web services, as valuable sources of data and additional functionality.

Aim of this research was to explore public web services and their usage areas and availability.

The result of the research showed that there are many public web services that are available to the programmers or common users for any kind of interaction. Areas of public use include APIs in various usage areas, such as health and sport related web services, social networking gaming and others. There is a large number of free web services available without registration.

Second type of web services include business-related functionalities and data, which requires registration and tokens for access. Business-related APIs include e-commerce platforms and solutions, as well as financial services, such as currency exchange APIs.

Finally, it could be concluded that wide use of web services enable integration of information systems and different platforms, such as desktop applications, client/server applications, web applications, as well as mobile applications and their data exchange, which improves their usability.

REFERENCES

[1] Sheth A.P. “Changing Focus on Interoperability in Information Systems:From System, Syntax, Structure to Semantics”. In: Goodchild M., Egenhofer M., Fegeas R., Kottman C. (eds) Interoperating Geographic Information Systems. The Springer International Series in Engineering and Computer Science, vol 495. Springer, Boston, MA, 1999.

[2] Web services https://searchapparchitecture.techtarget.com/definition/Web-services

[3] What is SOAP? https://www.guru99.com/soap-simple-object-access-protocol.html

[4] What is RESTfull web service? https://www.guru99.com/restful-web-services.html

[5] What is e – commerce? https://www.shopify.com/encyclopedia/what-is-ecommerce

[6] Advantages and disadvantages of e – commerce https://www.doofinder.com/en/blog/what-is-e-commerce

[7] Mobile e – commerce https://searchcio.techtarget.com/definition/e-commerce

[8] Ankolekar A. et al. DAML-S: Web Service Description for the Semantic Web. In: Horrocks I., Hendler J. (eds) The Semantic Web — ISWC 2002. ISWC 2002. Lecture Notes in Computer Science, vol 2342. Springer, Berlin, Heidelberg, 2002

[9] N. Milanovic and M. Malek, "Current solutions for Web service composition," in IEEE Internet Computing, vol. 8, no. 6, pp. 51-59, Nov.-Dec. 2004, doi: 10.1109/MIC.2004.58.

[10] Roman, Dumitru et al.’Web Service Modeling Ontology’. 77-106 print, 1 Jan, 2005

[11] https://www.sciencedirect.com/science/article/pii/S1567422302000157

[12] https://www.sciencedirect.com/science/article/abs/pii/S157082680400006X

[13] https://dl.acm.org/doi/10.1145/1013367.1013379

[14] https://sites.cs.ucsb.edu/~su/papers/2005/SCC2005b.pdf

[15] http://www.ra.ethz.ch/CDstore/www2004/docs/1p621.pdf

[16] http://csis.pace.edu/~marchese/CS865/Papers/30120059.pdf

[17] https://www.sciencedirect.com/science/article/pii/S1532046408000269

[18] https://www.dis.uniroma1.it/~degiacom/papers/2005/BCDHM_VLDB2005.pdf

[19] https://www.semanticscholar.org/paper/When-are-Two-Web-Services-Compatible-Bordeaux-Sala%C3%BCn/e09bb2ee683351403ad3384cc7c8a14a1c98568e#paper-header

[20] Soap data transfer, https://www.sololinux.es/servicio-web-sencillo-con-soap-y-php/

[21] Comparison of SOAP and REST http://violin.qwer.tk/dusty/thesis/html/node79.html

[22] Differences between SOAP and REST, http://yadna05.blogspot.com/2015/10/difference-between-rest-and-soap-based.html

[23] Most used web services in South Asia region, http://aec.utcc.ac.th/major-expenditure-in-e-commerce-to-continue/

[24] M.Blazic: Availability and business use of public web services in ditributed information systems, (In Serbian:Raspoloživost i poslovna primena javnih veb servisa u distribuiranim informacionim sistemima), Master Thesis, University of Novi Sad, Technnical Faculty “Mihajlo Pupin” Zrenjanin, menthor: Ljubica Kazi, 2019.

[25] National Bank of Serbia web services, https://webservices.nbs.rs/CommunicationOfficeSiteDoc/SerCyrl/default.html

[26] The most popular B2B e-Commerce Platforms, https://www.namogoo.com/blog/ecommerce/top-10-b2b-ecommerce-platforms-for-2019/

Fuzzy Inference System for SARS-CoV-2 Diagnostics

I. Vecštejn*, V. Brtka*, E. Brtka*, M. Mazalica*
*University of Novi Sad, Technical faculty "Mihajlo Pupin", Đure Đakovića bb, 23000, Zrenjanin, Republic of Serbia
igor.vecstejn@gmail.com; brtkav@uns.ac.rs; eleonorabrtka@gmail.com; milicamazalica@gmail.com

Abstract: The paper investigates the feasibility of applying fuzzy logic controllers to the domain of decision making, especially in decision making processes related to SARS-CoV-2 virus diagnostics. The issue refers to a system that makes decisions based on the opinion of several experts and several input criteria. The implementation of such a system is presented and initial tests are performed. The boundaries within which the system makes decisions have been established. Some of the benefits of the system are easy adaptation, implementation and easy and cheap use.

I. Introduction

Managing complex business processes, organizing projects, production and research, often mobilizes significant organizational resources. These resources are the material, human and conceptual, but their selection is in itself a challenging task. In situations where it is necessary to investigate phenomena or procedures that are not sufficiently known, while their mere existence requires a rapid response, the problem of organization is further complicated. A specific example of a complex phenomenon that is not sufficiently known and requires a rapid response is the occurrence of a pandemic caused by the SARS-CoV-2 virus. There are no definitive and fully concretized procedures that should be performed in case of infection with this virus, nor are there precise tests or a definitive way of timely determination of the existence of infection in a person. This leads to differences in the opinions of experts dealing with this topic, but it happens that their opinions are often partially or completely contradicted.

The problem that this research deals with is the derivation of a common decision based on the opinions of individual experts. Experts make their opinion about the probability of a person's infection based on several symptoms. Therefore, the procedure of Multi-Criteria Multi-Expert Decision Making (MCMEDM) was used. This procedure under fuzzy environment has been shown to be an efficient technique to obtain general decision. Currently, it is not clear which symptoms, ie which combination of symptoms indicates early diagnosis of the infection. Although there are many scientific papers on this topic, and many researches are currently being conducted, in this paper the symptoms recommended by the World Health Organization (12) are observed: fever, dry cough, and tiredness. There are less significant symptoms like: aches and pains, nasal congestion, headache, conjunctivitis, sore throat, diarrhea, loss of taste or smell or a rash on skin or discoloration of fingers or toes. As in [12] "These symptoms are usually mild and begin gradually. Some people become infected but only have very mild symptoms". The previous quote tells us why early diagnosis of infection is difficult. Therefore, a model of a Fuzzy Logic Controller (FLC) is proposed in the paper, which infers a general decision based on the opinion of individual experts and several symptoms. The benefits of the system are easy adaptation which results in quick decision making.

The paper is organized as follows: Chapter two contains a description of some previous cases of fuzzy application for MCMEDM. Chapter three describes the specific implementation of FLC, and Chapter four contains case studies. Conclusions and comments are given in the last, fifth chapter.

II. Previous Work

This chapter provides a partial overview of the application of MCDM and MCMEDM methods in various domains in order to confirm the importance of this method. The domains of energetic, transportation and medicine were selected. Based on this type of review, first of all, one can gain insight into the general mechanisms of fuzzy MCDM implementation, and then the implementation of the situation with several experts. It also gives insight in a number of papers dealing with the SARS-CoV-2 virus.

In [11] the energy decision and policy-making problems are tackled, including selecting among energy alternatives, evaluating energy supply technologies, determining energy policy and energy planning. These issues can be evaluated by using multi-criteria decision making (MCDM) methods. In order to address uncertainties in human opinions, the fuzzy set theory can be successfully used together with the MCDM methods to get more sensitive, concrete and realistic results. As in [11]: "The results of this study indicate that fuzzy Analytic Hierarchy Process (AHP), as an individual tool or by integrating with another MCDM method, is the most applied MCDM method and type-1 fuzzy sets are the most preferred type of fuzzy sets". The research deals with multiple criteria concerning energy decision making problems. Each criteria has its own assessment of the importance of influencing the final decision: Environmental issues 23%, Quality 2%, Efficiency 3%, Geographical and structure 6%, Economical 27%, Technological 17%, Social/Political 18% and Risk 4%. A percentage distribution of the importance of criteria for renewable energy evaluation problems as well as for power plant site selection problems was performed. An analysis of the literature has established that various MCDM techniques have been researched in the field of energy in 150 recently published scientific papers.

In [8] fuzzy MCDM techniques are applied to three hybrid wind farms to determine the best strategy. Machine learning method is utilized for wind speed forecasting and penalty cost estimation. Fuzzy TOPSIS [1] and Fuzzy COPRAS methods are applied. Fuzzy TOPSIS involves multiple decision makers and "...is capable of reflecting approval or rejection, and ambiguity displayed by decision-makers through the consideration of real-life uncertainties and linguistic human decisions". The paper presents the steps of application of fuzzy TOPSIS method. Fuzzy CORPAS as an extension of CORPAS method is used to identify the best alternative [2]. Both methods use the opinions of multiple decision makers (experts) which leads towards MCMEDM system.

Descriptive study presented in [10] deals with selecting qualified nurses. The study was conducted in three stages: Identification of criteria, Weighting the criteria and Assessing the performance of nurses. The DEMATEL and ANP [3, 4] methods were used to discover the weighted criteria, while VIKOR method [5, 6] is used to select nurses. The VIKOR method enables the selection of the best option, ie a better decision.

In [9] is presented an approach for assessing and mapping maritime transportation risk by combining geospatial techniques and MCDM in order to "develop effective mitigation plans and strategies to improve navigation safety". This paper is important because it provides insight into criteria threshold evaluation by several experts. The range of values of each criterion was categorized into five levels, where 1 represents very low and 5 represents very high. The FAHP (Fuzzy AHP) technique [7] was applied to weight each criterion. Study results affect minimization of the severity of maritime accidents.

Multiple researches of SARS-CoV-2 virus are underway, but there are already corpus of knowledge that is available.

Investigation presented in [13] deals with identifying patients with acute SARS-CoV-2 virus. For now, Reverse Transcriptase real-time Polymerase Chain Reaction (RT-PCR) of respiratory specimens represents the standard for identifying patients with acute SARS-CoV-2 infection as well as asymptomatic carriers. As stated in [13]: "However, serologic assays are urgently needed to supplement the diagnostic repertoire in identifying patients with past SARS-CoV-2 infection". As this is laborious task, it is necessary to develop a system for early detection of virus carriers, which leads us to the [12] recommendations and the design of the system based on the opinions of several experts.

Also, as stated in [14]: "Patients infected by SARS-CoV or MERS-CoV were previously reported to have antibody responses but exhibited defective expression of types I and II interferons (IFNs), indicative of poor protective immune responses. However, to date, there were few studies in characterizing the immune responses , especially adaptive immune responses to SARS-CoV-2 infection".

In order to develop a system of fast and cheap diagnostics, we took into account all the above and decided to develop a system that takes into account the opinions of several experts and also, more criteria.

III. FLC Design

Previous studies indicate the success of using MCDM methods in different areas. All methods considered involve the use of fuzzy environments or fuzzy numbers. Some of the methods (AN, TOPSIS) are adapted to the ranking of decisions, while some methods (eg VIKOR) enable the selection of the best decision. However, in this research it was decided to use a Fuzzy Logic Controller, so that it is necessary to design and implement an FLC that will allow multiple experts to express their opinions (the opinions of different experts may coincide, but may also be contradictory). Also, the FLC should reach a general conclusion based on all expert opinions. The Mamdani FLC architecture suits this task: experts enter their opinions in the form of fuzzy if-then rules and adjust Membership Functions (MF) so that they reflect their specific opinion. The controller is shown visually, in Figure 1.

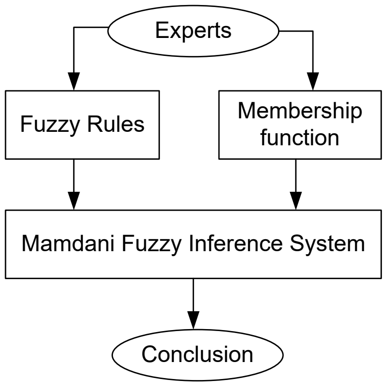


Figure 1. FLC Architecture in Decision Making

Mamdani Fuzzy Inference System (MFIS) generates decision based on the opinions of all experts. FLC must support the definition of membership functions, in this case usual triangular (trimf) and trapezoidal (trapmf) membership functions are used.

For simplicity, three symptoms were observed: fever, dry cough, and tiredness, while the output is the probability that the person is infected, expressed as a percentage. Symptoms are fuzzy variables, and their values are defined by MFs. All values are defined in the 0 - 100 range in percentages. The code illustrating the implementation of MFIS is shown below in Matlab/Octave.

```
% MFIS SARS-CoV-2

clc; clear;

outSpace=zeros(7,100);
step=1;
x=0;

% inputs
fever=89;
cough=21;
tiredness = 88;
```

```
% low = trapmf(-10,0,20,40,x)
% medium = trimf(20,50,80,x)
% high = trapmf(60,80,100,110,x)

for i=1:100
  x=x+step;
  outSpace(1,i)=mytrapmf(-10,0,20,40,x);%low
  outSpace(2,i)=mytrimf(20,50,80,x);%medium
  outSpace(3,i)=mytrapmf(60,80,100,110,x);%high

  outSpace(4,i)=mytrapmf(60,80,100,110,x);%high
  outSpace(5,i)=mytrapmf(60,80,100,110,x);%high
  outSpace(6,i)=mytrapmf(60,80,100,110,x);%high
end

alpha1=min(min(mytrapmf(-
10,0,20,40,fever),mytrapmf(-
10,0,20,40,cough)),mytrapmf(-10,0,20,40,tiredness))
alpha2=
min(min(mytrimf(20,50,80,fever),mytrimf(20,50,80,coug
h)),mytrimf(20,50,80,tiredness))
al-
pha3=min(min(mytrapmf(60,80,100,110,fever),mytrapmf
(60,80,100,110,cough)),mytrapmf(60,80,100,110,tirednes
s))

al-
pha4=min(mytrapmf(60,80,100,110,fever),mytrimf(20,50
,80,cough))
al-
pha5=min(mytrimf(20,50,80,fever),mytrapmf(60,80,100,
110,cough))
al-
pha6=min(mytrapmf(60,80,100,110,cough),mytrapmf(60,
80,100,110,tiredness))

for i=1:100
  if (outSpace(1,i)>alpha1) outSpace(1,i)=alpha1;end
  if (outSpace(2,i)>alpha2) outSpace(2,i)=alpha2;end
  if (outSpace(3,i)>alpha3) outSpace(3,i)=alpha3;end
  if (outSpace(4,i)>alpha4) outSpace(4,i)=alpha4;end
  if (outSpace(5,i)>alpha5) outSpace(5,i)=alpha5;end
  if (outSpace(6,i)>alpha6) outSpace(6,i)=alpha6;end
  outSpace(7,i) =
max(max(max(max(max(outSpace(1,i),outSpace(2,i)),out
Space(3,i)),outSpace(4,i)),outSpace(5,i)),outSpace(6,i));
end

maxVal=max(outSpace(7,:));
left=0;
for i=1:100
  if (left==0 && outSpace(7,i)==maxVal)
left=i;break;end
end

right=101;
for i=100:-1:1
  if (right==101 && outSpace(7,i)==maxVal)
right=i;break;end

end
```

```
left
right

out_com=round(((left+right)/2))
```

Figure 2. MFIS Implementation

The opinions of three experts were taken into account, each of them defined few fuzzy rules. In the general case, each of the experts can define a number of fuzzy rules. Also, for simplicity, it is considered that the experts agreed on the definition of membership functions:

```
low = trapmf(-10,0,20,40)
medium = trimf(20,50,80)
high = trapmf(60,80,100,110)
```

In the general case, each of the experts would express their view of the value of fuzzy variables. This would have a direct impact on the adaptability of the system.

IV. CASE STUDY

A case study was conducted for three experts and several rules. The assumptions are as follows: All experts agree that if the symptoms are mild, it is unlikely that the person is infected. Also, if all the symptoms are pronounced, the probability of infection is high, and if the symptoms are moderate, the probability of infection is about 50%. The above is formalized using three fuzzy rules:

IF fever is low AND cough is low AND tiredness is low THEN CoV2 is low
IF fever is high AND cough is high AND tiredness is high THEN CoV2 is high
IF fever is medium AND cough is medium AND tiredness is medium THEN CoV2 is medium

Disagreement in the opinions of experts refers to a situation in which the chance of infection is high: Two experts believe that tiredness has no effect on this, but they do not agree on the effect of fever and cough. One of the experts believes that a pronounced fever has a great influence on the presence of infection, and the cough less, while another expert believes that the situation is exactly the opposite. This situation is formally expressed by two fuzzy rules:

IF fever is high AND cough is medium THEN CoV2 is high
IF fever is medium AND cough is high THEN CoV2 is high
In contrast, a third expert believes that fever is not a true indicator but that cough and tiredness must be intensified:
IF cough is high AND tiredness is high THEN CoV2 is high.

The case study was conducted in the Octave environment (www.octave-online.net). Characteristic inputs were selected, and outputs were calculated for them taking into account all fuzzy rules, see Table 1. The mechanism is based on the Mamdani Fuzzy Logic Controller, and the program code has been previously shown.

TABLE 1. CASE STUDY

Person	Fever [%]	Caugh [%]	Tiredness [%]	SARS-CoV-2 presence [%]
1	30	50	90	51
2	2	0	0	11
3	89	51	88	90
4	89	76	88	88
5	89	21	88	81
6	27	98	2	83
7	13	24	38	20
8	100	100	100	90
9	32	4	2	17

The sensitivity of the system is in the range of 11% to 90%, which is understandable because it cannot be said with full certainty that a person is infected or not. Also, cases 3, 4 and 5 show how the variation of one criterion (in this case cough) affects the result. It has been shown that the system very rarely comes up with results ranging from 55% to 80%. Although at first glance this may be considered a shortcoming of the system, it should be borne in mind that the results from this range are often very confusing. Namely, low percentage values express a serious suspicion of the presence of a virus, conversely, high percentage values indicate the presence of a virus. Values of about 50% indicate cases for which the presence or absence of the virus cannot be established.

V. CONCLUSIONS

The issue of decision-making based on the value of several criteria has so far been very well researched, and a number of methods have been developed under the name Multi Criteria Decision Making (MCDM). If the decision rules generate more experts then we get Multi Criteria Multi Expert Decision Making MCMEDM. Based on the results achieved in the field of fuzzy MCMEDM, the design of Fuzzy Logic Controller was proposed, which, based on the opinion of several experts, evaluates the presence of SARS-CoV-2 virus, and based on several inputs. Such a system has been implemented and tested. The limits within which the system gives an estimate have been established: from 11% to 90%, so here is no complete certainty in the presence or absence of the virus that corresponds to the real situation. Also, the system "leans" towards higher or lower percentage values, avoiding the range of 25% to 45% and 55% to 79%, which are often considered confusing from the point of view of decision making.

The benefits of the system are reflected in the easy adaptation that must occur when research on the SARS-CoV-2 virus yields new results. Also, the system is easy to implement and easy and inexpensive to use. The advantage of the system is also reflected in the complex decision-making rules that are not limited to completing the questionnaire.

Future work involves the introduction of additional input criteria and more decision-making rules. Also, testing of the system on real-world data sample is needed.

REFERENCES

[1] Chen C-T. Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Set Syst 2000;114(1):1-9.

[2] Ghose D, Pradhan S, Tamuli P. Shabbiruddin, Optimal material for solar electric vehicle application using an integrated fuzzy-COPRAS model, Energy Sources, Part A: recovery, Utilization, and Environmental Effects. 2019. p. 1-20. https://doi.org/10.1080/15567036.2019.1668879.

[3] D. Vujanovic, et al., Evaluation of vehicle fleet maintenance management indicators by application of DEMATEL and ANP, Expert Syst. Appl. 39 (2012) 10552–10563.

[4] T. Saaty, Decision making with the analytic network process, Int. J. Serv. Sci. 1 (1) (2008).

[5] M. Chu, et al., Comparison among three analytical methods for knowledge communities group decision analysis, Expert Syst. Appl. 33 (4) (2007) 1011–1024.

[6] S.-P. Wan, Q.-Y. Wang, J.-Y. Dong, The extended VIKOR method for multi-attribute group decision making with triangular intuitionistic fuzzy numbers, Knowl. Base Syst. 52 (2013) 65–77.

[7] Chang, D.Y., 1996. Applications of the extent analysis method on fuzzy AHP. Eur. J. Oper. Res. 95 (3), 649–655.

[8] Harsh S. Dhiman, Dipankar Deb, "Fuzzy TOPSIS and fuzzy COPRAS based multi-criteria decision making for hybrid wind farms", Energy, Volume 202, 1 July 2020, 117755.

[9] Xiao Zhou, Liang Cheng, Manchun Li, "Assessing and mapping maritime transportation risk based on spatial fuzzy multi-criteria decision making: A case study in the South China sea", Ocean Engineering, Volume 208, 15 July 2020, 107403.

[10] Rahati Meghdad, Rohollahi Nayereh, Sakeni Zahra, Zahed Houriye, Nanakar Reza, "Assessment of the performance of nurses based on the 360-degree model and fuzzy multi-criteria decision-making method (FMCDM) and selecting qualified nurses", Heliyon 6 (2020) e03257.

[11] İhsan Kaya, Murat Çolak, Fulya Terzia, "A comprehensive review of fuzzy multi criteria decision making methodologies for energy policy making", Energy Strategy Reviews 24 (2019) 207–228.

[12] World Health Organization, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/coronavirus-disease-answers?query=symptoms+of+covid+19, accessed 05.06.2020.

[13] Alexander Krüttgen, Christian G. Cornelissen, Michael Dreher, Mathias Hornef, Matthias Imöhl, Michael Kleines, Comparison of four new commercial serologic assays for determination of SARS-CoV-2 IgG, Journal of Clinical Virology 128(2020) 104394.

[14] Ni et al., Detection of SARS-CoV-2-Specific Humoral and Cellular Immunity in COVID-19 Convalescent individuals (Report), 2020, Immunity 52, 1–7

Forecasting the total number of Internet of Things connections using trend analysis

M. Gaborov*, M. Blažić**
University of Novi Sad, Technical faculty "Mihajlo Pupin" Zrenjanin, Djure Djakovica bb, Serbia
majagaborov26@gmail.com, markoblazic93@gmail.com

Abstract: The paper presents the trend of the number of Internet of Things (IoT) connections by years. Using Microsoft Excel, an observed phenomenon is graphically represented, where the trend line is included. It is shown by using various mathematical functions. Based on that, it was predicted how many connections there will be in the future by applying mathematical functions as well as by applying built-in functions offered by Excel. According to the trend line on the graph, it is concluded with the help of mathematical functions that it is best to show the trend. Use of the coefficient of determination shows, which of the mentioned mathematical functions is the most reliable.

Keywords: linear function, non – linear function, forecast, connections, Internet of Things

I. INTRODUCTION

The paper presents the number of Internet of Things connections, which is taken from other sources, and based on these data, the number of future connections is predicted by using linear and nonlinear functions and built-in functions in Microsoft Excel. The ability for everyday devices to communicate with each other and/or with people is becoming widespread and is called the Internet of Things.

The Internet of Things is a concept that takes into account widespread things that have the ability to communicate with other things/objects in order to create new applications and services over wireless or wired connections[1]. Smart things have the ability to communicate wirelessly with each other and between interconnected objects within an ad-hoc network. Smart things can interact with the local environment through reading and activating existing capabilities [1]. Statistical analysis of time series aims to provide a basis for drawing conclusions about the characteristics of the development of phenomena over time. It is possible to show the time series graphically and based on that, tendencies can be noticed. The basic tendency of the development of a phenomenon in time is called a trend and is represented by some function of time[2].

II. THEORETICAL BACKGROUND

Internet of Things is a network of interconnected physical devices that exchange data and information through embedded sensors and actuators[3]. IoT refers to a network of physical objects with embedded electronics, software, sensors, or connectivity that allow objects to exchange data with another connected device[4].The devices are connected to the internet and so human life is made easier. It consists of a number of protocols, domains, and applications. Better connection between devices leads to better services in domain of use. The devices will enable automation in almost any environment[3]. "The Internet of Things is defined as a global network of interconnected smart devices/objects, which provide the possibility of mutual communication and communication with the environment, exchanging data collected from the environment, while initiating processes to reactions caused by the state of the environment can be realized with or without direct connection with a man"[5]. The advantages of IoT are: quality communication, automation, and control, easy access to information, saving time and money, efficiency, and a better quality of life[3]. "Quantitative data is based on counting or measuring. The numerical scale used to produce the figures forms the basis of the analysis of quantitative data[6]". A trend is overall movements of the time series values'. All the values will not necessarily appear along a smooth curve/line, i.e., there will most likely be some variations – a little higher or lower. One of the initial concerns is to discover the overall tendency of the data. In most instances only a linear trend is considered, but the trend may be non-linear [7].

The first tool is adding a trend line to a data series in a chart to show the trend, or direction, of the data in the series. Trend lines can be added to scatter charts. The first step in creating a trend line is to select the data series on the chart the trend line is to be associated. Then choose the trend line command from the insert menu. When Select the type of trend line from the type tab and choose the options from the options table. Types of trend lines are: linear, logarithmic, polynomial, power, exponential [8].

Trend is a method of forecasting variations in a dependent variable, Y, employing one independent variable, X. When graphed the linear factor is used to determine if there is a increase or decrease in the dependent variable, Y, as the independent variable, X, increases[9]. Logarithmic creates the trend line using equation $y = c\ lnx + b$. Polynomial creates the trend line using the polynomial equation $y = b + cx$. Power creates the trend line using the power equation $y = c * x^b$. Exponential creates the trend line using the exponential equation $y = c * e^{bx}$ [8]. Figure 1 shows the number of connected IoT devices, past status, current, and forecast until 2025 where constant growth is seen. Growth is projected at 11.6 billion in 2021.

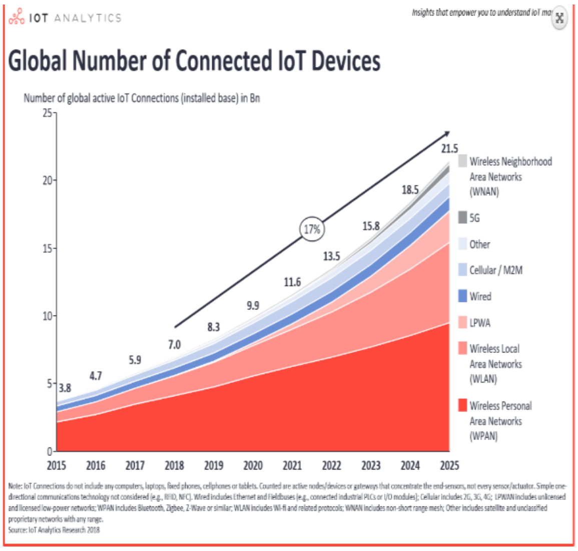


Figure 1. Number of connected IoT devices [10]

III. Forecasting with Linear Trend

A trend is a component that represents the basic direction of development over time and does not repeat itself within the time frame in which the data is repeated. The trend is actually through the average points through which the observed phenomenon would pass if there were no periodic phenomena. It is a component that represents the basis for predicting the future behavior of the observed phenomenon. Trend analysis is a technique that is based on the analysis of data in a certain period of time with the aim of discovering a trend that will enable forecasting in the future. Trends are usually modeled by mathematical equations, which contain a parameter by which various influences are represented. The linear trend represents a rectilinear movement of values and variables that represent the observed concept[6]. Table 1 presents number of IoT connections and forecast in billion.

TABLE 1. NUMBER OF IOT AND FORECAST

Number of IoT and forecast	Billion
2015	15,41
2016	17,68
2017	20,35
2018	23,14
2019	26,66
2020	30,73
2021	35,82
2022	42,52
2023	51,11
2024	62,12
2025	75,00

It is important to note that there are trend characteristics. Y_t denotes the value of the phenomenon at the observed moment t, and is the coefficient of the direction of the trend line, b is the segment on the axis where the values of the phenomenon are measured, e_T is the error or noise at the observed moment t. Assuming that there are n samples of the observed phenomenon in the observed time interval, regardless of the labels located on the time axis, labels in the form of ordinal numbers are often introduced so that the sum of these numbers is 0. The zero element is in the middle of the sample. When the noise is neglected, the equation of the linear trend $yt = at + b$ is obtained, where $a = \sum xy / \sum x^2$ [6] If the number of elements in the sample is odd then the middle element in the sample is declared zero, and if it is even then two values are assigned values -1.5 and 1.5, and the difference between adjacent elements remains 1. After that, the reliability of the assessment was determined by determining the coefficient of determination for the values taken for analysis. It has values in the range between [0, 1]. If it is closer to number one, then the estimate is more reliable[6]. The R^2 value is called the coefficient of determination and its value represents the fraction of the overall variance of the 'dependent' variable that is explained by the 'independent' variable. It is calculated from the sum of the squares of the residuals and the sum of the squares of regression[12]. By clicking on the selection field at the bottom of the menu in excel 'Display Equation on chart', you can see which function was used to approximate the data. Once the equation describing the trend for the observed sample is determined, it is possible to determine the future values of the observed variable. This is possible using the built-in forecast or trend functions, or by applying equations with coefficients 4 and 5[6]. The general form of the equation by which the approximation is performed is $y = ax + b$[6]. Based on table number 1, where you can see the number of internet connection items from 2015 to 2018, and how the author from the source stated the forecasts for the years from 2019 to 2025.

Based on the data from 2015, 2016 and 2017, a prediction for 2021 was made, with the help of the linear trend function as well as with the help of the already mentioned built-in functions in excel which give approximately the same result. The figure number 3 shows the trend line in the form of a linear equation. The table in figure 2 has the variables shown in the graph in figure 3. X and Y are used from table in figure 2.

IoT connection s iand forecast	bilion	x	y	x^2	x*y
2015	15.41	-1.00	15.41	1.00	-15.41
2016	17.68	0.00	17.68	0.00	0
2017	20.35	1.00	20.35	1.00	20.35
suma		0.00	53.44	2.00	4.94
		a=	2.47		
		b=	17.81333		
	Trend linija y=2.47x+17.8133y				

Figure 2. Function linear trend

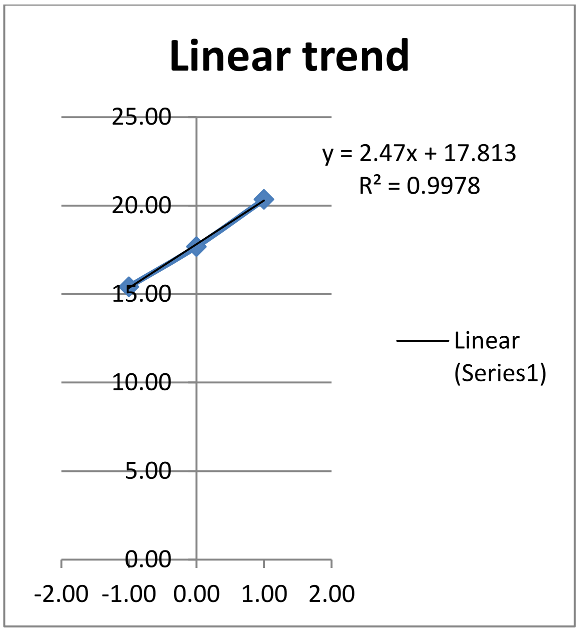


Figure 3. Graphic representation of a linear trend

IoT connection s iand forecast	bilion	x	y			
2015	15.41	-1.00	15.41			
2016	17.68	0.00	17.68			
2017	20.35	1.00	20.35			
2018		2.00				
2019		3.00				
2020		4.00		linear function	FORECAST	TREND
2021		5.00		30.1633	30.16333	30.16333

Figure 4. Prediction using linear function and built-in forecast and trend functions

Based on this figure, where you can see the pre-movement for 2021 (30,163) and when compared with the data from 2021, which was found from the source and is 35.82, it is concluded that it is not the same amount because this prediction was made without considering external factors, while the prediction in the literature was done in a different, probably more complex way, where some external factors were probably taken into account. It should be emphasized that no method is 100% reliable.

IV. Forecasting with non linear functions

Excel calculates a trend based on several mathematical functions. Each of these types is best applied to:

- *Exponential* – the exponential function approximation is performed when the data series increases or decreases with increasing acceleration with each subsequent data.
- *Linear* – linear approximation is best applied when the data series behave in a linear way: linearly decreasing or linearly increasing. The general form of the equation by which the approximation is performed is $y = ax + b$.
- *Logarithmic* – a logarithmic approximation when there is a series of data that first records rapid growth and then grows slower and slower.
- *Polynomial* – this approximation is best used when there is a series with fluctuating data. By clicking on Order, the order of the equation can be changed, which changes the general form of the function. For example, the general form of the equation for the second-order approximation is $y = ax2 + bx + c$, while for the third-order approximation $y = ax3 + bx2 + cx + d$, etc. Also, here, it is possible to use the Set Intercept option to influence the value of the last term of the equation.
- *Power*– we perform the quadratic function approximation when we have data in a series that grows at a regular pace[13].

In figures 5, 6, 7 and 8 the trend of the observed phenomenon is shown, where the following types of trend lines are in question: exponential, logarithmic, polynomial, and power. The number of IoT connections is best shown using the power type of the trend line because the coefficient of determination is 1, which means that it is a really reliable display, and it can be seen that the trend line follows the points on the graph.

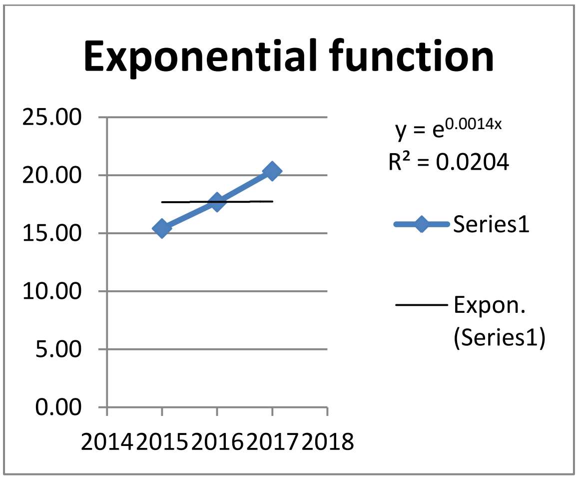


Figure 5. Exponential function

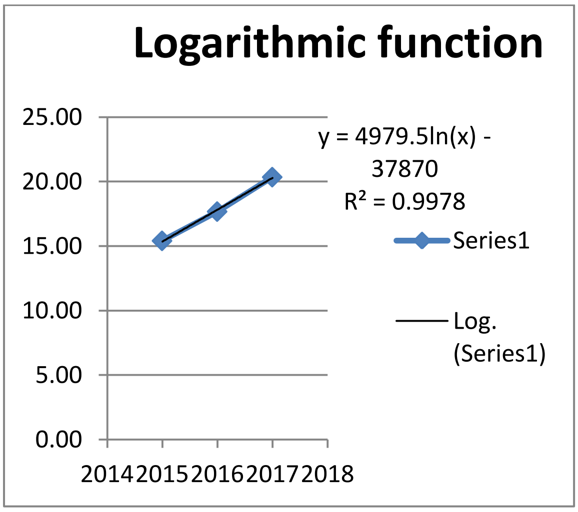


Figure 6. Logarithmic function

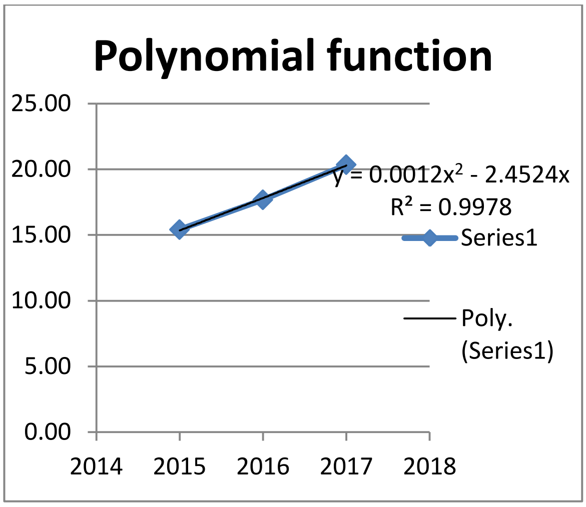


Figure 7. Polynomial function

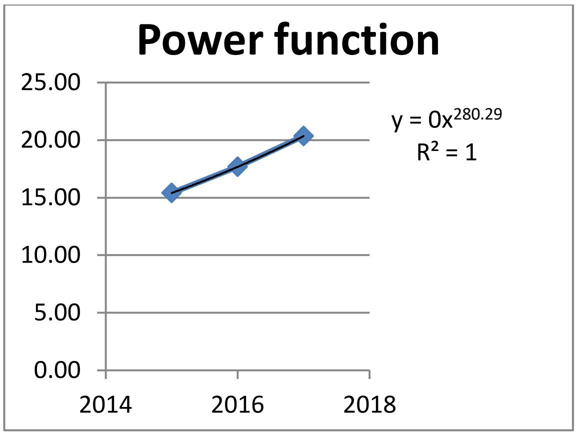


Figure 8. Power function

V. CONCLUSION

It is important to note that the noise is crucial. Noise is an unpredictable component and depends on external factors that affect the observed phenomenon. With the help of a linear trend, it is possible to show what the amount would be in the future without the influence of external factors, based on the existing data and their movements. It was found that in the observed period, the variables have a tendency to grow and it is best in this case to present the trend of the phenomenon using the power function type trend line. Also, the coefficient of determination is 1, which means that it is the most reliable of all other types.

REFERENCES

[1] Vujović V., Maksimoović M., Balotić G., Mlinarević P., Internet stvari – tehnički i ekonomski aspekti primjene, Infoteh – Jahorina, 14.03.2015.

[2] http://www.hrks.hr/skole/15_ljetna_skola/82.pdf Review date: 12.09.2020.

[3] Živković S., Primena internet stvari u različitim životnim okruženjima, Zbornik radova Fakulteta tehničkih nauka, Novi Sad

[4] Stevanović M., Đurđević D., Internet stvari, lična i materijalna bezbednost, Bezbednost, 13.02.2017.

[5] https://infoteh.etf.ues.rs.ba/zbornik/2015/radovi/RSS-4/RSS-4-3.pdf

[6] John Buglear, STATS Means Business a guide to business statistics, 2001.,

[7] https://www.sciencedirect.com/science/article/pii/S0169260700001243 Review date: 10.09.2020.

[8] Nelson L., Emmons, Jr., Mathematical modeling using Microsoft excel, 1997., Monterey, Kalifornija

[9] https://www.researchgate.net/profile/John_Kros/publication/27826995_Forecasting_with_Excel_Suggestions_for_Managers/links/54c10fe50cf28a6324a54ace.pdf Reveiwe Date: 12.09.2020.

[10] https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/ Review date 15.09.2020.

[11] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/

[12] Angus M. Brown, A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet, Computer Method and Programs in Biomedicine, 2000., Washington

[13] https://excelk.com/sr/linija-trenda/ Review date: 26.09.2020

APPENDIX

Conference Chairs

Dalibor Dobrilović, *Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

co-chairman Evgeny Cherkashin, *Institute of High Technologies, Irkutsk,Russia*

co-chairman Andrijana Bocevska, *Faculty of Information and Communication Technologies - Bitola, North Macedonia*

Organizing Committee

Dalibor Dobrilović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dragica Radosav, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Miodrag Ivković, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Biljana Radulović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Ivana Berković, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Duško Letić, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Branko Markoski, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Željko Stojanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Ljubica Kazi, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Eleonora Brtka, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vladimir Brtka, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Višnja Ognjenović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Zoltan Kazi, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Siniša Mihajlović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Velibor Premčevski, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Aleksandra Stojkov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dijana Karuović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vladimir Šinik, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Nadežda Ljubojev, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dušanka Milanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vladimir Karuović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Bojan Vujanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Evgeny Cherkashin, *Institute of System Dynamic and Control Theory SB RAS, Russia*

Anastasia Popova, *Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences*

Andriana Bočevska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Blagoj Ristevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Natasa Blazeska-Tabakovska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Tome Dimovski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Nikola Rendevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Maja Gaborov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Milica Mazalica, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Igor Vecštejn, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Marko Blažić, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Filip Tsvetanov, *South-west University "Neophyte Rilsky", Faculty of Engineering, Blagoevgrad, Bulgaria*

Goran Gecin, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Aleksandar Jašić, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Program Committee

Dalibor Dobrilović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia - president*

Ljubica Kazi, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dragica Radosav, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dragana Glušac, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Borislav Odadžić, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Duško Letić, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Miodrag Ivković, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Biljana Radulović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Ivana Berković, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Ljubica Kazi, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vladimir Brtka, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Branko Markoski, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Željko Stojanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Ilija Hristoski, *Faculty of Economics - Prilep, North Macedonia*

Zoltan Kazi, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Višnja Ognjenović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Eleonora Brtka, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Jelena Stojanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vesna Makitan, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Nadežda Ljubojev, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dijana Karuović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vladimir Šinik, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dušanka Milanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Igor Nedelkovski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Aleksandar Markovski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Violeta Manevska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Pece Mitrevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Ilija Jolevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Dragan Gruevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Konstadina Veljanovska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Sonja Mančevska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Snežana Savoska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Ramona Markoska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Mimoza Bogdanoska-Jovanovska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Božidar Milenkovski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Zoran Kotevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Nikola Rendevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Andrijana Bocevska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Blagoj Ristevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Monika Markovska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Tome Dimovski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Marina Blažekovik Toševski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Nataša Blažeska Tabakovska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Lela Ivanovska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Andrey Gachenko, *Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences. Irkutsk, Russia*

Andrey Mikhailov, *Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences. Irkutsk, Russia*

Anastasia Popova, *Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences. Irkutsk, Russia*

Alexey Daneev, *Irkutsk State Transport University, Irkutsk, Russia*

Denis Sidorov, *Melentiev Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia*

Viacheslav Paramonov, *Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia*

Andrey Dorofeev, *Institute of High Technologies, Irkutsk National Research Technical University, Irkutsk, Russia*

Gogolák László, *Subotica Tech - College of Applied Sciences, Subotica, Serbia*

Zlatko Čović, *Subotica Tech - College of Applied Sciences, Department of Informatics, Subotica, Serbia*

Zora Konjović, *University Singidunum, Centar Novi Sad, Serbia*

Siniša Nešković, *Faculty of organizational sciences, University of Belgrade, Serbia*

Nataša Gospić, *Faculty of transport and traffic engineering, Belgrade, Serbia*

Željen Trpovski, *Faculty of technical Sciences, Novi Sad, Serbia*

Branimir Đorđević, *Megatrend University, Belgrade, Serbia*

Slobodan Jovanović, *Faculty of Information Technology, Belgrade, Serbia*

Željko Eremić, *College of Technical Sciences - Zrenjanin, Serbia*

Rajnai Zoltán, *Obuda University, Budapest, Hungary*

Mirjana Pejic Bach, *University of Zagreb, Croatia*

Androklis Mavridis, *Aristotel University of Thessaloniki, Greece*

Madhusudan Bhatt, *R.D. National College, University of Mumbai, India*

Amar Kansara, *Parth Systems LTD, Navsari, Gujarat, India*

Narendra Chotaliya, *H. & H.B. Kotak Institute of Science, Rajkot, Gujarat, India*

Zeljko Jungic, *ETF, University of Banja Luka, Bosnia and Hercegovina*

Saso Tamazic, *Univerisity of Ljubljana, Slovenia*

Marijana Brtka, Centro de Matemática, *Computação e Cognição, Universidade Federal do ABC, São Paulo, Brazil*

Zoran Cosic, *Statheros, Split, Croatia*

Istvan Matijevics, *Institute of Informatics, University of Szeged, Hungary*

Slobodan Lubura, *Faculty of electrical engineering, University of East Sarajevo, Republic of Srpska, Bosnia and Hercegovina*

Edit Boral, *ASA College, New York, NY, USA*

Dana Petcu, *West University of Timisoara, Romania*

Marius Marcu, *"Politehnica" University of Timisoara, Romania*

Aleksej Stevanov, *South-west University "Neophyte Rilsky", Faculty of Engineering, Blagoevgrad, Bulgaria*

Petar Apostolov, *South-west University "Neophyte Rilsky", Faculty of Engineering, Blagoevgrad, Bulgaria*

Filip Tsvetanov, *South-west University "Neophyte Rilsky", Faculty of Engineering, Blagoevgrad, Bulgaria*

Francesco Flammini, *School of Innovation, Design and Engineering, Division of Product Realisation, Mälardalen University, Eskilstuna, Sweden*

Deepak Chahal, *Jagan Institute of Management Studies (JIMS, Rohini Sector-5), New Delhi, India*

Abdel-Badeeh M. Salem, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

Dragan Peraković, *University of Zagreb, Faculty of Transport and Traffic Sciences, Croatia*

Gordana Jotanović, *University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Herzegovina*

Goran Jauševac, *University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Herzegovina*

Invited Speakers Biographies

Alexander Feoktistov received the Ph.D. degree from Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences (ISDCT SB RAS) in 2000. He is currently a Head of the Laboratory of Parallel and Distributed Computing Systems of ISDCT SB RAS, and a professor in the Graduate school in ISDCT SB RAS. He leads a number of national research projects. His main interests include computational models, applied software packages, knowledge engineering, distributed computing, multi-agent technologies, and simulation modeling.

Sergei Gorsky received the Ph.D. degree from the Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the Russian Academy of Sciences (ISDCT SB RAS) in 2008. He is currently a Research Officer in the Laboratory of Parallel and Distributed Computing Systems of ISDCT SB RAS. He participates in a number of national research projects. His main interests include parallel and distributed programming, mathematical modeling, solving of Boolean Equations, and Service-Oriented Programming.

Invited Speakers Biographies

Dr Tünde Anna Kovács is an Associate Professor in the Department of Materials Technology of the Óbuda University, Hungary. Member of the editorial board of the Acta Materialia Transylvania, Safety and Security Sciences Review and Security Engineering of Anthropogenic Objects. Author and co-author of 109 publications with around 150 citations. Her research interests are in the field of the materials science and technologies, special welding processes (ultrasonic and explosive welding). Welder robots and collaborative welding robotics. Dr Kovács has h-index 4 in Web of Science and 7 in Google Scholar.

Dr Zoltán Nyikes is an Associate Professor in the Department of Methodology and Informatics of the Milton Friedman University, Hungary. He is a Computer Science Engineer, Dipl. Security Engineer, Doctor of Military Engineering, University Lecturer, Member of the Public Body of the Hungarian Academy of Sciences. Author and co-author of 65 publications with around 110 citations. His research focuses on INDUSTRY 4.0, user safety awareness and digital competence, the security of information transmission, buildings protection coatings innovation against explosive loads and safety of rail transport. Researcher on several research and development projects. Supervisor of engineering thesis works and diploma works. Dr Nyikes has h-index 3 in Scopus and 6 in Google Scholar.

Abdallah Kafi MSc graduated Mechatronics Engineer on Óbuda University Banki Donát Faculty of Mechanical Engineering and Safety Engineering, Budapest (Hungary). He focused on his research on the robotics and the robot's application in the welding process. Currently PhD student on Doctoral School on Safety and Security Sciences, Óbuda University University. In his doctoral thesis work, he is continuously working on the application of the collaborative robot for industrial working.

Invited Speakers Biographies

Dr. Andrej Šmid was born in Maribor, Slovenia, in 1971. He attended primary and secondary school in Maribor, Slovenia and then finished his architectural studies in Graz, Austria in 1998.

Since 1993 he was working in local architectural offices and after graduation in the local urban planning offices until 2001 when he started his architectural career at the Komunaprojekt office in Maribor. After the begin of the 2013 economic crisis he founded his own architectural office in the Studio Baza group and an architectural firm Arhitekt Šmid in 2016. His work in architecture and urban planning is very divergent, but oriented to all the urban detailed planning problems and the architectural design of mainly non residential buildings. The work of the Maribor - based office includes designs in all the states of central Europe, from Northern Macedonia to the Netherlands.

His academic career started in 2010 when he started working as a guest lecturer on the Architectural department of the Civil engineering Faculty of the Maribor University. The various studies and research on urban structures in former Yugoslavia led to the PhD completion in 2019 on the Faculty of Architecture at the Graz University of Technology (TU Graz). At this point he is employed as independent architect and also as a co-worker at the Faculty for civil engineering, traffic engineering and architecture of the Maribor University.

He is a father of two girls and two boys, a devoted chef, a certified skipper and reliable construction worker.

List of Reviewers

Željko Stojanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vladimir Brtka, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Eleonora Brtka, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vladimir Šinik, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Ivana Berković, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dalibor Dobrilović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Ljubica Kazi, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Vesna Makitan, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Zoltan Kazi, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Višnja Ognjenović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Dijana Karuović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Tome Dimovski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Evgeny Cherkashin, *Institute of High Technologies, Irkutsk,Russia*

Deepak Chahal, *Jagan Institute of Management Studies (JIMS, Rohini Sector-5), New Delhi, India*

Zoran Kotevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Blagoj Ristevski, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Biljana Radulović, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Duško Letić, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Jelena Stojanov, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Goran Jauševac, *University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Herzegovina*

Gordana Jotanović, *University of East Sarajevo, Faculty of Transport and Traffic Engineering, Doboj, Republic of Srpska, Bosnia and Herzegovina*

Dragana Glušac, *University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia*

Snežana Savoska, *Faculty of Information and Communication Technologies, University "Sv. Kliment Ohridski", Bitola, North Macedonia*

Tibor Szakall, *Subotica Tech - College of Applied Sciences, Department of Informatics, Subotica, Serbia*

Zlatko Čović, *Subotica Tech - College of Applied Sciences, Department of Informatics, Subotica, Serbia*

Andreja Samčović, *University of Belgrade, Faculty of Transport and Traffic Engineering, Belgrade, Serbia*

Dragan Peraković, *University of Zagreb, Faculty of Transport and Traffic Sciences, Croatia*

Index of Authors