# DESIGN OF DATA ANALYSIS SYSTEMS
# FOR BUSINESS PROCESS AUTOMATION

**Tomasz Rymarczyk**[1,2], **Tomasz Cieplak**[3], **Grzegorz Kłosowski**[3], **Paweł Rymarczyk**[1]

[1]Research and Development Center, Netrix S.A., Lublin, [2]University of Economics and Innovation in Lublin, [3]Lublin University of Technology, Faculty of Management, Department of Organization of Enterprise

*Abstract. The paper deals with the design of data analysis systems for business process automation. The main goal of the project is to develop an innovative system for analyzing multisource data, business data mining processes, and as a result the creation and sharing of new improved procedures and solutions.*

Keywords: Smart system, Multi-source Data, Agent Based Modelling

## PROJEKTOWANIE SYSTEMÓW ANALIZY DANYCH
## DO AUTOMATYZACJI PROCESÓW BIZNESOWYCH

*Streszczenie. Artykuł dotyczy projektowania systemów analizy danych do automatyzacji procesów biznesowych. Głównym celem projektu jest opracowanie innowacyjnego systemu do analizy danych wieloźródłowych, procesów eksploracji danych biznesowych, a co za tym idzie tworzenie i udostępnianie nowych ulepszonych procedur i rozwiązań.*

Słowa kluczowe: inteligentne systemy, dane wieloźródłowe, modelowanie agentowe

## Introduction

In this paper, the concept of an information system for analysis, preservation, optimization and generating innovative business data mining processes was presented. In modern and complex business world, companies must develop pioneering ways to distinguish themselves from other market players by becoming more cooperative, effective, precise, flexible and agile. They need to be capable to quickly reply to market requirements and alterations. Depending on the company's competitive advantage which may be a novelty, price, outstanding web content, or presence in social media, specific online strategies should be applied to achieve the desired market. Many companies have observed that the data they preserve and how they utilize it can build their market advantage. Data and information are turn into primary resources for many organizations [6]. For example, Walmart's databases located in cloud are estimated to contain overall more than 40 PB and process 2.5 PB of data every hour consisting information about customer behaviour and preferences, network and device activity, and market trends data [17]. Business models are based on data from external sources, data warehouses and web resources. They create a process base that is a resource for computational intelligence algorithms. The functionality of intelligent decision support and business process modelling is the management of strategic information, the analysis of information coming from the immediate and indirect business environment and influencing its strategic development. Traditional approach to systems development is concentrated on building applications, not services. Systems built that way are not flexible and are difficult to scale [19]. Nowadays, one of the major issues is that how to capture users' frequently changing needs and expectations and to support those with dynamic business processes. Moreover, analytical systems design must take into consideration that there are multiple channels of data that may need to be accessed and incorporated for further exploration. One of key characteristics of cloud is on-demand and secure access to scalable data exploration resources. Recently, main players of cloud-based services market introduced wide range of tools for data scientists. They enable analytical systems to aggregate a number of services with computing algorithms that are capable to provide dynamic solutions of data analysis tasks. Cloud-based analytical systems provide unlimited access to various data types and sources. It gives us an ability to process high volume and wide variety of data with high velocity and veracity, which means Big Data solutions.

The emergence of new technologies and tools of data collection and manipulation now opens areas of research on understanding customers-business relations and interactions. Conventional user interest modelling approaches are not created for integrating and studying interests from multiple sources, therefore, they are not very successful for obtaining comparatively complete description of user interests in the dispersed conditions. In this context data acquired from sources like social networks, digital journalism, mobile telephony, online gaming, online shopping, etc. empower business analysts to find more precise explanation of phenomena's driving contemporary business [5]. This process can be strongly empowered by AI methods. According to Gartner, Inc. report [4] for several next years main concern of IT companies will be creation of systems supported by artificial intelligence that can be applied in one form or another. Another important technological trend mentioned in the report is called "digital twins". This term occurs more frequently with progression of digitalization of our life but also digitalization of places, processes and "things". The process of "digital twins" development will offer productive space for new event-driven business processes and digitally enabled business models as well as ecosystems. Thanks to technological advancement, such systems will be able to provide a coherent set of components that will act as a living organism. In most cases, an intelligent system is able to percept and act to altering conditions as well as collect and deposit information in its memory and learn from prior experiences and knowledge. What is crucial, intelligent system is being able to adapt its actions to implement new tasks and achieve its pre-determined or developing objectives. Intelligent actions of those systems include listening and responding to participants, inspecting the markets, collecting and analysing data, creating and broadcasting knowledge, learning, and efficient decision making.

## 1. Smart system – design remarks

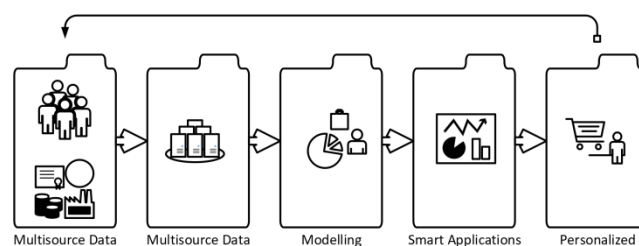A diagram of system's data flow is shown on Fig. 1.



Fig. 1. Multi-source smart application data flow pipeline

Proposed approach involves following steps: data sources determination (mandatory – user's feedback and reactions on propositions sent by the system) as well as representing key information (e.g., entities, relations, events, sentiments); accumulating information from multiple sources, identifying inter-

document relationships, and storing the information in a structured knowledge databases; exploring and filtering existing data including entities, relations, and events, building analytical models and then identifying system efficiency with appropriate simulation methods (e.g. agent based modelling); upon on selected model suited for user's needs, an adequate type of application will be served to system's user.

### A. Multi-source information integration

Perception of any particular state (e.g. purchase decision) involves incorporating many bits of information. In case of user interests, they are typically spread in various systems on the Web. This information has numerous attributes as a rule cases and is gained from multiple data sources within multiple time periods. Consumers' purchasing decision is stimulated by their knowledge about the brand, which includes innovativeness, impression of the brand, self-perceived brand value from web content, and e-WOM from social media [10]. State evaluators' understanding and preference will obviously affect the outcome of information integration, and therefore impact the understanding generated for a given state. Process of a multisource data analysis demands an accurate and precise data integration with evaluator's effect deprived. A learning-based information integration approach, which embeds the fuzzy least squares support vector machine (FLS-SVM) technique, was described in a research [15]. In the cited research, a state can be assessed through incorporating an inference acquired information and analyzing correlated data sources. In series of experiments it was shown that the suggested method has a precision learning capability depending on evaluators' skill in the information incorporation for producing awareness for a state. Another method of user interest modelling based on multisource personal information fusion and semantic reasoning was presented in a research [16]. Authors give different fusion strategies for interest data from multiple sources. Moreover, investigation on the semantic relationship between users' explicit interests and implicit interests by reasoning through concept granularity was presented. Authors gave illustrative examples based on multiple sources on the Web (e.g. microblog system Twitter, social network sites Facebook and LinkedIn, personal homepage, etc.) showing that offered approach is possibly effective.

### B. Simulation modelling

Leaving aside a diversity of analyzed data and their huge number the smart applications are still very complex to design. Based on the above-mentioned statements the application design can be based on interactions of system actors. Following assumptions of essential functionality are made: all actors can exchange information with each other to understand continuously changing social situation; outgoing motives are generated to help other actors to understand the preferences and important factors in the decision-making process; at the same time respond to incoming motives from other actors to update its own attitude to the process [18]. In such scenario many dynamics and discrete events take place. Understanding a particular system is difficult without appropriate modelling method – a way of solving problems that occur inside the real system. Mentioned system consists of a number of dynamic processes that occur between actors of the system. Also, the actor's state is dynamic and changes in time as well as under interaction with environment. Analytical or static modelling solution for dynamic system does not always exist or may be very hard to find [3]. Nowadays, multi-agent simulation is increasingly being promoted – and used – as a tool to study the dynamics in various kinds of systems where human behaviour plays a critical role [9]. That's why, as a best solution, Agent Based Modelling (ABM) can be taken under consideration. Application of ABM allows system optimization prior its implementation.

### C. Agent Based Modelling

Agent-based model (ABM) is a bottom-up methodology and has advantages in modelling complex systems. It has capability to symbolize the macro-level dynamics of a system by gathering individual behaviours and interactions among agents at the micro-level. ABM has been applied to engineering, sociology, economics, and management fields. Agent based modelling treats system as a collection of independent decision making objects so-called agents. Each agent separately evaluates its state and makes decisions on the foundation of a set of instructions. Agents may perform various actions suitable for the system they symbolise — for instance: producing, consuming, or selling. Recurring competitive relations among agents are a characteristic of agent-based modelling, which depend on the power of computers to investigate dynamics without application of pure mathematical methods [2].

Upon the research paradigm of study depicted on Fig. 2, a simplified agent-based model to study factors affecting consumers' adoption decision in multiple brands competition was developed. The modelling solution was based on a framework developed by Wander Jager. Understanding these decisions allows one's to say which factors ultimately affects brand's market share. The model simplification limits factors to a self-perceived value (price, rating, quality), a self-perceived utility from online information, e-WOM and social media [8]. In case of the self-perceived value, buyers often evaluate the value of a product by searching its online information, which includes price, online rating, quality ranking, etc. [12]. Wordof-mouth (WOM) is an important aspect that can influence consumers' opinion of a product and promote product diffusion [1].

Though, typical WOM is limited to social communication margins, and the influence weakens quickly over time and space [11]. The online review (e-WOM) performs a more significant role in consumers' purchase decision as the online groups are becoming more widespread and persistent. Buyers can easily retrieve the online evaluations in social network to reach first-hand data about the product [22]. The empirical investigation works also implies that there occurs a positive association between e-WOM and sales yield.
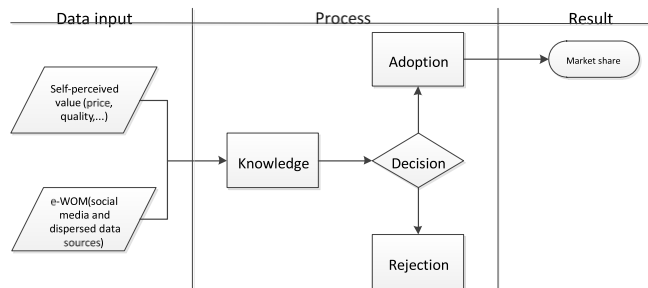


*Fig. 2. Simplified model of factors affecting consumers' adoption decision in multiple brands competition*

### D. Simulation study

The agent-based model gives possibility of system observation, where consumers make decisions and adopt a brand upon its price, services, products quality and information taken from WOM and e-WOM sources. Often, customers may lack necessary product data, even after discussing with off-line groups. Often, customers may face a lack of necessary product data, even after discussing with off-line groups (WOM). They possibly will enter an online society to investigate relevant information, which can decrease an uncertainty and prevent incongruities between presumed and actual product functioning. Therefore, e-WOM communications are likely to have strongly effect opinion and purchase intent when customers meet new goods [7]. On the other hand, the more e-WOM is being searched, the less it is being utilized in the final purchase decision [20]. The online reviewing

———— IAPGOŚ 3/2018 ————

system can act as an amplifier of information received through WOM, but also if e-WOM is also negative, it will have strong negative effect on customers' purchase decisions [14]. Taking into account information given by Liu [13], in the particular simulation model the scale-free network was employed to represent the interindividual interactions in the population to provide diversity of individuals in searching and information dissemination ability.

The calibrated consumer agent population to lead several simulations was used. Simulations were carried out to test market shares evolutions with a variant e-WOM level of information and different levels of perception of feedback (influenced by a number of searches). The following section reports these simulations and interpretations of the obtained results.

The interface of simulation system is shown in Fig. 3. All computational experiments were conducted in the same environment. The proof-of-concept of investigated approach was programmed in NetLogo 6.0.2 environment [21] on standard laptop (Intel core i7-5500U, CPU 2.4 GHz, and 16 GB of RAM), with Microsoft Windows 10 operating system.
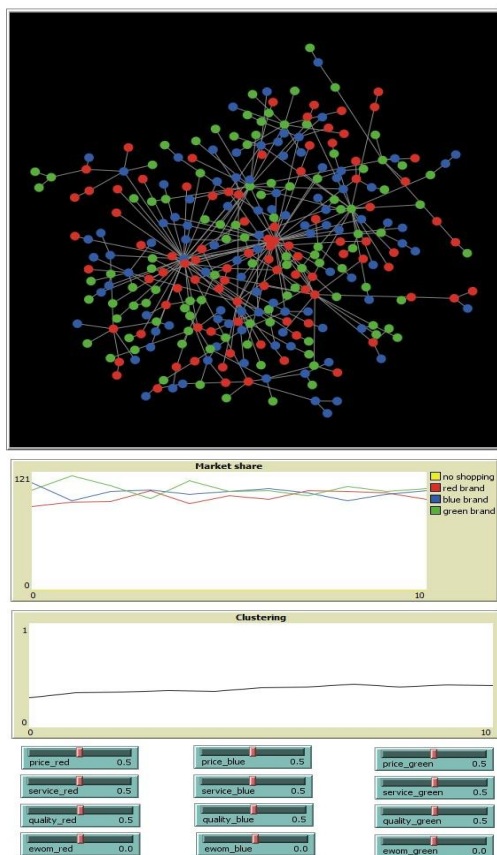


*Fig. 3. Interface of the simulation system with depicted equilibrium state of customers' purchase intents*

E. Simulations results

The simulation was started by creating a consumer agent population whose individuals have initially set behavioural attitudes intensities, like a preference of product, an ability to receive information from other agents, a level of market exploration. Then agents are gathered together establishing a scale-free network. Two experimentation contexts were tested. The first one was carried without an influence of e-WON on customers' purchase instances, the second included that influence. In both cases simulation contained a population of 300 consumer agents within a virtual market including 3 competing brands. Main parameters ranges of brands and customers agents are shown in Table I. When the simulation progress advances one "tick", consumers will make the purchase decision and trigger the choice event based on their purchase requirements (generated randomly Poisson-distributed integer values).

Consumers can retrieve brand information in WOM, and in the second research' scenario, the off-line WOM is amplified by the influence triggered by e-WOM. After the simulation process starts, the simulation system will run continuously until one hundred of repetition condition is met. In the following three cases of results relative to the evolution of market shares in the virtual market are presented. In the first experiment an initial market situation dominated by one brand was considered. In this case "Red Brand" overcome other brands by a mean of better price in comparison to competitors. The difference was set in an instance of price parameter. For "Red Brand" it was set to level 0.9, were for other brands the price level was 0.5. In such conditions, the mean market share was 0.44% in case of "Red Brand" and other two took almost the same position with 0.28% for each. The summary statistics of the simulation results of case 1 are shown in Table II.

*Table I. Agents parameters ranges*

|  | Parameter | Min | Max |
|---|---|---|---|
| Brand | Price | 0 | 1 |
|  | Services | 0 | 1 |
|  | Quality | 0 | 1 |
|  | e-WOM | -0.5 | 0.5 |
| Customer | WOM influence | 0 | 1 |
|  | Perception ability | 0 | 1 |
|  | e-WOM search level | 1 | 2 |

*Table II. Descriptive statistics of first case of first experimentation context*

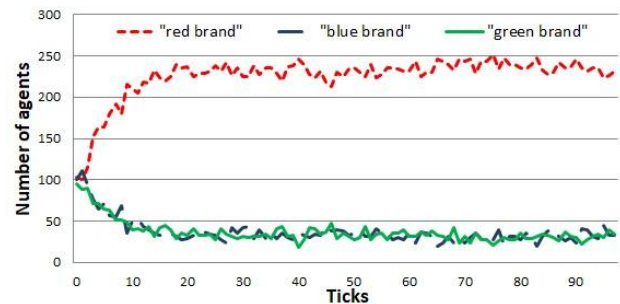| Brand | Min | Max | Mean | SD |
|---|---|---|---|---|
| Red | 0.34 | 0.53 | 0.44 | 0,04 |
| Blue | 0.22 | 0.36 | 0.28 | 0,03 |
| Green | 0.20 | 0.34 | 0.28 | 0,03 |



*Fig. 4. Interface of the simulation system with depicted equilibrium state of customers' purchase intents*

Following two cases describe the influence of the positive e-WOM on the market share. This is the aim of the research to understand how to eliminate information noise to increase the positive e-WOM influence on brand's position on a market. With the first e-WOM influence case, "Red Brand" has a well-developed system of online reviews and opinions about its product. Users were able to find quickly corresponding opinions, so the e-WOM search level was low and equalled 1.0 and e-WOM amplification was set to 0.1. "Blue Brand" has in the same time no information and no positive influence of e-WON on buyers' decisions, trying to conquer the market with lower price level competing with "Red" adversary. With a low diffusion coefficient of the information about "Blue" and "Green" brands over the market participants, "Red Brand" quickly gets almost 75% of market share. This phenomenon is depicted on Fig. 4.

The situation has opposite course when "Red Brand" information quality spread around online community become worse comparing to the previous state. Then users must spend more time and have to search through many more reviews to find sufficient and satisfactory enough information about brand's products. On the other hand, "Blue Brand" is developing much better services and product quality with lower price in contrast to "Red Brand". This situation is represented by the model setting: "Red Brand": price 0.5, services 0.5, quality 0.5, eWOM: 0.1,

e-WOM search level: 2; "Blue Brand": price 0.9, services 0.9, quality 0.9, e-WOM: 0.0, e-WOM search level: 2. Due to that users can't rely on the brand's information collected from Internet, they need to develop a self-perceived value of brands. So, that's why after some time needed for the user to investigate the market "Blue brand" overcome "Red Brand". This can be observed on Fig. 5.
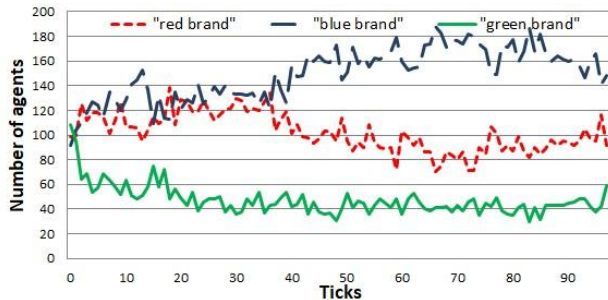


*Fig. 5. Interface of the simulation system with depicted equilibrium state of customers' purchase intents*

## 2. Conclusion

A goal for designed system is better market penetration and long-tail marketing, targeted and personalized recommendation, increased sale and customer satisfaction. To reach this aim, the influence of the information spread over the internet in the form of product's users' opinions, reviews or recommendations on customer purchase intents were investigated. The investigation shows that well developed and reliable e-WOM has very high influence on customer's decisions.

Systems that will be developed in future will be based on research areas containing five critical technical areas – big data sets analytics, text analytics, web analytics, network analytics, and mobile analytics. Characteristics of user's data can be described as follows: structured web-based, user-generated content, rich network information, unstructured informal customer opinions. Data sources identification gives a potential way for application of different analytical methods like: association rule mining, database segmentation and clustering, anomaly detection, graph mining, social network analysis, text and web analytics, sentiment and affect analysis. Introduction of multi-source, multi-user data analysis allows to develop or reinvent systems for social media monitoring, crowd-sourcing data gathering and monitoring, social and virtual games and better recommender systems.

## References

[1] Baker A.M., Donthu N., Kumar V.: Investigating how word-ofmouth conversations about brands influence purchase and retransmission intentions. Journal of Marketing Research 53(2)/2016, 225–239.

[2] Bonabeau E.: Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 99(3)/2002, 7280–7287.

[3] Borshchev A., Filippov A.: From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. Proceedings of the 22nd international conference of the system dynamics society 22. Citeseer, 2004.

[4] Cearley D.W., Burke B., Searle S., Walker M. J.: Top 10 strategic technology trends for 2018. https://www.gartner.com/doc/3811368/top–strategic-technology-trends

[5] Chang R. M., Kauffman R. J., Kwon Y.: Understanding the paradigm shift to computational social science in the presence of big data 63/2014, 67–80.

[6] Demirkan H., Delen D.: Leveraging the capabilities of service oriented decision support systems. Putting analytics and big data in cloud 55(1)/2013, 412–421.

[7] Huang J.-H., Hsiao T.-T., Chen Y.-F.: The effects of electronic word of mouth on product judgment and choice: The moderating role of the sense of virtual community. Journal of Applied Social Psychology 42(9)/2012, 2326–2347.

[8] Hudson S., Huang L., Roth M.S., Madden T.J.: The influence of social media interactions on consumer–brand relationships: A three country study of brand perceptions and marketing behaviors. International Journal of Research in Marketing 33(1)/2016, 27–41.

[9] Jager W.: Simulating consumer behaviour: a perspective. Rapport in het kort, 2006.

[10] Krizanová A., Stefániková L.: Importance of the brand for consumer purchasing decision in the slovak republic. Verslo Sistemos ir Ekonomika 2(2)/2012.

[11] Li F., Du T.C.: The effectiveness of word of mouth in offline and online social networks. Expert Systems with Applications 88/2017, 338–351.

[12] Lien C.-H., Wen M.-J., Huang L.-C., Wu K.-L.: Online hotel booking: The effects of brand image, price, trust and value on purchase intentions. Asia Pacific Management Review 20(4)/2015, 210–218.

[13] Liu C., Du W.-B., Wang W.-X.: Particle swarm optimization with scale-free interactions PloS one, 9(5)/2014, e97822.

[14] López M., Sicilia M.: Determinants of e-wom influence: the role of consumers' internet experience. Journal of theoretical and applied electronic commerce research 9(1)/2014, 28–43.

[15] Lu J., Yang X., Zhang G.: Support vector machine-based multisource multi-attribute information integration for situation assessment. Expert Systems with Applications 34(2)/2008, 1333–1340.

[16] Ma Y., Zeng Y., Ren X., Zhong N.: User interests modeling based on multi-source personal information fusion and semantic reasoning. International Conference on Active Media Technology, Springer, 2011, 195–205.

[17] Marr B.: Really big data at walmart: Real-time insights from their 40+ petabyte data cloud. https://www.forbes.com/sites/bernardmarr/2017/01/23/really-bigdata-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud

[18] Pan A., Choi T.-M.: An agent-based negotiation model on price and delivery date in a fashion supply chain 242(2)/2014, 529–557.

[19] Polakowski K., Filipowicz S., Sikora J., Rymarczyk T.: Tomography Technology Application for Workflows of Gases Monitoring in The Automotive Systems. Przegląd Elektrotechniczny 84(12)/2008, 227–229.

[20] Pöyry E., Parvinen P., Salo J., Blakaj H., Tiainen O.: Online information search and utilization of electronic word-of-mouth. Proceedings of the 13th International Conference on Electronic Commerce ACM 2011, 2.

[21] Wilensky U.: Netlogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston.

[22] Yan Q., Wu S., Wang L., Wu P., Chen H., Wei G.: E-wom from e-commerce websites and social media: Which will consumers adopt? Electronic Commerce Research and Applications 17/2016, 62–73.

**Ph.D. Eng. Tomasz Rymarczyk**
e-mail: tomasz@rymarczyk.com

He is the director in Research and Development Center in Netrix S.A. and the director of the Institute of Computer Science and Innovative Technologies in the University of Economics and Innovation, Lublin, Poland.. He worked in many companies and institutes developing innovative projects and managing teams of employees. His research area focuses on the application of non-invasive imaging techniques, electrical tomography, image reconstruction, numerical modelling, image processing and analysis, process tomography, software engineering, knowledge engineering, artificial intelligence and computer measurement systems.

**Ph.D. Tomasz Cieplak**
e-mail: t.cieplak@pollub.pl

Doctor of Engineering, acting assistant professor in the Department of Enterprise Organization in the Faculty of Management, Lublin University of Technology. He has extensive experience in preparing and conducting lectures and training. What's more, he has research experience gained in the prepared doctoral dissertation at Saga University, Japan. The area of research interest is the use model of cloud computing in the enterprise, creating systems based on SOA and the cloud computing model.

**Ph.D. Grzegorz Kłosowski**
e-mail: g.klosowski@pollub.pl

Assistant Professor in Department of Organization of Enterprise at the Faculty of Management of Lublin University of Technology. The research interests of the author include artificial intelligence, simulation and modeling of engineering and business processes. Leader and participant in several implementation projects. Chairman of the Board of the Special Purpose Vehicle of POLLUB-Invest Ltd., whose purpose is indirect commercialization.

**M.Sc. Paweł Rymarczyk**
e-mail: pawel.rymarczyk@netrix.com.pl

In 2003, He graduated from IT studies at the University of Information Technology and Management in Rzeszów. In the years 2012-2016 he was a PhD student at the Electrotechnical Institute in Warsaw. Research interests are mainly focused on IT applications.