Information Systems Development and Applications

Edited by
Stanisław Wrycza

UNIWERSYTET GDAŃSKI

Faculty of Management
University of Gdańsk
Sopot 2015

www.wzr.pl

Reviewer
Prof., Ph.D. Hab. Eng. Celina M. Olszak

Cover and title page designed by
ESENCJA Sp. z o.o.

Technical Editor
Jerzy Toczek

Typeset by
Mariusz Szewczyk

ISBN 978-83-64669-06-4

Published by
Faculty of Management
University of Gdańsk
81-824 Sopot, ul. Armii Krajowej 101

Printed in Poland by
Zakład Poligrafii Uniwersytetu Gdańskiego
Sopot, ul. Armii Krajowej 119/121
tel. +48 58 523 13 75, +48 58 523 14 49, e-mail: poligraf@gnu.univ.gda.pl

Table of Contents

Preface

This monograph is the collection of the selected papers from Gdańsk EuroSymposium 2015 on SAND – Systems Analysis and Design. SAND is the classical field of research and education in the area of management information systems (MIS) or, as it is called more frequently in Europe – Business Informatics, almost from its origins.

The objective of the EuroSymposium on Systems Analysis and Design is to promote and develop high quality research on all issues related to SAND. It provides a forum for SAND researchers and practitioners in Europe and beyond to interact, collaborate, and develop their field.

Therefore, there were three organizers of the 8th EuroSymposium on Systems Analysis and Design:

- SIGSAND – Special Interest Group on Systems Analysis and Design of AIS,
- PLAIS – Polish Chapter of AIS,
- Department of Business Informatics of University of Gdansk, Poland.

This monograph contains 8 papers selected from the Eurosymposium 2015 submissions. They regard the following innovative topics:

- social media application for recruitment,
- the prerequisites of SysML use for furniture manufacturer,
- information security in public administration,
- the possibilities of emotions modeling in IT applications,
- the research of consumerization in IT companies,
- big data quality framework for unstructured data analysis,
- the formal frameworks for simulation in information systems,
- the assessment of Enterprise Architect CASE tool used in academic didactics.

The first chapter shows the areas of social media application and the benefits that an organization can be achieved by applying social media tools. The main goal of the chapter is to present the concept of social recruiting as a part of social business. The activities of recruiters in social networking sites are also evaluated.

The second chapter is presenting the basic principles and difficulties of application of different types of SysML diagrams including requirements diagrams, block data diagrams, use case diagrams, parametric diagrams, activity diagrams, sequence diagrams and state chart diagrams applied in the furniture firms.

The results of the research regarding the authorisation of users in IT systems, conducted in Polish self-government institutions is the area of interest of the third chapter. The survey is based on the interviewing the employees of these institutions. Research allowed to identify the key factors which influence the construction of weak passwords. It became the basis for an alternative authorisation method. Effectiveness of this method was verified and users were asked to assess the prototype in the form of a programme.

The fourth chapter regards emotional states representation and modeling for software systems, that deal with human influence. A review of emotion representation models is provided, including discrete, dimensional and componential models. The chapter provides also analysis of emotion models used in diverse types of affect-aware applications: games, mood trackers or tutoring systems. The analysis is supported with two design cases. The study allowed to reveal, which models are most intensively used in affect-aware applications as well as to identify the main challenge of mapping between the models.

The aim of the fifth chapter is to examine the trend of consumerization in Polish IT companies. Theoretical analysis of the phenomenon of consumerization and BYOD has been conducted. In addition, a survey has been done for the people responsible for IT in 56 Polish companies.

In the sixth chapter Big Data is perceived as an alternative way of conducting business and marketing analysis based on unstructured data analysis. The key issue of the chapter is to define the attributes of Big Data quality. Several dimensions of Big Data quality have been identified based on a case study performed in Big Data ecosystem.

The chapter seventh presents an ontological modeling tool that may serve as a visual editor for the DEMO methodology and simulator of business-processes. The whole design process is briefly described. The study is an attempt to fill the gap in conditions of the present deficiency of visual editors for DEMO with process modeling capacities.

Unified Theory of Acceptance and Use of Technology (UTAUT) is widely used for the assessment of acceptance factors in different fields of business and IT. A number of studies focus on university teaching with regard of technology acceptance as well. The last chapter is devoted to development of the original, enhanced UTAUT model that takes into account the specificity of Systems Analysis & Design discipline in order to verify CASE tool acceptance within university teaching process. Model introduced in the chapter was verified by carrying out a two-stage study among regular and extramural students of MIS.

It is expected that the knowledge presented in this monograph will inspire the readers in conducting their own research and encourage them to submit the papers for the next EuroSymposium.

Stanisław Wrycza

Chapter 1

Social recruiting – the use of social technologies in the area of recruitment

Dorota Buchnowska

Introduction

Great popularity of social media (SM) causes that more and more business executives regard social technologies (in particular social networks sites – SNS, microblogs and blogs) as a source of competitive advantage. They indicate many benefits among which the most important are: increasing speed of access to knowledge and local expertise, improved sharing of knowledge, increasing customer and employee satisfaction, lower costs and higher revenue [McKinsey& Company, 2009].

According to the global survey conducted by MIT Sloan Management Review and Deloitte [Kane et al., 2014], companies across industries are creating value with social business and 75% of respondents believe social business is important or somewhat important. The concept of social business refers to companies that use social media, social software and technology-based social networks to advance their business objectives [Walker, 2013] and "to enable more efficient, effective and mutually useful connections between people, information and assets" [IBM, 2013]. Authors of the study found that the advance of social business is a global trend.

The use and importance of social media tools is growing in any industry, any sector (B2C, B2B, C2C), any size of enterprise and also across all functional areas of company. Although social tools are used most for customer engagement, customer relationship management and executing marketing activities by employee from marketing, sales and IT departments [McKinsey&Company, 2015], more and more frequently they are used in other processes, such as new product development, talent management or recruiting. This paper shows the areas of social media application and the benefits that an organization can achieve by applying social media tools. The main goal of the article is to present the concept of social recruiting as a part of social business, the possibilities of using different SM tools and benefits of this. The purpose of this article is also to evaluate the activity of recruiters in social networking sites.

1.1. Social business – the use social media in the enterprise

The concept of social business refers to the companies where employees from each department use social media technologies to perform their duties. However, first of all SM are used in the field of customer relationship management (marketing, sales and customer services) [Buchnowska, 2013]. According to the data from Eurostat, 38% of European Union enterprises used social media for business in 2013 and almost 60% of them (22% of total) used SM for developing the enterprise's image or marketing products (Table 1.1). Larger companies make greater use of social media than smaller.

Table 1.1. Types of social media used and purposes of using social media by enterprises in 2013

Kind of activity on social media	% of total EU enterprises in a group			
	total	small	medium	large
Using social media:	38	37	42	55
– social networks (e.g. Facebook, LinkedIn, Xing, Yammer),	28	26	33	46
– enterprise's blogs and microblogs,	10	8	14	24
– multimedia content sharing websites,	11	10	16	27
– wiki based knowledge sharing tools	6	5	9	18
Developing the enterprise's image or marketing products (e.g. advertising products, goods, services)	22	21	29	41
Obtaining or responding to customer opinion, reviews	15	14	20	30
Involving customers in development or innovations of products (goods or services)	9	8	11	17
Collaborating with business partners or other organisations (e.g. public administration, NGO)	9	8	11	16
Recruiting employee	9	7	15	27

Source: Own elaboration based on [Eurostat 2013].

Management of customer relationships is the main but not the only way to use of social media in the enterprise. SM technology is used in such business areas as: research and development (R&D), supply chain management (SCM) and human resource management (HRM) [Buchnowska, 2013]. There are presented examples and benefits of social media application in different business areas in the Table 1.2.

Table 1.2. Examples and benefits of social media application in different areas

Examples of social media using		Benefits of SM using
Marketing	– corporation blog can be used for building reputation; – internal wiki can be used to share market knowledge that has been collected from conversations with customers; – viral advertising propagated online, encouraging word-of-mouth reference on video sharing websites; – messages to announce special offers and discounts on microblogging services; – spreading of advertising campaigns within the communities of Clients on Personal Social Networking Sites.	– reduced marketing expenses, – improving the company's image, – increasing exposure, – increasing traffic, – reduced marketing expenses.
Sales	– deals with members of a community for the acquisition of products and services on corporation blog; – shared presentations on sales and common knowledge about new sales leads using internal wiki. – launching of exclusive product offers for followers (e.g. Twitter) as a way of looking for new sales opportunities; – launching of new product and benefit campaigns for community members on Personal Social Networking Sites (e.g. Facebook).	– improved sales process, – reduced sales expenses, – making customers into distributors, – increasing geographical coverage.
Supply Chain Management	– add social layer to supply chain management. – build purchasing coalitions. – hire "virtual" contractors. – monitoring the here and now of supply chain production and needs, tracking of the logistical updates, sharing the data, monitoring the progress. – pooling resource and ideas from across the entire network of suppliers.	– increase the effectiveness of SCM. – increase productivity. – building a network of profitable business relationships. – problems are dealt more quickly and effectively.
Customer services	– corporate blogs can be used for communicate product development information and gather comments from consumers or for capture of comments regarding claims or requests by customers to act accordingly. – microblogs can be used use for customer feedback, – creation of a knowledge base of customer service procedures. – publication of educational videos on how to use certain product, extending the user manual online (e.g. YouTube); – response to support inquiries and product complaints, monitored by an exclusive team. – resolution of enquiries among community members. – opinion gathering regarding services and products.	– reduced customer services expenses, – shorten the process of customer service. – increasing the quality of customer service, – improved customer satisfaction.

Examples of social media using		Benefits of SM using
Research & development	– monitoring to gather intelligence about the company, product or service, competitors or industry. – conducting research (turn to customers into focus groups). – spread the word about new products and services to workforce. – solicit ideas, opinions, and feedback to incorporate them into existing and/or new products and services. – demonstration a series of new product (e.g. YouTube). – reach larger numbers of audience and receive quick feedback (e.g. Twitter).	– development of products and services in line with customer expectations. – faster introduction of new products. – increase in innovation of company.

Source: Own elaboration based on: [CISCO, 2013; Gehman, 2011; Falls, Deckers, 2011; Handfield, 2013; Deloitte, 2012; Buchnowska, 2013].

In R&D area social technologies allow to use crowdsourcing to the creation and development of products and services. Customers may become a valuable source of innovations such as the inventing, design, improvement or testing of ideas and new product concepts [Jelonek, 2012]. In the case of SCM, social media can improve the cooperation between enterprise and its suppliers or other business partners. This results in higher efficiency of the entire supply chain.

1.2. Social recruiting – the use social media at the HRM department

Business areas in which the social media technology is used more and more frequently is human resource management (HRM). Web 2.0 technologies have many applications in this field. HR departments increasingly make use of social media (particularly LinkedIn and Facebook) to gather information when making hiring and promotion decisions [The ACE Group, 2011]. SM tools can be also use to [CISCO, 2013]:

- promote company among potential employees;
- identify/gather information on job candidates;
- train employees;
- tap into pool of "passive" job seekers via professional social networks (e.g. LinkedIn, Xing);
- discuss ideas, post news, ask questions and share links by employees.

In Table 1.3 there are presented some definitions of the concepts connected with the use of social media in human resource management area. They are often incorrectly considered as synonyms.

Table 1.3. Examples and benefits of social media application in different areas

Term	Synonyms	Author	Definition
e-HRM	e-HR, electronic Human Resource Management	Stefan Strohmeier	The planning, implementation and application of information technology for both networking and supporting at least two individual or collective actors in their shared performing of HR activities.
Social HRM	HRM 2.0	Erol Erena, Pelin Vardarlier	The use of social media in processes related to human resources management
E-recruitment	online, Internet, web-based recruiting	http://www.hrzone.com/	Using electronic resources – the internet and HR software – to guide or assist the hiring process in order to reduce the administrative and financial burden of recruitment and gain access to a wider pool of talent
Social recruiting	recruiting 2.0, recruitment 2.0 s-recruiting	Paul Jacobs	Harnessing the evolution of Web 2.0 technologies and social media tools to communicate, engage, inform, and recruit our future talent

Source: Own elaboration based on: [http://www.hrzone.com/hr-glossary/what-is-e-recruitment; Strohmeier, 2007; Eren, Vardarlier, 2013].

One of the primary purposes of using social media in the area of HRM is to support the recruitment and talent acquisition [Chwiałkowska, 2013]. According to the data from Eurostat, 24% of European Union enterprises which use social media, use SM for recruiting. Use SM technologies in process of staff recruitment is defined as social recruiting.

Social recruiting (recruiting 2.0) means sourcing candidates for jobs via social media channels and social media networks and "harnessing the evolution of Web 2.0 technologies and social media tools to communicate, engage, inform, and recruit our future talent" [Vicknair et al., 2010].

Rise in popularity of social recruiting results from a number of advantages it can bring. The most important benefits of using social media are:

- job visibility, increasing the pool of applicants reached (potential employees) and increase in quantity of candidates [Massplanner, 2015; Broughton et al., 2013];
- opportunity to target recruitment at specific groups of potential candidates – filter of candidates depending on qualifications, skill levels, salary range, experience and titles [Massplanner, 2015; Broughto et al., 2013];
- access to numerous details and information about a potential candidate – the possibility of using big data to select the best candidate (increase in candidate quality) [Massplanner, 2015];
- the ability to reach passive or potential candidates who are not necessary searching for a new job [Massplanner, 2015];

- the ability to verify the accuracy of the information about candidates [Microsoft, 2010];
- shorter time needed to hire employees;
- sourcing for candidates via social media platforms is cost–effective compared to traditional methods of recruitment [Massplanner, 2015];
- the possibility of using non-standard methods of recruitment as a scouting, competitions, crowdsourcing talent or a gamification;
- ability to implement Employee Referral Program (increase in quantity and quality of employee referrals).

Through the use of social media, recruitment can be not only a hiring a new employee when someone leaves, but it can be a continuous process. It helps organizations to get the appropriate employee and effectively manage of talents.

However, it should be aware, that with the upside of using SM for recruiting, exists also downside which can occur in three major areas of risk: reputational, legal and operational [The ACE Group, 2011].

1.3. The state of social recruiting

The importance of social media in conducting the recruitment process is great and continues to grow. According to a study carried out annually since 2008 by Jobvite, 93% of recruiters have used or planned to begin use social media for recruitment process in 2014, i.e. over 10% more than in 2009 (Table 1.4). The Jobvite Social Recruiting Survey 2014 was completed by 1,855 recruiting and human resources professionals spanning across industries.

Table 1.4. Use of social media for recruitment

Use social media for recruitment	2009	2010	2011	2012	2013	2014
Yes	68%	73.3%	80.2%	92%	94%	93%
Plan to begin	13.7%	9.3%	8.7%			
No	17%	13.8%	10.5%	8%	6%	7%
I don't know	2%	3.5%	0.5%			

Source: Own elaboration based on: [Jobvite, 2009; 2010; 2011; 2012; 2013; 2014].

Social recruiting is a recruiting method which the largest number of employers plan to increase investment in (Table 1.5). The data also shows that more and more employers declares increased investment in social media as a tool for recruitment – 46% in 2010 and 73% in 2014. Recruiting leaders believe social and professional networks (37%), employer brand (33%) and passive candidate recruiting

(26%) are the most essential and long-lasting trends in recruiting [LinkedIn, 2015].

Table 1.5. Recruiting methods which employers plan to increase investment in

Recruiting methods which employers plan to increase investment in	2010	2011	2013	2014
Social media	46%	54.5%	73%	73%
Referrals	19.8%	29.8%	62%	63%
Corporate career site	28.8%	27.7%	61%	60%
Direct sourcing	20.5%	25.5%	57%	57%
Mobile career site	–	–	–	51%

Source: Own elaboration based on: [Jobvite, 2009; 2010; 2011; 2012; 2013; 2014].

The rise in popularity of social recruiting due to an increased use of social media by job seekers. Adecco, the biggest employment agency, has published the comprehensive global study on the use of social media in recruiting and job search. They have gathered the answers of over 17,000 job seekers and more than 1,500 recruiters from 24 countries. The data shows [Adecco Group, 2014] that 55% of job seekers use social media for job search purposes (and 73% of recruiters use social media for their daily HR activities). 63% of job seekers use SM for job searching, 55% for researching potential employers' pages, 49% – to distribute their CV online, 47% – for checking what other say about potential employers and 40% – for personal branding. 29% of job seekers were contacted through social media by a recruiter at least once and 9% received a job offer. Largely the most used social networking site (SNS) for job search purposes (35%) is LinkedIn, followed by Facebook (17%). Figure 1.1. shows how many job seekers use social networking sites in general and for job search.

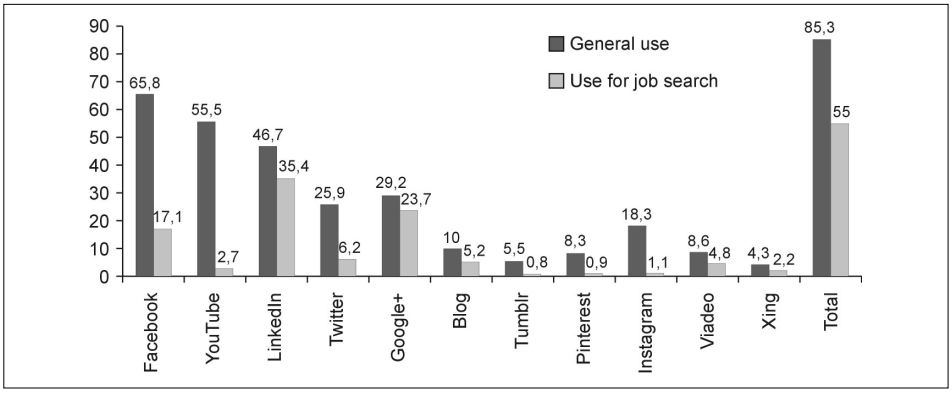

Figure 1.1. Social media used in general and for job search by job seekers

Source: [Adecco Group, 2014].

As well as among job seekers, LinkedIn is the most popular social networking site among recruiters. 94% of them used LinkedIn for recruiting in 2014 – up 16% since 2010 [Jobvite, 2014]. Facebook was used by 66% of recruiters and Twitter by 52% of them (Table 1.6). LinkedIn and Facebook are regarded as the most efficient social media platforms in the recruitment process [Adecco Group, 2014].

Table 1.6. Social network sites used by recruiters

SNS	2010	2011	2012	2013	2014
LinkedIn	78%	87%	93%	94%	94%
Facebook	55%	55%	66%	65%	66%
Twitter	45%	47%	54%	55%	52%
Google+	–	–	–	18%	21%
Blog	19%	16%	–	20%	20%
YouTube	14%	12%	–	15%	15%
None	14%	9.4%	–	–	14%

Source: Own elaboration based on: [Jobvite, 2010; 2011; 2012; 2013; 2014].

The most recurrent activity that is pursued through SNS by recruiters is job advertising. 65% of recruiters and HR professionals use these types of sites for advertising jobs [Adecco Group, 2014]. Other popular activities that are pursued by recruiters through SNS are: active sourcing of passive candidates, checking the accuracy of an applicant's CV, receiving job applications, checking an applicant's network, employer branding, checking content posted by an applicant and checking references of an applicant. Actions taken by recruiters are different, depending on the SNS (Table 1.7).

Table 1.7. Tactics for recruiting used on social network sites by recruiters

SNS	Tactics for recruiting	2013	2014
LinkedIn	Search for candidates	96%	95%
	Contact candidates	94%	95%
	Keep tabs on potential candidates	93%	93%
	Vet candidates pre-interview	92%	93%
	Post jobs	91%	92%
Facebook	Showcase employer brand	65%	59%
	Generate employee referrals	51%	51%
	Post jobs	48%	48%
	Vet candidates post-interview	35%	35%
	Vet candidates pre-interview	31%	32%

SNS	Tactics for recruiting	2013	2014
Twitter	Showcase employer brand	47%	44%
	Post jobs	43%	39%
	Generate employee referrals	31%	32%
	Vet candidates post-interview	18%	18%
	Search for candidates	–	17%
	Contact candidates	19%	–

Source: Own elaboration based on: [Jobvite, 2013; 2014].

As seen from the data in the table, LinkedIn is the main social networking portal which is used to search for workers and contact with candidate. On the other hand, Facebook and Twitter are primarily the channels of showcase employer brand.

In the period July-October 2015 year, study of hundred largest companies in Poland [http://www.forbes.pl/100-najwiekszych-firm-w-polsce-2014,ranking,175066,1,1.html#] was conducted. Objective of the study was a comprehensive analysis and assessment of the level of use of e-recruitment by the largest Polish companies. As part of the study, Content analysis and observation of activities undertaken on corporate websites and social media profiles by the 100 largest companies in Poland were carried out.

The study shows that active business recruitment activities on business social networking sites (LinkedIn, GoldenLine) are actively conducted primarily by those companies that have profiles in other SM. These companies more often publish job offers on business SNS and keep continuous communication with social media user-potential employees. They also enable the sharing of job offers, posted on the company website, on various social networking sites.

Due to the extensive scope of the survey, the detailed results of its realization will be presented in separate articles.

Conclusion

Presented in this article data show that social media are widely used by both: job seekers and companies in the recruitment process. The use of social media in the recruitment process can bring many benefits to organization.

However, organizations using social recruiting should be aware of two aspects. First, effective use of SM in the activities of recruitment requires the preparation of a coherent strategy that must fit organization's business strategy. Company's activities in area of social recruiting shouldn't be based only on ad hoc projects but every company should have a policy on recruiting using social media.

Managers should set business aims for social media use, identify measures its success and systematically measure the level of achievement of the objectives.

Secondly, the use of social media is associated with risks in the area such as legal, operational or reputational. To reduce the level of risk, the organization should create and implement a comprehensive social media policy and monitor what is being said about the company and how social media is used.

The article shows that there is a growing interest in social media as a tool to support recruitment processes. However, the presented data do not allow comprehensively assess the level of use of social technologies in recruitment of employees. That is why, it is necessary to carry out detailed and comprehensive scientific research of selected companies in order to evaluate their activity in social media. It will include two main stages. In the first stage the observation on the corporate websites will be carried out. This will assess the presence of companies in social media (based on links to social media network on web page) and will identify the IT tools used by companies in the recruitment process. In the second stage the observation of corporate profiles on social networking sites will be conducted. This will help to assess the activity of companies on social networks and to observe the reaction of users of social networking to business recruitment activities.

References

1. Adecco Group (2014), *Social recruiting. A global study 2014*, http://www.adecco.com/en-US/Industry-Insights/Documents/social-recruiting/adecco-global-social-recruiting-survey-global-report.pdf.
2. Broughton A., Foley B., Ledermaier S., Cox A. (2013), *The use of social media in the recruitment process. Research paper*, Acas, http://www.acas.org.uk/media/pdf/0/b/The-use-of-social-media-in-the-recruitment-process.pdf.
3. Buchnowska D. (2013), *Social Business – conceptual framework*, "Business Informatics", No. 4, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
4. Chwiałkowska A. (2013), *Social HRM – media społecznościowe jako narzędzie wspierające zarządzanie zasobami ludzkimi w organizacji*, in: *Systemy wspomagania organizacji SWO 2013*, Porębska-Miąc T., Sroka H. (ed.), Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
5. CISCO (2013), *Social Media: Cultivate Collaboration and Innovation*, http://www.cisco.com/en/US/services/ps2961/ps2664/CiscoServicesSocialMediaWhitePaper.pdf.
6. Deloitte (2012), *CRM 2.0 or social CRM for Financial Industry*, http://www.deloitte.com/assets/Dcom-Croatia/Local%20Assets/Documents/2012/FSINews03.02-Social_CRM.pdf.
7. Eren E., Vardarlier P. (2013), *Social Media's role in Developing an Employees Sense of Belonging in the Work Place as an HRM Strategy*, in: *The Proceedings of 9th International Strategic Management Conference*, Özşahin M. (ed.).

8. Eurostat (2013), *Special module 2013: enterprises – use of social media*, http://ec.europa.eu/eurostat/data/database?node_code=isoc_cicce_bnf.
9. Falls J., Deckers E. (2011), *How to Use Social Media for Research and Development*, http://www.entrepreneur.com/article/220812.
10. Gehman S. (2011), *3 Ways to Use Social Media for Product Research and Development*, http://www.socialmediaexaminer.com/3-ways-to-use-social-media-for-product-research-and-development.
11. Handfield R. (2013), *Effective supply chain management through social networking: Guest Blog by Nicholas Moores*, http://scm.ncsu.edu/blog/2013/08/01/effective-supply-chain-management-through-social-networking-guest-blog-by-nicholas-moores/.
12. IBM (2013), *Collaboration and access for Smarter Education*, http://public.dhe.ibm.com/common/ssi/ecm/en/ege12345usen/EGE12345USEN.PDF.
13. Jelonek D. (2012), *The choosen determinants of customer co-created innovations*, in: *Creativity support systems, Methods, and applications*, Sroka H., Stanek S., Pańkowska M. (ed.), Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice.
14. Jobvite (2009), *2009 Social Recruiting Survey Results*, http://www.slideshare.net/martinosmar/2009-jobvite-social-recruitment-survey1.
15. Jobvite (2010), *2010 Social Recruiting Survey Results*, http://web.jobvite.com/rs/jobvite/images/Jobvite%202010%20Social%20Recruiting%20Report_2.pdf.
16. Jobvite (2011), *2011 Social Recruiting Survey Results*, http://web.jobvite.com/rs/jobvite/images/Jobvite-SRP-2011.pdf.
17. Jobvite (2012), *2012 Social Recruiting Survey Results*, http://web.jobvite.com/rs/jobvite/images/Jobvite_2012_Social_Recruiting_Survey.pdf.
18. Jobvite (2013), *2013 Social Recruiting Survey Results*, http://web.jobvite.com/rs/jobvite/images/Jobvite_2013_SocialRecruitingSurveyResults.pdf.
19. Jobvite (2014), *2014 Social Recruiting Survey*, https://www.jobvite.com/wp-content/uploads/2014/10/Jobvite_SocialRecruiting_Survey2014.pdf.
20. Kane G.C, Palmer D., Phillips A.N., Kiron D, Buckley N. (2014), *The 2014 MIT Sloan Management Review and Deloitte social business study*, http://dupress.com/articles/social-business-study-mit-smr/.
21. LinkedIn (2015), *2015 Global Recruiting Trends*, https://business.linkedin.com/content/dam/business/talent-solutions/global/en_US/c/pdfs/recruiting-trends-global-linkedin-2015.pdf.
22. Massplanner (2015), *Benefits of Recruiting via Social Media*, http://www.massplanner.com/benefits-recruiting-via-social-media.
23. McKinsey&Company (2009), *How companies are benefiting from Web 2.0: McKinsey Global Survey results*, http://www.mckinsey.com/insights/business_technology/how_companies_are_benefiting_from_web_20_mckinsey_global_survey_results.
24. McKinsey&Company (2015), *Transforming the business through social tools*, http://www.mckinsey.com/insights/high_tech_telecoms_internet/transforming_the_business_through_social_tools.
25. Microsoft (2010), *Online Reputation in a Connected World*, http://www.job-hunt.org/guides/DPD_Online-Reputation-Research_overview.pdf.
26. Strohmeier S. (2007), *Research in e-HRM: Review and implications*, "Human Resource Management Review", No. 17.

27. The ACE Group (2011), *Social Media: The Business Benefits May Be Enormous, But Can the Risks — Reputational, Legal, Operational — Be Mitigated?*, http://www.acegroup.com/us-en/assets/ace-progress-report-social-media.pdf.
28. Vicknair J., Elkersh D., Yancey K. (2010), *The Use Of Social Networking Websites As A Recruiting Tool For Employers*, "American Journal of Business Education", Vol. 3, No. 11, http://cluteinstitute.com/ojs/index.php/AJBE/article/viewFile/57/55.
29. Walker L. (2013), *Social Business Definition. Capitalism with a Heart, or Corporate Social Media?*, http://personalweb.about.com/od/social-marketing/a/Social-Business-Definition.htm.
30. http://www.forbes.pl/100-najwiekszych-firm-w-polsce-2014,ranking,175066,1,1.html.
31. http://www.hrzone.com/hr-glossary/what-is-e-recruitment.

Chapter 2

SysML in practice. A furniture manufacturer case study

Michał Firek

Introduction

Systems Modeling Language (SysML) is a general-purpose graphical modelling language that supports the analysis, specification, design, verification, and validation of complex systems. These systems may include hardware, software, data, personnel, procedures, facilities, and other elements of man-made and natural systems [Wrycza, Marcinkowski, 2010].

The language was created by system engineers for system engineers. Systems engineering was a discipline that has not had a uniform modelling language. It requires a language that is independent of specific disciplines like software, hardware, or mechanics. INCOSE (International Council of Systems Engineering – the worldwide systems engineering organisation) was the originator of creating a language for system engineering [Weilkiens, 2007].

UML 2 was a good candidate for that role. First, it meets the most important requirements of systems engineering, and second, it is popular and widely used. In addition, there is a considerable amount of literature, and many seminar programs are offered. The powerful extension mechanisms of this language allow it to adapt to the needs of systems engineering. The adapted UML is called Systems Modelling Language, or SysML for short [Weilkiens, 2007].

SysML was created as an extension of UML but UML consists of 14 [Object Management Group, 2015] diagrams and SysML consists of only 9 diagrams [Object Management Group, 2013]. This is the result of removing from SysML elements explicit to UML not required in systems engineering.

2.1. Main research trends related to the SysML language

SysML language is widely used as a tool in the process of creating object-oriented systems. Notation of the language is based on certain syntax, semantic and pragmatic rules. Apart from that SysML does not limit the designer in his work. The language allows the user to customise and adapt it to the current problem,

thanks to the extension mechanism. The responsibility that lies with designers and analysts for creating a system that fits the requirements in a specified time period and budget, leads them use a language that is easy to use and is supported by available tools [Wrycza, Marcinkowski, 2010].

SysML language meets these requirements and it is often chosen as a design standard. It can be used as a design language in a wide range of projects for a variety of companies [Wrycza, Marcinkowski, 2010]. In the context of current, shared design work, clearly specified areas of use can be set for SysML language and are:

1. Modelling and analysis of information systems [Tolk, Turnitsa, 2012; Fan et al., 2013].
2. The application of SysML language in industries [Dumitrescu et al., 2013; Rahm et al., 2012].
3. The use of SysML language as simulation process [Rahm et al., 2012; Batarseh, McGinnis, 2012].
4. Extending the language with functional and non-functional requirements [Tsadimas et al., 2012; Chouali et al., 2013].
5. Integration with other standards and languages [Colombo et al., 2008; Bousse et al., 2012].
6. Extending SysML language with new elements [Tsadimas et al., 2012; Knorreck et al., 2011].
7. Using SysML in the process of system design [Udagawa, 2011; Wille et al., 2013].

Using SysML in a computer added system engineering process [Steinbach, 2012; Schönherr, Rose, 2011].

2.2. Furniture manufacturer case study

An ISO 9001certified furniture manufacturing company of Goods standing, with a long tradition and years of experience in the market, wants to order an information system that will support the staff in their everyday work. The main parent company is a manufacturer of furniture for offices, banks and hotels. The furniture is manufactured in accordance with pre-existing models presented in the catalogues, and at individual custom orders. All the furniture is made to order, and is not available to hand. Thanks to this firm does not have a warehouse with finished goods. Usually implementation of an order is 14 working days, due to subcontractors and the supply time of production. This period may be lengthened or shorten, depending on the volume of furniture ordered, the complexity of the design and availability production materials. The main raw material used in furni-

ture production is chipboard melamine, the colours of which can be chosen from a colour palette. The chipboard used to produce the furniture comes in thicknesses of 18mm, 25mm and 36mm. The furniture is finished using a PVC border in the same colour as the chipboard, or in a different colour, in accordance with the customer's requirements. The company offers a wide range of fittings in aluminium, plastic and steel, which are customisable with text, not only to finish the piece of furniture but also to decorate it. The company also offers delivery services to an address specified by the customer.

The information system should be based on the analysis of processes within the company, related to the operations of the production order. Also helpful is the ISO book that describes the quality management system, management procedures system and interactions between processes occurring in the organisation. Also the book describes the organisation structure and the history of the company.

2.3. Selected research outcomes

System design presents the functionality and modules associated with complex production orders by staff of a furniture manufacturer. The main task of the system is the make it easier, faster and to improve the process of issuing internal documents associated with the production order. Thanks to the system, the exchange of information between employees of different departments of the company, between the company and subcontractors, as well as the company and customers, will be quicker and easier, and the information will be more reliable and accurate.

The system should be approachable and intuitive to use, saving time and effort in the implementation. The system should consist of associated modules, allowing for modifications in any individual module. The modular construction of the system will increase its flexibility and give the ability to add new modules and functionalities as the company expands. The system, based on existing procedures, must be adapted to the requirements of users and the specifics of the company and will take three departments of the company: sales, production and purchasing, into account. The project has been expanded with additional functionalities, containing ready-made systems.

The first diagram that describes the basics structures of the system and the relationship between the modules is the requirement diagram. This diagram is the basis for other diagrams in the modelled system. The requirement diagram contains also blocks, flows and the relationships between the blocks are presented on separate diagrams and internal block diagrams.

The main element of the system is the production order, and its operation, therefore, diagrams presented later in this chapter refer to this. The parametric diagram describes the functions of the system that estimates the time of order execution, the second diagram presents how the cost of the furniture is to be calculated. Next are use case diagrams that show the functionality of each department of the company. The division was initially presented at the block diagrams.

The activity diagram, due to the requirements of the block order, describes the cycle of a production order inside the company. During the circulation, the order can take the different states that are shown on the state machine diagram. The sequence diagram describes the hierarchy of activities related to the order issuing.

The system model presents the requirements, components and operations associated with their implementation. The included diagrams depict elements that should be in the system and the relations between them that should occur.

2.3.1. Requirements diagram

The requirement diagram shows requirements that the supporting system in operating the production orders should fulfil and the relationships between the diagram and others system elements.

The diagram shown in Figure 2.1 describes functional requirements that the system must fulfil to cope with the demands of the user. The main requirement of the system is to support the process of handling orders, which has been presented as a requirement called *Production orders management*. This requirement was decomposed into four other requirements, i.e., *Invoice, Customer orders, Production orders*, and *Information flow*. The *Invoice* requirement states that the system should register the entire invoice and facilitate the process of invoicing the customer. This requirement will be met by *the Invoice* block. In the case of *Information flow* requirement, the system should facilitate the exchange of information between departments and the employees of each department. The above mentioned requirement will be satisfied by the *Information* block. The *Customer orders* requirement needs the system to register all the orders placed by customers, allowing to them to be viewed and modified, responsible for it will be block *Customer orders*. The *Customer offers* requirement is a component of the *Customer order* requirement, and stores and registers all versions of offers presented to customers. The functionality will be satisfied by the *Offer* block.

The *Production orders* requirement needs the system to register, inspect, and modify orders and the production of goods. The system should register all the orders and allow their modification, whenever possible: the, block *Production order* block is responsible for this functionality.

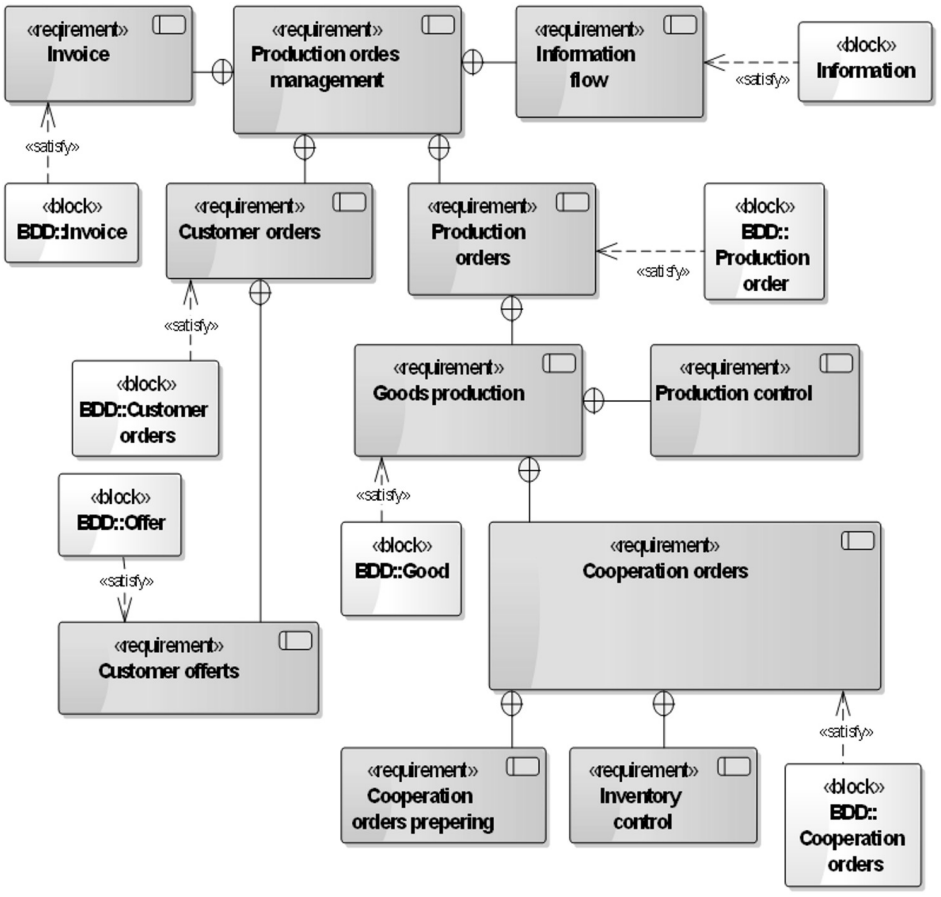

Figure 2.1. Requirements diagram
Source: Own elaboration.

The *Production orders* requirement includes the *Goods production* requirement that will force the system to analyse any carried out orders and prepare the necessary documents needed to produce goods. The functionality will be satisfied by the *Goods* block. *Production control* is a separate requirement demanding the system to provide information about order execution status. This requirement is also satisfied by the *Goods* block. The *Goods production* requirement includes the *Cooperation orders* requirement that asks the system to control the stock status and prepare cooperation orders. These requirements are represented by the *Inventory control* and *Cooperation orders preparing* requirement. These requirements will be satisfied by the *Cooperation orders* block.

2.3.2. Block definition diagram for the furniture manufacturer

The block definition diagram (Figure 2.2) presents the inner structure of the furniture manufacturer and the relationship between each of its departments.

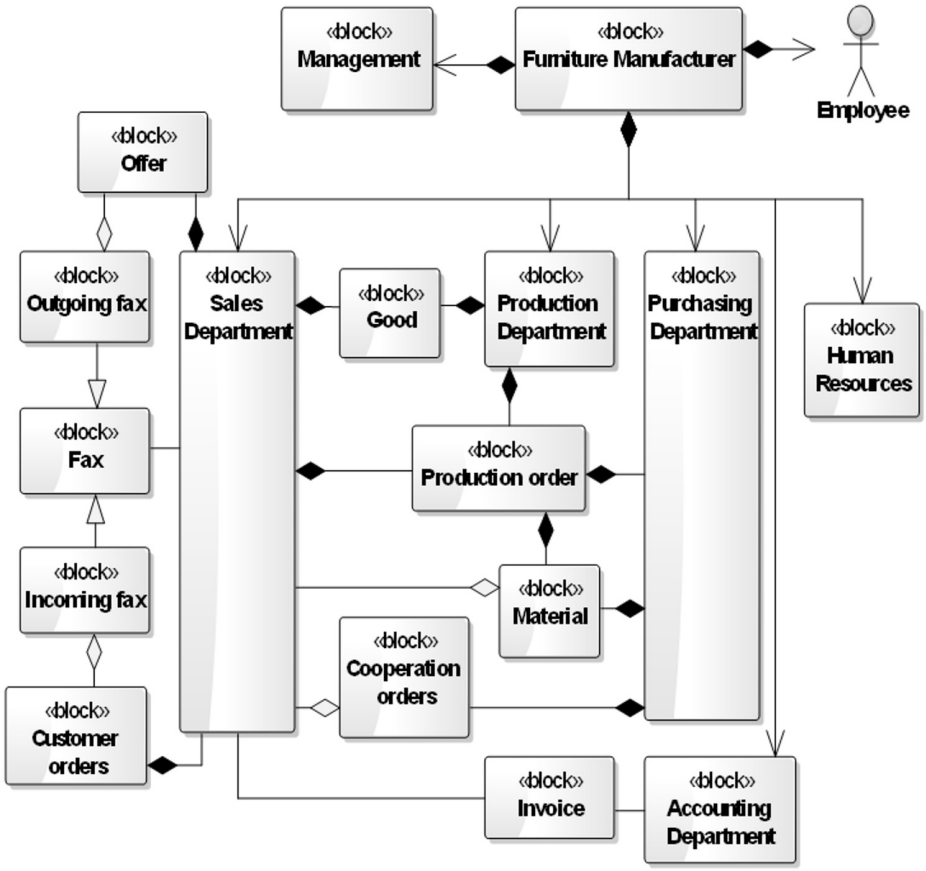

Figure 2.2. Block definition diagram

Source: Own elaboration.

The *Sales Department* block is connected with the *Offer* block with a composite association because, an offer is created as an answer to customer price inquiry only by the sales department. An offer is connected by a reference association with the *Outgoing fax* block, because not all information sent outside the firm is in answer to price inquires. Between the *Customer order* block and the *Sales Department* block a composite association occurs, resulting from the fact that customer orders concern not only the sales department but also the whole company. The *Sales Department* block is connected with the *Production order* block, because in the sales depart-

ment each production order begins and ends its path through the furniture manufacture. Here it is started, put into execution, and closed at the end by an invoice. The *Goods* block is also connected with the *Sales department* block with use of a composite association, because the sales department employee is responsible for goods created by the production department. The reference association connection connects the *Sales Department* block and the *Material* block that results from the fact that sales department may sell material. The same type of connection is used between the *Sales Department* and *Cooperation order* blocks, because finished goods purchased for the purpose of further resale are ordered by sales department employees. The *Production Department* block represents the production department of furniture manufacturer. Between the *Production Department* and *Goods* blocks is a composite association connection, because the production department produces goods using material and components that are described in the production order. From this another composite association connection results, between production department and the *Production Order* block. This connection results from the fact that production department is responsible for the realisation of the production order, producing new goods according to the technical description and quantify qualitative statement.

The third department that is included in the design of the system is the *Purchasing Department. T*he composite association between the *Production order* and *Purchasing Department* blocks results from the fact that the purchasing department is responsible for the organisation of materials necessary in the production process. The *Purchasing Department* block is connected with the *Material* and *Cooperation Order* blocks by the use of a composite association. This is due to the purchasing department organising materials needed to produce the goods and keep's minimal inventory levels.

2.3.3. Use case diagram for Production Department

The production department is shown in Figure 2.3, as a case diagram, describing the functionality of the system and its relationship with the actors representing department employees.

The diagram shows two actors: the *Production department employee* and the *Production department manager*. The Manager is connected with the the *Analyse order* use case by an association with multiplicity, resulting from the fact that the manager analyses one or more orders at the same time. After analyses, the manager creates a production card, that is described by an association between the *Production department manager* actor and the *Make production card* use case.

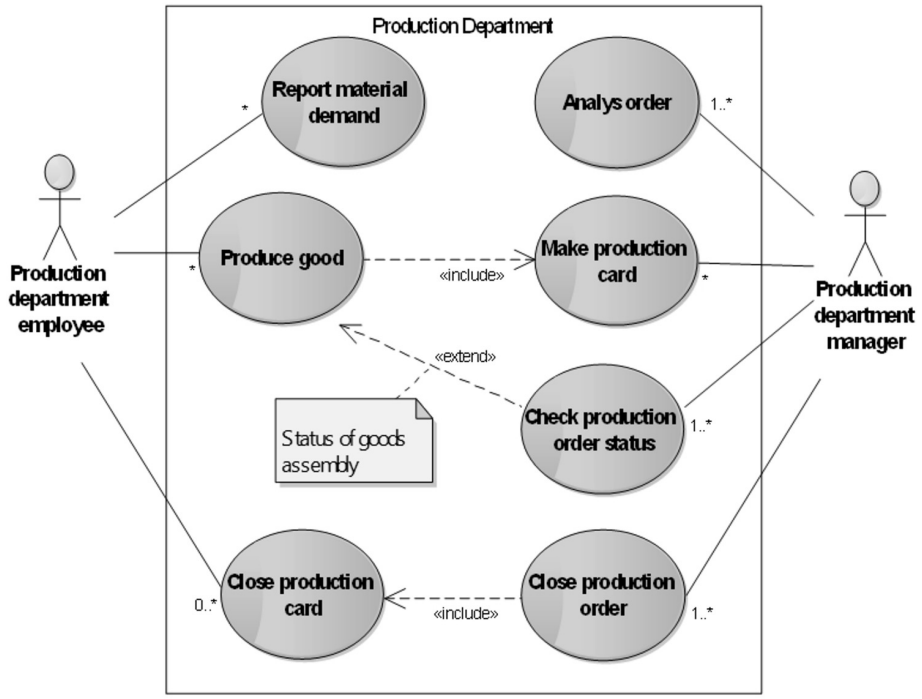

Figure 2.3. Use case diagram for Production Department
Source: Own elaboration.

In a situation where an orders status is needed the manager uses the *Check production order status* use case. This use case is extended by the *Produce Goods* use case. When all goods from one production order have been completed, the manager closes the production card, a functionality a described by the *Close production order* use case. This use case includes the *Close production card* use case, because the production cards must be closed before the production order is closed.

The second actor in the diagram is *the Production department employee*, who is connected with three use cases. The first use case is *Report material demand, a* connection which has multiplicity, as the employee can repeatedly report a demand for materials. The second use case is *Produce Goods.* The association between this use case and the actor also has multiplicity, as the employee may be involved in processing many orders. The last use case is *Close production card.* When goods are assembled, it is the duty of employee is to close the production card. An employee may close none of production cards or any of those shown by the use of an association with multiplicity.

2.3.4. Activity diagram Production Order

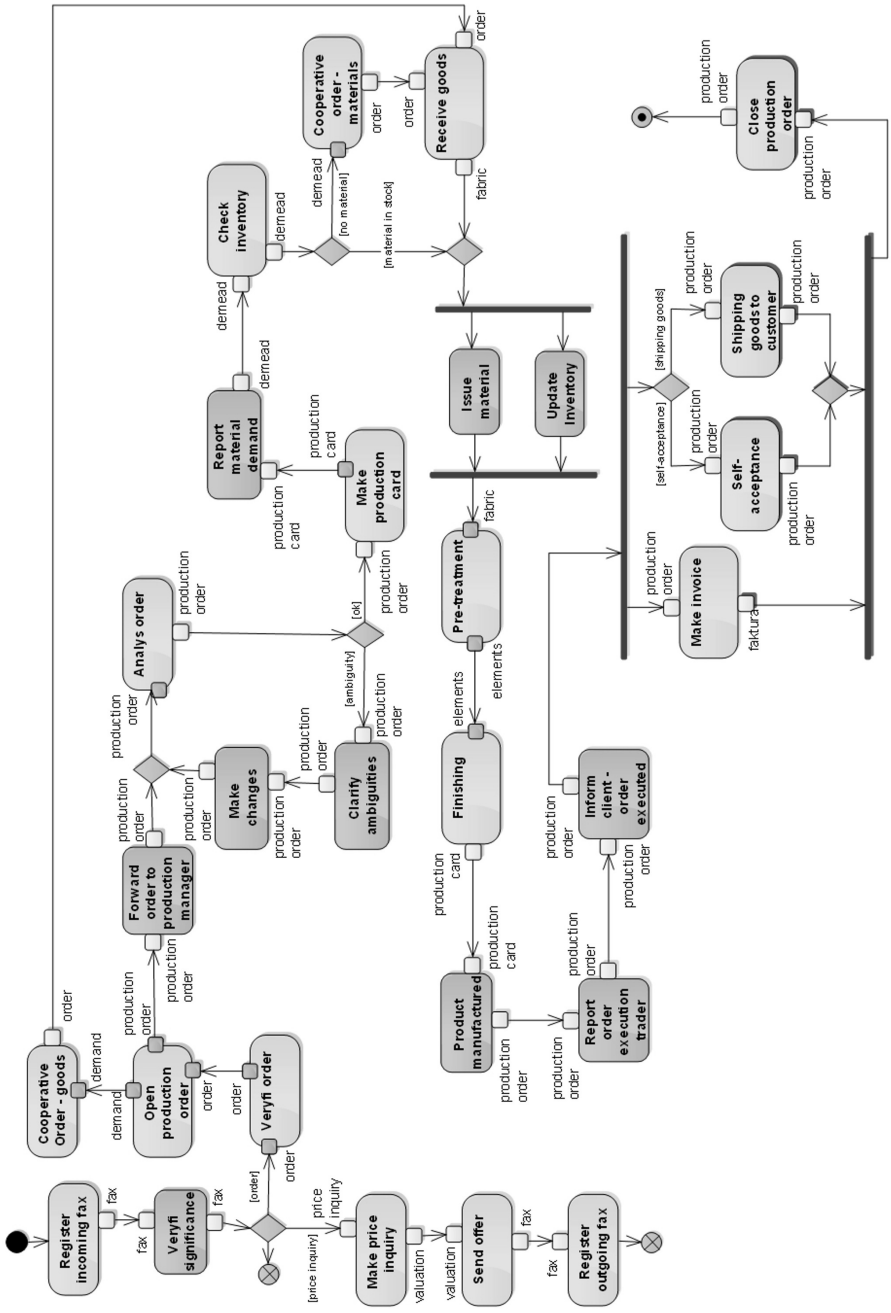

Figure 2.4. Extended activity diagram Production Order

Source: Own elaboration.

The activity diagram showing the circuit of a production order through the various furniture manufacturer departments is presented in Figure 2.4. The first activity in the diagram is *Register incoming fax, an* activity which is decomposed. Next, the registered fax is passed to activity *Verify significance* that checks if given fax is an order, price inquiry or contains any other information. If it the fax is qualified as 'other information' the activity ends with a flow final.

Any fax qualified as a price inquiry is sent to the *Make price inquiry* activity. A prepared estimate is then sent to the customer with the use of the *Send offer* activity. After this, it is marked as an outgoing fax and sent to the *Register outgoing fax activity*, where it is registered in the system.

Qualifying a fax as an order starts the handling process. The order is send to the *Verify order* activity, in order to check that all the information necessary to issue the order is given. When all needed information is collected, the order is send to the *Open production order* activity, in order to issue a production order. This activity also generates any requirements for goods purchased for the purpose of further resale. Such information is send to the *Cooperative order-goods activity*. This activity is responsible for generating orders and sending them to the external company. From here, the order is send to the *Receive Goods activity and t*he issued production order is send to the production manager. The *Forward order to production manager* activity is responsible for this process. All described activities and actions concerning the sales department of the furniture manufacture are presented in the diagram by use of the *Sales department* partition. The Purchasing Department is responsible for supplying materials used to manufacture the product. The reported demand is transmitted to the *Check inventory* activity and, if the materials are in stock, they are released. If not, the demand is forwarded to the *Cooperative order – materials* activity, which is then responsible for creating an order. The order is transferred to the *Receive Goods* activity, including information checking whether the goods delivered, whether the material is consistent with the order, whether it is accompanied by evidence of an external issue, and whether it is undamaged. After the verification of the delivered goods, they are taken on the status of storage. Next the necessary materials can be sent to the production department. The *Issue material* activity is responsible for this action. This action is in a concurrent area between a fork and join node. The other action that is in the area is *Update inventory*, which makes a quantitative change in the state of materials.

Realised materials are forwarded to the production department to the *Pre-treatment* activity. This is responsible for pre-preparation items that will be needed to carry the goods. From the *Pre-treatment* activity, the elements are sent to the *Finishing* activity, which is responsible for assembling the goods from the elements according to the production card. When the goods are assembled, this is reported to the production manager, as shown in the diagram, by means of the *Product manufac-*

tured activity. Additionally, this activity stores information about the status of goods and after all goods from one production order have been assembled, this is marked on the order and forwarded to the *Report execution of order to trader* activity.

The order comes back to sales department in order to finalise the customer service process. The completed order is forwarded to the *Inform client-order executed* activity which reports to the client that his order had been completed. Control is transferred to concurrent area with the final steps of handling orders. To issue an invoice, the control is forwarded to the *Make invoice* activity. The goods can be either sent to the client or be made available for self-collection, by use of the *Self-acceptance* and *Shipping Goods to customer* action. The last activity in the diagram is *Close production order*, which is responsible for the termination of the order and the notation of this in the system.

2.4. Discussion

The information system presented in this work covers only a portion of the fundamental processes associated with handling production orders in the furniture manufacturer. It presents modules that are part of a future system that will allow the appropriate elaboration of production orders, selecting relevant, important information for employees of various departments.

The SysML is an easy to use language for system engineers in creating a model of the system based on the specified requirements. The model is described graphically and is easily understood by stakeholders, users and programmers. This gives the possibility for the validation and adjustment of the system, until an acceptable architecture is defined in early stages of creation. The requirements diagram is the first diagram created in the design process, requirement are described by using text. There are no limitations for describing a requirement and can lead to a situation where the requirements cover too much space on a diagram. The decomposition of the requirement diagram may solve this problem, as may limiting the words used for describing a requirement, or organising requirements into packages.

Available tools that support SysML do not cover the whole specification during the design process, as the most common problem is describing the requirements. Tolls supported required text fields leads to the entire text being written in one line, affecting the readability of the diagram.

Conclusion

The most important thing that allows the furniture manufacture to function is the production order. Therefore, the manufacturer requires an information system that will integrate all features and steps that must be taken in order to support the production order during its life cycle.

Creating appropriate software that will support the production order will concentrate all elements in an integrated, centralised system, with the required functional modules. This will accelerate and facilitate the work of various departments and ease the exchange of relevant information that affects the order execution process. This will also avoid bugs and save time spent on routine, reproducible activities. Departmental staff will perform their actions quickly and efficiently, thus increasing the production capacity translating into the ability to service more clients. This will increase profitability and enhance the image of the company in the market.

To create an efficient system, the design process needs to cover the company as a whole. Case studies in other companies of the same industry are needed to compare and identify common requirements. The functionality of designed system should also be verified with the use of parametric diagrams. Most importantly, during the design process, the need for a dedicated methodology was noticed: a methodology that will support the whole design process of a system with the use of SysML language.

References

1. Batarseh O., McGinnis L. (2012), *SysML to Discrete-event Simulation to Analyze Electronic Assembly Systems*, TMS/DEVS '12 Proceedings of the 2012 Symposium on Theory of Modeling and Simulation – DEVS Integrative M&S Symposium, art. No. 48.
2. Bousse E., Mentré D., Combemale B., Baudry B., Katsuragi T. (2012), *Aligning SysML with the B Method to Provide V&V for Systems Engineering*, MoDeVVa '12 Proceedings of the Workshop on Model-Driven Engineering, Verification and Validation.
3. Chouali S., Hammad A., Mountassir H. (2013), *Assembling Components using SysML with Non-Functional Requirements*, Electronic Notes in Theoretical Computer Science.
4. Colombo P., Bianco V., Lavazza L. (2008), *Towards the Integration of SysML and Problem Frames*, IWAAPF '08 Proceedings of the 3rd international workshop on Applications and advances of problem frames.
5. Dumitrescu C., Mazo R., Salinesi C., Dauron A. (2013), *Bridging the gap between product lines and systems engineering: an experience in variability management for automotive model based systems engineering*, SPLC '13 Proceedings of the 17th International Software Product Line Conference.

6. Fan Z. Yue T., Zhang L. (2013), *A generic framework for deriving architecture modeling methods for large-scale software-intensive systems*, SAC '13 Proceedings of the 28th Annual ACM Symposium on Applied Computing.
7. Friedenthal T., Moore A., Steiner R. (2008), *A Practical Guide to SysML*; MK/OMG Press.
8. Knorreck D., Apvrille L., Saqui-Sannes P. (2011), *TEPE: a SysML language for time-constrained property modeling and formal verification*, ACM SIGSOFT Software Engineering Notes.
9. Object Management Group (2012), *OMG Systems Modeling Language (OMG SysML™)*, http://www.omg.org/spec/SysML/1.3/, accessed 06.08.2015.
10. Object Management Group, *Unified Modeling Language (OMG UML)*, http://www.omg.org/spec/UML/2.5/, accessed 06.08.2015.
11. Rahm T., Sadri K., Koch Ch., Thewes M., Konig M. (2012), *Advancement simulation of tunnel boring machines*, WSC '12 Proceedings of the Winter Simulation Conference, art. No. 65.
12. Schönherr O., Rose O. (2011), *A general model description for discrete processes*, WSC '11 Proceedings of the Winter Simulation Conference.
13. Steinbach D. (2012), *Guidelines for model based systems engineering*, DATE '12 Proceedings of the Conference on Design, Automation and Test in Europe.
14. Tolk A., Turnitsa Ch. (2012), *Conceptual modelling with processes*, WSC '12 Proceedings of the Winter Simulation Conference, art. No. 234.
15. Tsadimas A., Nikolaidou M., Anagnostopoulos D. (2012), *Extending SysML to explore Non-Functional. Requirements The Case of Information System Design*, SAC '12 Proceedings of the 27th Annual ACM Symposium on Applied Computing, p 1057-1062.
16. Udagawa Y. (2011), *Enhancing information retrieval to automatically trace requirements and design artefacts*, iiWAS '11 Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services, p 292-295.
17. Weilkiens T. (2007), *Systems Engineering with SysML/UML*, MK/OMG Press 2007.
18. Wille R., Gogolla M., Soeken M., Kuhlmann M, Drechsler R. (2013), *Towards a generic verification methodology for system models*, DATE '13 Proceedings of the Conference on Design, Automation and Test in Europ.
19. Wrycza S., Marcinkowski B. (2010), *Profile UMl 2.x w praktyce. Język inżynierii systemów SysML Architektura i zastosowania*, Helion, Gliwice.

Chapter 3

Authorisation conditions in IT systems used by self-government units

Przemysław Jatkiewicz

Introduction

At least once a month the media publish alarming information about IT systems break-ins and theft of users' passwords and logins. One can also find a number of combinations of the most popular passwords [Vance, 2010]. Despite the fact that each system, developed according to the generally accepted principles, stores encrypted passwords they are nevertheless decrypted or recovered by hackers. The opinion that passwords are dead and should be replaced by multi-factor authentication becomes more and more widespread. It is confirmed by an increasing technological possibility to decrypt codes. During the Passwords^12 Conference in 2012 the presented device in as much as 5.5 hours was able to break an 8-character password encrypted in one of the most popular algorithms used in the Windows Server systems [Gosney, 2012].

Not many publications indicate the necessity to conduct research concerning the further use of passwords in the process of verification and authentication. Cormac Herley and Paul C. van Oorschot conducted research which led them to the conclusion that the complete elimination of passwords is currently impossible. No other method of authorisation meets all the expectations of users. The opinion that passwords are dead should be considered premature or even harmful, as it leads to the discontinuation of activity aimed at improving this authorisation method [Herley, Van Oorschot, 2012].

It is necessary to emphasise that the majority of the abovementioned alarming information mostly concerns internet social networks or internet shops. One should not use this information to make an authorisation assessment in systems used by governmental or self-government organisations. These systems, as well as their usage process, are subject to legal regulations which mostly concern security and directly refer to the ISO 27001 norms.

Self-government institutions process data which is significant to citizens living in the area of operation of these entities. According to a popular opinion, self-government officials, obliged by numerous normative acts, attach particular importance to the protection of information entrusted to them. This group mostly

consists of individuals with higher education who work according to specific and clear rules and who are sensitive to any attempts of fraud.

Each institution is obliged to develop and establish the Information Security Policy which incorporates the authorisation principles [Calder, 2006]. The Information Security Officer appointed for this purpose ensures that its provisions are observed. Apart from security monitoring his responsibilities also include the training of IT system users [Mikkonen, 2014].

Bearing in mind numerous confirmed press reports concerning the theft of passwords of business systems users as well as legal restrictions referring to the IT systems of institutions, the initiation of research regarding the authorisation process in self-government units seems justified. These units constitute the most numerous group of public organisations which process a wide range of information about the citizens.

The aim of research presented here was the identification of conditions of the authorisation process in IT systems used by local self-government units. Research results indicated the possibility to use logging methods alternative to passwords.

3.1. Methodology of research

During study the following research methods were used [Kothari, 2014]:

- experiments,
- surveys,
- polls,
- interviews.

The experiment consisted in implementing the safe authorisation principles in the IT system of one of the organisations on the basis of catalogue services [Allen et al., 2008]. These principles were implemented gradually. Password quality control, conducted after each stage, involved attempts to decrypt passwords using the LOphtCrack v6.0.15 programme. Similar tests were conducted by J. A. Cazier and B. D. Medlin [2006] but they concerned the e-commerce systems. A completely different categorization and the password strength grade scale were used as well.

The implementation process was preceded by the survey used to collect information concerning the construction, change and remembering passwords by users. The group of respondents consisted of all clerks in organisation whose systemic passwords were tested. Anonymous respondents also expressed their opinion concerning the alternative logging methods such as biometric and behav-

ioural techniques, tokens or cryptographic cards. They also assessed the validity of password complexity control.

Surveys were sent to all the self-government units in Poland. These surveys were not anonymous. The respondents were IT specialists or managers. Survey questions mostly concerned the used catalogue services, encryption algorithms and authorization methods used currently and planned in the future. Respondents were also asked about the budget allocated for information technology and the size of the department working with new technologies.

A short survey was conducted in order to determine the number of alphanumeric characters which a clerk is able to remember. Time used to remember the string was measured as well. As it was established in the mid-20th century, human short-term memory is limited to about 7 pieces of information [Miller, 1956]. However, this research was conducted when the information systems, and hence the passwords used in the logging process, were not commonly used. Tests presented here were conducted in the natural working environment. Clerks detached from daily activities were performing under time pressure. Natural disturbances such as telephones or clients considerably influenced the results.

An alternative authentication method was selected on the basis of the extensive research material. The distributed survey verified its usefulness and functionality. Respondents marked the true sentence:

- The method is useful and I will use it.
- The method is useful but I prefer my own.
- The method is useless.

Respondents who expressed their negative opinion of the suggested method were interviewed in order to specify their expectations. After the modification, the logging module prototype was developed. A simple software, made by author, which simulated the process of logging as well as setting and changing its parameters was presented to respondents to evaluate.

3.2. The course of research

The research was initiated in January 2013 by distributing anonymous sur-veys among all the employees of the selected institution. They consisted of 6 questions – 2 semi-open-ended and 4 closed-ended. As many as 181 surveys were returned which amounts to 56.56%. Mistakes which consisted in the lack of answers to some of the questions appeared in 5 surveys. Such a low rate of return may stem from the fact that the respondents either did not attach particular importance to

this research or they were afraid to give answers despite the assurance of the anonymity of the survey [Couper et al., 2009].

At the same time password strength testing by attempts to decrypt them was enabled. It should be noted that the Security Policy was established in the researched organisation. Its provisions clearly stated that users should use at least 8-character passwords which include 3 out of 4 available groups of characters. These groups are:

- lower case,
- upper case,
- digits,
- special characters.

At the same time users were neither allowed to use names, surnames or dates connected with their lives or family members nor dictionary words. Clerks were obliged to change their passwords every month. The IT system based on the catalogue services Windows server 2003 operating in this organisation did not control the complexity or history of passwords.

All the users whose passwords were broken were informed via e-mail that their passwords do not meet the requirements stipulated in the Security Policy. After a month tests were repeated and respondents once more received e-mails with information supplemented with password construction and usage principles found in the Security Policy. In the next month tests were conducted as well and then the control of password complexity and history was enabled in Active Directory. Attempts to break passwords were made three more times at monthly intervals.

At the end of 2013 all the self-government units operating at the commune level, Voivodeship Offices and Marshal's Offices received surveys which consisted of 12 closed-ended questions. The number of surveys and answers provided by respondents divided by type of institutions is presented in Table 3.1.

Table 3.1. Surveys divided by type of researched organization

Type	Surveys	Answers	%
Marshal's Office	16	7	43.75
Voivodeship Office	16	2	12.50
Municipal Office with county rights	66	13	19.70
Municipal Office	742	83	11.19
Town and Community Council	42	6	14.29
Community Council	1 542	191	12.39
Total	2 424	302	12.46

Source: Own elaboration.

E-mails were signed by a digital certificate. They also included a phone number where additional information connected with surveys could be obtained. These elements were used to authenticate conducted research. In five cases they proved to be insufficient. Documents which confirmed the connection of the person conducting the study with the University of Gdańsk were requested.

A relatively small survey return rate can be explained by the unwillingness to provide authorisation details, which may pose a threat to the IT system security. The choice of the e-mail form was also disadvantageous. Electronic mail is not considered to be the official method of communication. It was observed that some email boxes of institutions reject the e-mail messages because the boxes are full, which indicates that clerks check them rather seldom. For this reason in 4 cases it was necessary to use ePUAP (Electronic Platform of Public Administration Services), the official system made available by the Ministry of Administration and Digitization in order to communicate with public organisations. It was not used in large-scale research due to the lack of electronic ePUAP address lists and difficulties connected with sending automation.

Among 302 answers as many as 29 (9.6%) contained only general data which was not directly connected with security issues. They also included the information according to which the Security Policy operating in this organisation does not allow to transfer information which may facilitate unauthorised access to the IT system. This form of reply was essential for the conducted research since it indicated that the Security Policy was enacted in the institution and its provisions are complied with.

In February 2014 as many as 10 people, accidentally encountered in the corridor of one of the institutions, were invited to participate in tests which consisted in attempts to remember random alphanumeric strings. These test were based on software specially designed for this purpose. As many as 10 strings of 4, 5, 6, 7 and 8 characters were drawn. Each string had to include at least 1 unique character from each group. These strings were common for each sample. Single string starting from the short ones were displayed one after the other to the participants of the study. Participants decided how much time they needed to remember the string. The task was to reproduce the string after it disappeared. The participant could stop the test if he or she decided that remembering strings of certain length was no longer possible. Due to the fact that these tests were conducted in the office hours and in participant's posts, all the distracting factors were included.

In March 2014 all the IT system users were sent the suggestion of a new password establishment method. A table 10x10 with 100 alphanumeric characters – one in each cell, presented in Figure 3.1, was generated. The suggestion was to print the table and use it as the basis to visualise shapes in order to establish

passwords. The suggestion also included the request for opinion which was categorised in the following way:

- a useless method,
- a useful method but I will use my own because it is simpler,
- a useful method and I will use it.

	1	2	3	4	5	6	7	8	9	10
1	m	e	B	g	s	3	]	L	2	-
2	q	F	S	e	3	,	q	,	Q	G
3	N	%	D	4	=	a	[	!	i	4
4	r	f	,	a	t	W	T	h	;	d
5	s	G	O	I	9	b	v	{	;	H
6	<	v	L	u	?	=	f	?	5	5
7	U	4	A	V	R	a	%	5	8	%
8	E	/	%	_	%	9	q	_	p	S
9	m	A	w	L	]	+	L	N	Z	d
10	p	0	<	E	k	@	7	t	)	r

Password 1: **mFDa9=%_Z**
Password 2: **,OLAVRa**
Password 3: **Qi;;58pZ**

Figure 3.1. Password establishment support table

Source: Own elaboration.

Only 79 people out of 320, which is 24.69%, followed the suggestion. All those who considered this method useless were interviewed in order to collect their remarks.

A logging prototype in the form of an application was developed (Figure 3.2) and presented to 10 respondents who expressed their negative opinions concerning the new authorisation method and to 10 respondents who preferred passwords despite the usefulness of the new method. The application generated a table 10x10 with alphanumeric characters which was unique for each user. It allowed users to select particular fields using a keyboard or a mouse. Selected fields read from left to right created a password which was transferred to the system. The password as well as the table were kept for the logging purpose. Users' opinions were collected and passwords constructed by them analysed.

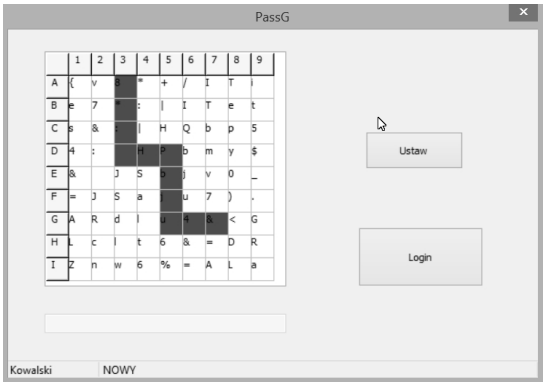

Figure 3.2. Password establishment support table

Source: Own elaboration.

3.3. Research results

The survey conducted among users allows to state that only 12% follow the provisions of the Security Policy concerning the safe password construction methods, operating in their organisation. These principles are consistent with good practices and forbid password establishment based on dates, names or words in Polish or English as well as characters arranged in a predictable order. Their aim is to prevent a quick decryption using dictionary, statistical [Herley et al., 2010] or sociotechnical attacks [Hadnagy, 2010]. Survey participants who want to use safe passwords build them on the basis of characters which come to their mind at random. Only one of them used the password generator whereas the remaining five constructed passwords on the basis of a sentence, i.e. used the first letter of each word which composed the sentence. Detailed comparison of techniques used to construct a password is presented in Table 3.2.

Table 3.2. Password establishment techniques

Declared password construction method	Quantity	%
I use password generators	1	0.55
I base my password on subsequent numbers, letters, characters (in the alphabet or on the keyboard)	37	20.44
My password includes numbers or dates important to me	13	7.18
My password includes surnames or names of my close friends and relatives	31	17.13
My password includes words	39	21.55

Declared password construction method	Quantity	%
I build passwords on the basis of sentences (poems, sayings etc.)	5	2.76
I use more than one method mentioned above at the same time	39	21.55
Random characters	16	8.84

Source: Own elaboration.

It was decided to check the influence of the declared password establishment method which was inconsistent with the security principles on the possibility to break them. In the half of the year attempts to break passwords to 254 accounts were made each month. Accounts of senior management and staff responsible for confidential data processing were omitted. The results of attempts to break passwords are presented in Figure 3.3.

The first attempt was made without prior notice to users. Over 90% of passwords were broken and in the next 5% at least one character was identified. Subsequent tests conducted after the users were notified of the weakness of their passwords and informed of the Security Policy provisions, led only to a slight improvement of the situation. A considerable decrease of the number of broken passwords was observed only in test 4 which was preceded by enabling of complexity control in Active Directory. Even then almost 65% of passwords cannot be considered safe. One of the causes is the use of an outdated LM or NTLM authentication protocol instead of the available and much safer NTLMV2 [De Ciercq, 2006].

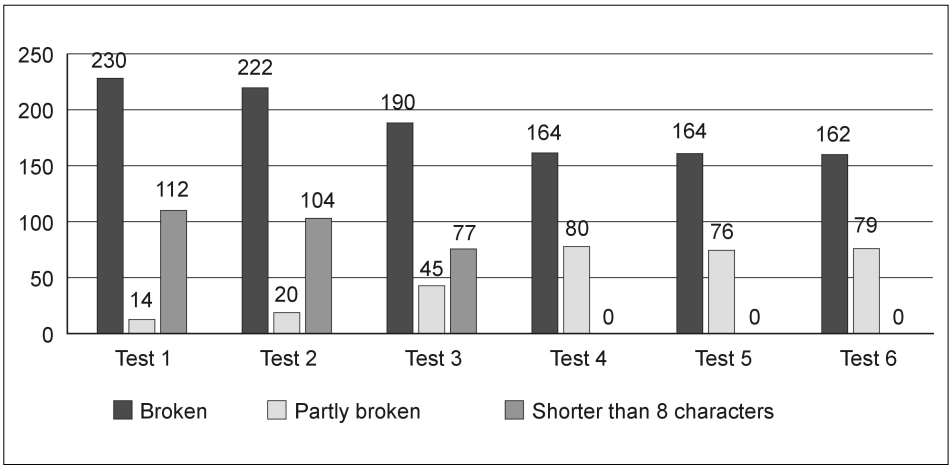

Figure 3.3. Password decryption results

Source: Own elaboration.

Complexity principles alone do not constitute a sufficient protection against hybrid attacks which use passwords taken from a dictionary and completed, preceded or interleaved with digital sequences. Thus it was decided to verify password establishment methods declared by users and check how they change them. The number of controlled accounts amounted to 132 and these were the accounts of users whose passwords, regularly changed, were broken during all 6 tests.

Users definitely include their own or their friends' and relatives' names and surnames in passwords. Only in 17% of cases it was impossible to determine the password establishment method. Detailed data is presented in Figure 3.4.

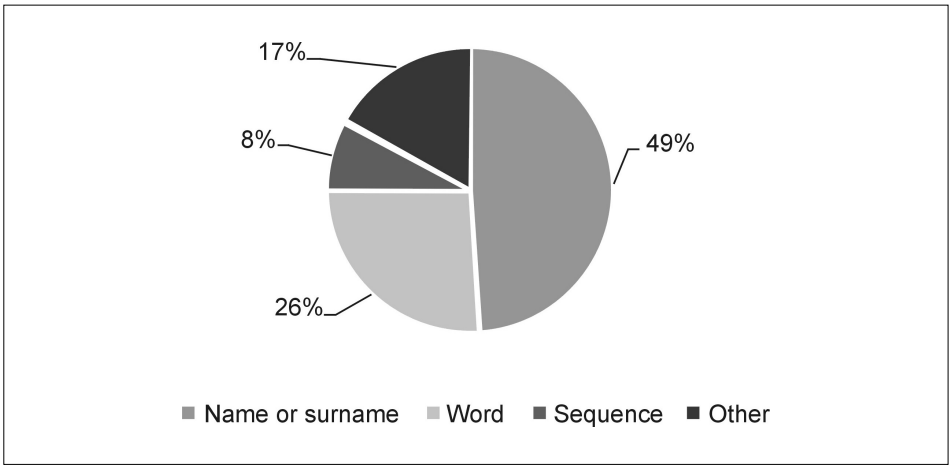

Figure 3.4. Verified password construction method

Source: Own elaboration.

As presented in Figure 3.5, users change their password by applying the prefix or suffix incrementing technique and 9% in this group use a fixed segment and complete it using different characters in a way known only to them. In 27% of cases subsequent passwords had no common features.

Security Policy and the presented evidence of password weakness do not convince users to construct safer passwords. Clerks do not identify with a problem of potential unauthorised access to data, conducted using their accounts in the IT system [Benbasat et al., 2010].

One of the causes of establishing weak passwords may be the problem to remember complex alphanumeric strings. From among 10 tested individuals only 3 people made an attempt to remember sequences composed of 8 random characters. The number of mistakes for a particular number of characters is presented in Table 3.3.

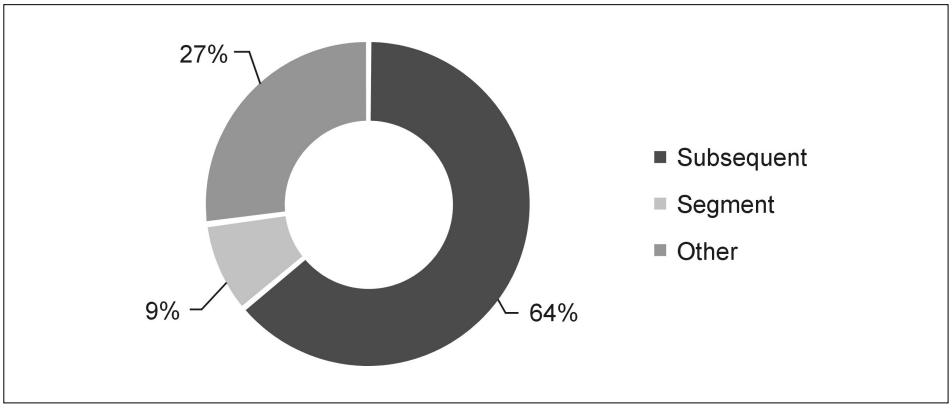

Figure 3.5. Password exchange methods

Source: Own elaboration.

Table 3.3. Incorrectly remembered alphanumeric sequences divided by their length

The number of characters	The number of attempts	% of incorrectly remembered
4	100	16.00
5	100	31.00
6	74	41.89
7	36	33.33
8	29	44.83

Source: Own elaboration.

Fewer than 70% of sequences were remembered and reproduced correctly even though the test participants were not bound by any time limits. They felt the pressure, however, because they wanted to return to their work as soon as possible. One may observe a considerable increase in the number of mistakes along with the length of the string. The 7-character strings are the exception. But in this case there is a substantial decline of the number of attempts. The test was continued only by people with unusual remembering capacity. As many as 66% of clerks admitted in the survey that their memory is not good enough and they have to use other methods in order to remember their passwords. Detailed results are presented in Table 3.4.

In view of the results presented above it was decided to check the issue of authorisation to IT systems in other self-government units. The majority – 76% use passwords as the only authorisation element. Others also use tokens, cryptographic cards and one person even uses behavioural techniques descripted in [Fred & Gamboa, 2004].

Table 3.4. Remembering passwords

Method	% of respondents
Mnemonic (associations, visualisation)	12.71
Writing down	18.23
Password construction based on information important to the user	35.36
Good memory	33.70

Source: Own elaboration.

However, it was observed that the prevalence of password usage does not entail the issue of password security. What was surprising to the researcher was the fact that some units do not have a server on the basis of which one could control the complexity principles. A considerable number of the remaining units use LM or NTLM authentication protocols. These are outdated and susceptible to cryptanalysis. They are weak due to the following reasons [Bakker, Van Der Jagt 2010]:

1. Passwords consist of maximum 14 characters (the remaining characters are omitted during control).
2. Passwords are completed with spaces to obtain 14 characters.
3. Letters in LM passwords are capitalized.
4. Passwords are stored and encrypted as two independent 7-character strings.
5. Passwords are encrypted using a symmetric algorithm with the same key for each user.

Detailed data concerning protocols used in the authorisation process in researched self-government units is presented in Table 3.5.

Table 3.5. Authentication used in the studied units

Authentication	% of studied units
Lack of authentication server	15.81
LM, NTLM	13.24
Lack of knowledge	36.76
NTLMV2, other safe	34.19

Source: Own elaboration.

An interesting fact is the lack of knowledge concerning the used algorithms observed in the studied units. It probably results from low skills of IT staff, or the fact that the staff is either overburdened with responsibilities or negligent. The argument regarding negligence is con-firmed by the results of penetration tests conducted in office information systems – the results indicated that computers used by IT specialists are most prone to threats due to the lack of update of many critical applications [Jatkiewicz, 2013].

The scope of activity of IT departments in self-government units is too wide. A considerably small staff is responsible for:

- preparation of the IT system development plans,
- procurement and settlement of contracts connected with the IT system,
- trainings,
- end user support,
- supervision of the services outsourced to third parties,
- administration and maintenance of the IT system,
- creation of new elements of the IT system,
- repairs of components of the IT system,
- IT system security management.

So one department consisting of no more than several people has both executive and supervisory functions. These require various skills, which may be difficult to obtain bearing in mind the level of salaries and scarcity of funds allocated for trainings. The role of IT specialists in these institutions is quite difficult to specify. It is difficult to state whether they are clerks or technical support. Data concerning the number of IT staff is presented in Table 3.6.

Table 3.6. The number of IT department staff

The average number of IT specialists	1.71
The median of the number of IT specialists	1
The maximum number of IT specialists	42
The minimum number of IT specialists	0

Source: Own elaboration.

As it arises from the presented information, one should rather consider IT related posts rather than departments formed only in larger units. Smaller entities, such as community councils do not employ IT specialists but tend to use services provided by external companies. Apart from performing one's duties connected with IT post, a single worker also frequently performs responsibilities of a PR specialist or OHS worker.

So if the password usage is connected with considerable problems then one should consider the possibility to use the alternative authentication methods.

The basic categories of authentication information include [Weaver, 2006]:

1) "Something you know about",
2) "Who you are",
3) "Something you own".

The technological development enabled the identification and authorisation using the biometric techniques. These belong to the category "Who you are" and are based on the measurements of unique characteristics of living organisms. The

first scientific approach to the issue was presented in 1879 by a French clerk Alphonse Bertillon. He developed the method of criminals' identification on the basis of measurements of over ten features such as height, head circumference, length of fingers etc. [Renneville, 2010].

The most well-known biometric techniques are based on measurements of [Jain et al., 2004]:

1) Fingerprint,
2) Retina,
3) Iris,
4) Unique voice characteristics,
5) Facial geometry,
6) Vein patterns,
7) DNA.

In the process of selection of biometric technology one should pay attention to the following indicators: FRR (false rejection rate) also called type I error is the percentage of times a valid subject has been falsely rejected; and FAR (false acceptance rate) also called type II errors is the percentage of times an invalid subject has been falsely accepted by the system. The crossover error rate (CER) which is the percentage of cases in which FRR equals FAR is also important.

The difference between behavioural and biometric techniques is that the former are not used to measure innate but acquired forms of behaviour. Identity verification based on the analysis of an individual's handwriting is the most natural and thus the most accepted identification method. A digital representation of the signature contains the visual characteristics as well as the dynamics of the pen movement, pressure, angle, azimuth and inclination. The number of additional parameters collected in the process depends on the device used to put one's signature (a graphics tablet) as well as the algorithm. The most advanced algorithms combine the statistical and dynamic approach, and they are characterised by low FAR, FRR and ERR indexes [Impedovo et al., 2012].

Statistical approach is based on the graphic analysis of the scanned signatures. Applied by G. Dimauro [Dimauro et al., 1992], it consists in distinguishing the characteristic features of an image after making it scale- and orientation- independent and dividing it into parts. The dynamic approach consists in the temporal analysis of signature elements in time. A case in point is the algorithm developed by T. Wessels, which uses Markow's five-dimensional observations [Omlin, Wessels, 2000].

Techniques described earlier require adequate devices whose quality influences the system security level. Thus the method based on the keystroke dynamics and mouse movements is an attractive alternative. It uses rule-based classifiers or artificial intelligence based on neural networks which enable to create the pro-

file of a mouse or keyboard user on the basis of analysis of keystroke intervals, keystroke dynamics, the speed of mouse cursor or uncoordinated but characteristic movements.

Effectiveness of this method increases with the number of registered samples, which in turn involves a large amount of computing power necessary to construct the pattern. The achieved error indicators are satisfactory (FAR 1.8%, FRR 3%) and give grounds to further analysis. Keystroke dynamics is most frequently conducted in connection with the password based authentication method. In such case not only the password compatibility but the way it is entered is tested as well. This considerably limits the number of samples necessary to generate the pattern.

A token is a cryptographic device which generates one time passwords. Thus the name frequently encountered in literature of the subject – OTP Token (One Time Password). The token activity is based on algorithms and cryptographic keys. It generates a sequence of digits using a private key on the basis of time or sequence of digits which constitute a password. The authentication system based on keys and sequence generated by the token identifies and authenticates the user.

There are both hardware and software tokens. Soft-tokens, also called virtual tokens have the same functionality as the hard tokens. They practically emulate hardware but they are much less safe because its theft may remain undisclosed. Security of a virtual token is as strong as the security of the computer on which it is installed. Therefore they are seldom used.

Modern identification systems are based on microprocessor cards whereby memory access verification through a Personal Identification Number is possible and strong cryptographic algorithms can be applied. The software for microprocessor cards is developed on the basis of special Java implementation, the so called Java Card. A Java virtual machine which handles software written for it is mounted in the card memory.

The basic authentication method in all, tested by the survey, units are passwords. Cryptographic cards and tokens are used in some organisations for a group of selected users. Biometric and behavioural techniques are used to a somewhat smaller extent. Detailed information concerning the authentication techniques is presented in Figure 3.6.

Only 17.65% of the surveyed offices would decide to completely eliminate passwords and replace them with other techniques. The majority would use cryptographic cards. Detailed information is presented in Table 3.7.

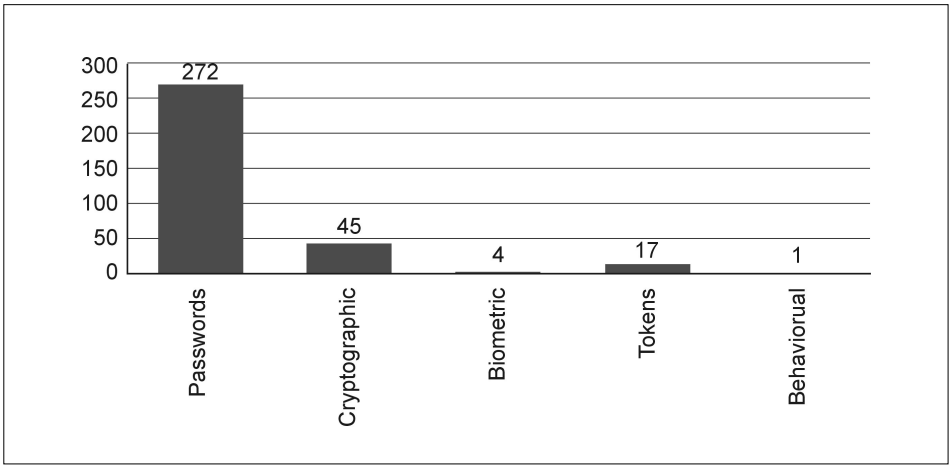

Figure 3.6. Used authentication techniques

Source: Own elaboration.

Table 3.7. The declared authentication techniques replacing passwords in the surveyed organisations

Authentication techniques	% of organisations
Cryptographic cards	13.24
Biometric techniques	1.47
Tokens	1.47
Cryptographic cards and biometric techniques	1.10
Cryptographic cards and tokens	0.37

Source: Own elaboration.

In view of the low skills of IT staff or the fact that they are overburdened with official duties, one should not expect a prompt implementation of these technologies. What is also significant is that institutions have relatively small budgets which may be allocated for the implementation of IT solutions. Data concerning the budget of analysed units collected during the study is presented in Table 3.8.

Table 3.8. Budgets allocated for IT solutions in the researched institutions

The average % of budget allocated for information technology	1.39
The median % of budget allocated for information technology	0.25
The average budget allocated for information technology	EUR 58,295.36
The median of budget allocated for information technology	EUR 12,500.00

Source: Own elaboration.

Budget amounts have been converted from Polish zlotys to euros at the average exchange rate of the National Bank of Polish 10.03.2014 (1 EUR – 4.1991 PLN).

As many as 51.93% of clerks believe that passwords constructed by users should still be used and 8.84% prefer to obtain passwords from IT specialists. Over 58% would not approve to use biometric techniques due to privacy security. In the case of behavioural techniques it amounts to 70.72%.

Conclusion

The results of research presented here lead to the conclusion that passwords are in fact the dominating authentication method in IT systems used by self-government institutions. Bearing in mind the scarcity of financial resources which offices are able to allocate for the implementation and maintenance of IT solutions, problems with IT staff as well as unwillingness on the part of the users, introduction of the popular authentication techniques available on the market does not seem realistic in the immediate future.

Thus one should focus on the main problems associated with password usage, i.e. the weakness of the used algorithms and the ability of users to establish and remember passwords which meet the security requirements.

One should remind that weak algorithms were used due to the lack of knowledge of the possibility to implement stronger ones. It is therefore concluded that skills of IT specialists in self-government institutions should be by all means verified. Education itself is no longer the sufficient evaluation criterion since the level of higher education institutions is very diverse. Producers of IT system components should in principle use the strongest cryptographic algorithms as default parameters. If they are updated it is necessary to effectively inform users about the possibility to use new options.

The process of remembering complex alphanumeric sequences which constitute safe passwords can be supported. Therefore the users were offered to use the password establishment method based on the table of characters presented in table 3.1. Collected opinions are presented in Figure 3.7.

Interviews with 10 random respondents revealed difficulties connected with reading the character on the table and finding it on the keyboard. This activity considerably prolongs the logging process and leads to numerous mistakes. What is more, a printed sheet of paper may be easily lost among a considerable number of processed documents. The process would be significantly facilitated if the tables were displayed on the screen and the characters could be selected using a mouse or a finger on the touch screen.

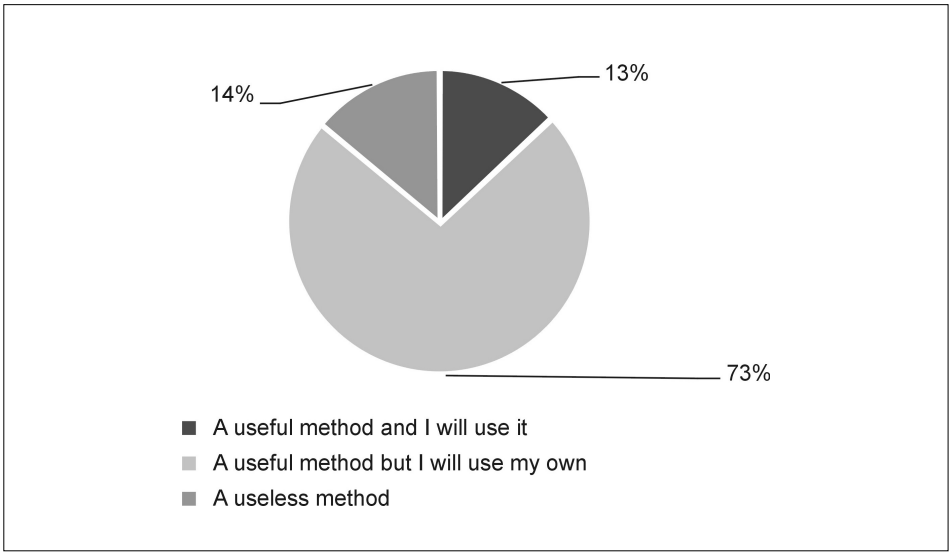

Figure 3.7. Opinions of the password establishment method based on the table of characters
Source: Own elaboration.

According to the respondents' suggestions, the authorisation system prototype in the form of the application which simulated the logging process was developed. The group of respondents who were asked to test this solution and express their opinion included 10 respondents who considered this method to be useless and 10 respondents who considered this method to be useful but preferred their own method. All testers were supposed to generate 10 passwords each and use them to log in.

None of the passwords established by them was shorter than 10 characters and the average length was 13 characters. For obvious reasons all of them complied with the strictest complexity principles by including at least 1 out of 4 types of characters. They were not based on names, surnames or words. They also did not refer in any way to the previous ones. Since the content of the table of characters was randomly selected for each user and each password change, none of the 200 passwords were repeated.

Even though the presented method is not devoid of flaws – apart from password encryption one should safely store tables of characters and it is easier to peep the shape drawn on the table – one may go as far as to say that the opinion stating that the passwords are dead is premature. Their potential may still be used and establishment as well as usage methods can be improved.

References

1. Bakker M., Van Der Jagt R. (2010), *GPU-based password cracking*, University of Amsterdam. System and Network Engineering, Amsterdam.
2. Bulgurcu B., Cavusoglu H., Benbasat I. (2010), *Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness*, "Management Information Systems Quarterly", Vol. 34, No.3.
3. Calder A. (2006), *Implementing Information Security Based on ISO 27001/ISO 17799. A Management Guide*, Van Haren Publishing.
4. Cazier J.A., Medlin B.D. (2006), *How secure is your password? An analysis of e-commerce passwords and their crack times*, "Journal of Information System Security", Vol.2, No. 3.
5. De Ciercq J. (2006), *Deflecting Active Directory Attacks. in ISSE 2006-Securing Electronic Busines Processes*, Vieweg.
6. Desmond B., Richards J., Allen R., Lowe-Norris A.G. (2008), *Active Directory: Designing, Deploying, and Running Active Directory*, O'Reilly Media.
7. Dimauro G., Impedevo S., Pirlo G. Salzo G.A. (1992), *A Multi-Expert Sig-nature Verification System for Bankcheck Processing*, "International Journal of Pattern Recognition and Artificial Intelligence".
8. Gamboa H., Fred A. (2004), *A behavioral biometric system based on human-computer interaction, in: Defense and Security*, International Society for Optics and Photonics.
9. Gosney J. (2012), *Password Cracking HPC*, In Passwords^12 Conference. Oslo.
10. Groves R.M., Fowler F.J., Couper M.P., Lepkowski J.M., Singer E., Tourangeau R. (2009), *Survey Methodology*, John Wiley&Sons Inc.
11. Hadnagy C. (2010), *Social engineering: The art of human hacking*, John Wiley & Sons.
12. Herley C., Van Oorschot P. (2012), *A research agenda acknowledging the persistence of passwords*, "Security & Privacy", Vol. 10, No. 1.
13. Impedovo D., Pirlo G., Plamondon R. (2012), *Hand-written signature verification: new advancements and open issues*, in: *Frontiers in Handwriting Recognition (ICFHR)*, 2012 International Conference on, IEEE.
14. Jain A.K., Ross A., Prabhakar S. (2004), *An introduction to biometric recognition*, "Circuits and Systems for Video Technology". IEEE Transactions on, Vol. 14, No. 1.
15. Jatkiewicz P. (2013), *Identifying Factors of an Information Security Management System of Local Self-government Bodies in Information Systems: Development, Learning, Security*, Springer Berlin Heidelberg.
16. Kothari C.R. (2014), *Research Methodology: Methods and Techniques*, New Age International (P) Ltd., Publishers.
17. Mikkonen T. (2014), *Perceptions of controllers on EU data protection reform in: A Finnish perspective*, "Computer Law & Security Review", Vol. 30, No. 2.
18. Miller G.A. (1956), *The magical number seven, plus or minus two: some limits on our capacity for processing information*, "Psychological Review", Vol. 63, No. 2.
19. Renneville M. (2010), *Alphonse Bertillon and the Identification of Persons.*
20. Schechter S., Herley C., Mitzenmacher M. (2010), *Popularity is everything: A new approach to protecting passwords from statistical-guessing attack*, in: Proceedings of the 5th USENIX conference on Hot topics in security. USENIX Association.

21. Vance A. (2010), *If your password is 123456, just make it hackme*, "The New York Times".
22. Weaver A.C. (2006), *Biometric authentication*, "Computer", Vol. 39, No. 2.
23. Wessels T., Omlin C. (2000), *A hybrid system for signature verification.* Transactions of IJCNN. Italy.

Chapter 4

Modeling emotions for affect-aware applications

Agata Kołakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, Michał R. Wróbel

Introduction

Since the dawn of civilization, philosophers reflected on the nature of emotions. Starting with Aristotle, who distinguished four humors, the Enlightenment philosophers attempted to name human emotions and moods. However, the development of psychology in the twentieth century led to a thorough study of this aspect of human life [Smith, Ellsworth, 1985]. The developed discrete emotion theory assumes that there is a small number of core emotions. Most scientists agree that happiness, anger and fear are easily recognizable and fundamentally different from each other [Smith, Ellsworth, 1985]. However human emotions are much more sophisticated than those three groups. Therefore, many affect models were created in order to describe emotional reaction of a human.

An affect-aware system can be defined as a program of any main functionality, that additionally recognizes emotional state of a user and has control mechanisms and logic able to handle the information on affect [Landowska, 2013a]. The affect-aware applications require some model for affect representation in order to automatically process emotional states of a user. Multiple psychological models of emotions have been adapted in such system designs, however there is no methodological approach to modeling emotions for software systems. This paper reviews the emotion representation models and provides examples of applications. The main research question of the paper might be formulated as follows: How to model emotions and which emotion representation models are used in affective systems design? The study is supported with insight into two cases of an affect-aware game and an affective tutoring system.

4.1. Emotion representation

There are three major model types of emotional state representation: discrete, dimensional and componential. Discrete models distinguish a set of basic emotions (named with word labels) and describe each affective state as a combination of the basic ones. A significant group of emotion recognition algorithms uses emo-

tion representation based on labels only and that kind of representation causes serious interpretation problems due to fuzziness of linking concepts with words and a problem of semantic disambiguation. Dimensional models represent an emotional state as a point in a two or three-dimensional space. Componential models use a number of factors that constitute or influence the resulting emotional state.

4.1.1. Discrete emotion representation models

At first psychological studies aimed at identifying a set of emotions that are universal among all people, independently on the ethnicity and civilization level. One of the most popular theories about basic emotions that is used until now was proposed by Paul Ekman. He is an American psychologist, one of the most influential psychologists of the 20th century, who studied emotions and their correlation with facial expressions. Ekman, who traveled to Papua New Guinea to study the facial expressions of isolated populations, discovered that some of facial expressions are cross-cultural. He identified a set of six basic, core emotions: anger, disgust, fear, happiness, sadness, and surprise [Ekman and Friesen, 1971] (Figure 4.1).

Figure 4.1. Six basic emotions according to Ekman

Source: Own elaboration.

Carroll Izard proposed a similar approach, but extended with neuropsychological aspects. His discrete emotion theory postulates ten core emotions: interest, joy, surprise, sadness, anger, disgust, contempt, fear, shame and guilt [Izard, Izard, 1977; Allen et al., 1988].

Based on these basic emotions, various scales were developed and used in different studies. However, it has led to situations that research results were incomparable. Therefore, to describe positive and negative affect Watson et al. proposed universal PANAS scale, consisting of 20 labels describing emotions (10 for positive and 10 for negative) [Watson et al., 1988]. It is widely used, and was later extended to PANAS-X scale (with 60 emotion labels) [Watson, Clark, 1999].

There is also a number of emotional scales intended to be used in specific applications, e.g. JES and JAS scales were proposed for measuring job satisfaction [Brief et al., 1988; Fisher, 1998].

The advantage of the label-based emotion representation is its comprehensibility for human and as a result, they are frequently used in surveys. However, direct application of label-based models in computer systems might be misleading. First of all, human use diverse words to describe similar emotional state, e.g. positive state after good news might be named with joy, delight, merriment, rejoicing, gladness, excitement and so on. Moreover, perceived meaning of a specific word e.g. "angry" might differ depending on culture, context and even individuals. The more sophisticated emotion label, the fuzzier might be human perception of it. This fuzziness and complexity of human emotions cannot be perceived by computer systems, in which label "anger" is context-independent.

4.1.2. Dimensional models

Circumplex model of affect, one of the most popular dimensional models, was proposed by Russell [1980]. It assumes, that any emotion might be described with two continuous dimensions of valence and arousal. The valence (pleasure) dimension differentiates positive from negative emotions, and as it is continuous, both directions might be graded. Values close to zero correspond to neutral emotional states. The dimension of arousal (activation) allows to differentiate active and passive emotional states, while a value of zero on the axis means intermediate states. In the dimensional models a specific affective state is represented as a point in dimensional space with coordinates' values ranging from –1 to 1 (sometimes –10 up to 10 or –100 up to 100 scales are used).

Whissel divided this two dimensions into quadrants, shown in Figure 4.2 [Whissell, 2009].

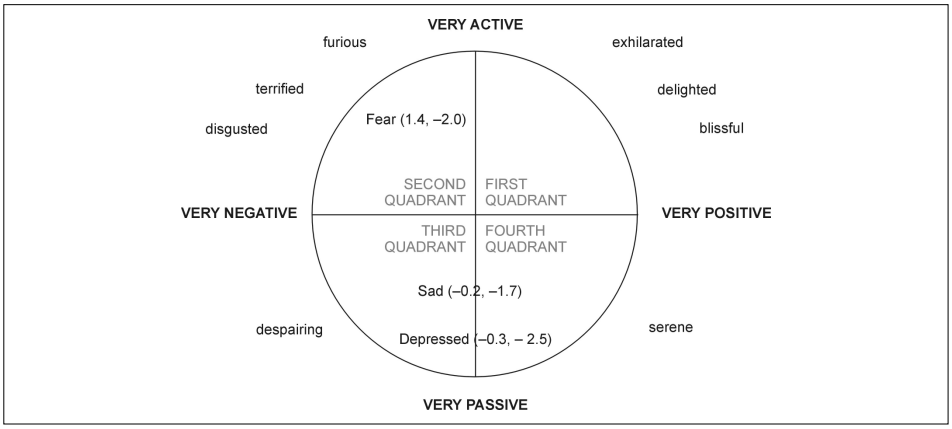

Figure 4.2. Whissell Wheel emotion representation model

Source: [Garcia-Rojas et al., 2006].

Quadrant I (positive valence and positive arousal) indicates positive and proactive emotional states such as joy, happiness, love, contentment, satisfaction. Quadrant II (–V +A) indicates negative states of high arousal, and among them are anger, disgust, surprise and fear. The third quadrant (–V –A) represents negative states of low arousal, such as boredom, sadness, up to depression states. Quadrant IV (+V –A) represents positive states of low arousal, such as relaxation, bliss, up to sleepiness.The circumplex model of affect was criticized for at least three reasons: (1) it is unnatural for human representation of emotions (people do not think about emotions as points), (2) it is hard to represent ambivalent emotional states and (3) fear and anger are undistinguishable, as these emotions both lie in the same quadrant of high arousal and negative valence.

To provide distinction between fear and anger it is necessary to add the third dimension. One of the 3D models that gained popularity in affective computing is PAD model (Pleasure-Arousal-Dominance) proposed by Mehrabian and Russell [Mehrabian, Russell, 1974; Mehrabian, 1996]. The pleasure dimension corresponds to valence in the circumplex model, and the dimension of arousal remains. The third dimension is dominance representing 'fight or escape' reaction to stimuli. The model allows to differentiate anger (positive dominance) from fear (negative dominance). The PAD model is visualized in Figure 4.3.

The PAD space might be divided into eight sub-spaces, that are named after extreme emotions represented by the extreme points of a scale [Mehrabian, 1996]:

- exuberant (+P +A +D) vs. bored (–P –A –D),
- dependent (+P +A –D) vs. disdainful (–P –A +D),
- relaxed (+P –A +D) vs. anxious (–P +A –D),
- docile (+P –A –D) vs. hostile (–P +A +D).

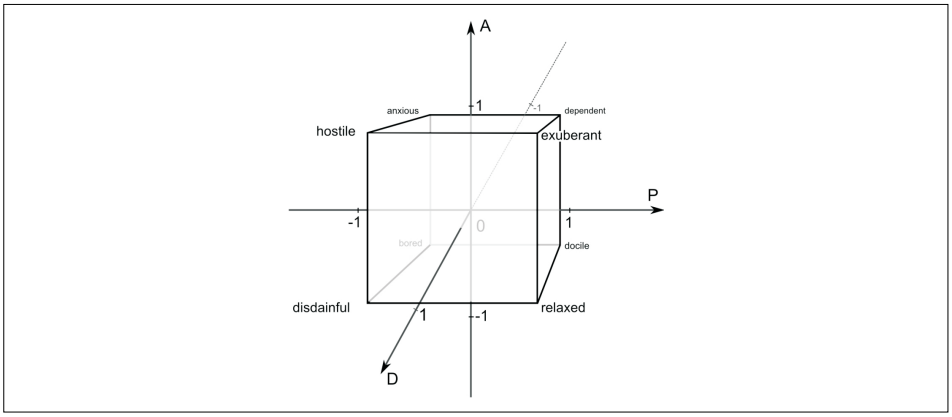

Figure 4.3. PAD emotion representation model

Source: Own elsboration.

Dimensional models with continuous scales are suitable for computer processing and their mathematical representation enables processing by algorithms for analysis or synthesis of emotions.

The lexicon of affect-annotated words that was developed by Russel and Mehrabian provides average as well as standard deviation of the assigned dimension value for each word. In some cases (for some ambiguous words) standard deviation even exceeds half of the adopted scale (0.5). There is an issue of inaccuracies and blur concepts related to representing emotions in terms of dimensional models. Categories that are used by people, often limit the possibilities of expression. This means that the term 'angry' could be assigned by different people with values that deviate from the average point (–0.51, 0.59, 0.25).

Moreover, for computer systems the difference between 0.546 and 0.547 is recognizable, but it's impossible to describe in such detail the emotional state of a person. Any application adopting the dimensional model must therefore take into account the uncertainty and fuzziness of the emotional states representation.

4.1.3. Plutchik componential model

Plutchik developed a psychoevolutionary theory of emotion and created a model of emotions where he proposed distinction between basic and complex ones [Plutchik, 2001]. Different researchers define different basic affects and Plutchik in his model proposed eight basic, fundamental emotions: joy, trust, fear, surprise, sadness, anticipation, anger, and disgust. According to his theory, other emotions are a combination or mixtures of the basic emotions, such as love is in this model is a combination of joy and trust. A mixture of any two primary emotions may be called dyad. Primary dyads, e.g. optimism=anticipation+joy, are often felt, secondary dyads, e.g. guilt=joy+fear, are sometimes felt and tertiary dyads, e.g. delight=joy+surprise, are seldom felt. Each of the basic emotion has an opposite one: the most primitive emotions surprise vs. anticipation as a respond to 'new-old' conditions, fear vs. anger as a reaction to danger, enabled to take or give up fighting, trust vs. disgust as reactions that protect against contact with something abhorrent or to motivate to do something pleasant, and the latest of evolutionary emotions joy vs. sadness, sadness cause calm and joy mobilize to contact with somebody or something [Jarymowicz, Imbir, 2010]. Each emotion may have different level of intensity or arousal.

The Plutchik model is interestingly visualized. Plutchik presented these basic affects and also other advanced, composed of the basic ones, on the color wheel of emotions (Figure 4.4). The eight sectors indicate that there are eight basic emotions arranged as four pairs of opposites. The color indicates the intensity of an emotion: the darker the shade, the more intensive the emotion is. So the intensity

of emotion increases towards the center of the wheel and decreases in the other direction. For example, the lower level of joy is serenity and the highest level of it is ecstasy. The two dimensional wheel may be visualized with 3D object where the top middle point represents neutral emotion.

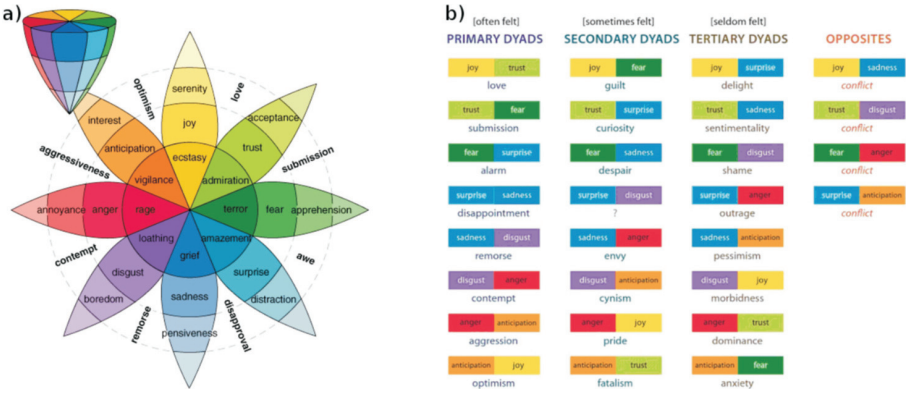

Figure 4.4. a) Plutchik's color wheel of emotions, b) dyads of emotions
Source: [Mihalinec et al., 2008].

4.1.4. OCC componential model

The OCC model proposed by Ortony, Clore and Collin in 1988 [Ortony, 1990] defines a hierarchy of 22 emotion types representing all possible states which might be experienced. In contrast to discrete models of emotions, the OCC one takes into account the process of generating emotions. Each emotion is a result of an affective reaction, which occurs after evaluating the aspects of a situation as positive (beneficial) or negative (harmful) [Ortony, 1990]. The reactions may differ depending on three factors described below.

1. Consequences of events, which are perceived as desirable or not for achieving one's goals. The consequences might be either beneficial to the goals or not and thus an affective reaction of being pleased or displeased may be generated. The emotions belonging to this category are divided into those regarding consequences for others and they form a group called *fortune-of-others* (happy-for, resentment, gloating, pity) or for self forming two groups, i.e. *prospect-based* (hope, fear, satisfaction, fears-confirmed, relief, disappointment) and *well-being* (joy, distress) emotions.

2. Actions of agents, which may result in affective reaction of approving when the actions are compatible with one's standards, or disapproving otherwise. They may evoke emotions connected with the subject's actions (pride, shame) or other agent's (admiration, reproach) and they all form a group of *attribution* emotions.

The model also takes into account the situation when two eliciting conditions occur simultaneously, i.e. when the actions of agents are evaluated in terms of their consequences. The emotions corresponding to this group called *well-being/attribution compounds* are: gratification, remorse, gratitude and anger.

3. Aspects of objects, which may be either appealing or not depending on one's attitudes and thus may result in affective reaction of liking or disliking respectively. The emotions evoked in this way are love and hate and they form a group of *attraction* emotions.

Apart from the hierarchy, the model provides precise description of each of the 22 emotional states [Steunebrink et al., 2009]. The description contains a short sentence presenting the eliciting conditions, e.g. for hope it is as follows: "pleased about the prospect of a desirable event". Then a number of synonymous words, which represent similar emotions belonging to the same group, are given, e.g. for hope these are: anticipation, excitement, expectancy. Finally variables which affect the emotion intensity are given, e.g. the variables for hope are: degree of desirability, likelihood.

Despite some ambiguities highlighted in [Steunebrink et al., 2009] for example, the OCC model is widely used to simulate affect in artificial agents [Bartneck, 2002], because it covers majority of possible situations which could be modeled in an affect-aware system.

4.2. Emotion models in applications

In a growing number of affective and affect-aware applications different emotion representation models are used. Selection of the model depends largely on the type of application, therefore in this paper two application types were analyzed: affect-aware games and tutoring systems. The studies provide more insight into the reasons for the emotion model choice.

4.2.1. Affect-aware games

Video games belong to the wide area of entertainment applications and one of their main goals is to engage players emotionally through both gameplay techniques as well as interaction with virtual agents or human players in multiplayer games [Adams, 2013]. Although video games are among some of the most natural applications of affect-aware software, not many of them seek to incorporate their players' affective state into the gameplay. Moreover, this affect-awareness is usually statically built in the gameplay at the development stage, basing on the assumed model of a so called representative player. Unfortunately, such approach

does not take into account that each player differs to a certain extent from that averaged model [Landowska, Wróbel, 2015], and more importantly, a player's affective state can change even radically from session to session making it almost impossible to predict the play-time emotions at the development stage.

In the last several years only a few truly affect-aware games have been developed, mainly as non-commercial academic projects. For example, Feed The Fish takes a player's facial expressions as input and dynamically responds to the player by changing the game elements [Obaid et al., 2008]. The goal of this system is to use human expressions to build a communication channel between the game and the players, so playing the game can be more enjoyable [Obaid et al., 2008]. In this game three emotional user states are recognized, namely *neutral*, *happy*, and *frustrated*, which switch the gameplay's difficulty, respectively, to normal, hard and easy level. Reported user study results demonstrated greater satisfaction level of the player of affect-aware version of the game that seemed to be more challenging and dynamic.

This example demonstrates that in most affect-aware video games rather simple emotional models are used which are labeled or dimensional ones. These models usually use two or three labels representing positive (desirable), negative (unwanted), and sometimes neutral emotional state of the player. This emotional state may be represented as two or three discrete labels of emotions or a float value indicator along an axis in one-dimensional emotion model.

A good example of using such approach is ErfDemo video arcade game developed under supervision of one of the paper's authors at Gdansk University of Technology [Szwoch, 2015]. The game has been designed for rather young players and the goal is to shoot as many balls as possible in a limited gameplay time, through a hoop hanging in the air at some distance. A gameplay similarity to individual basketball training, should give the player additional motivation and understanding of the gameplay challenges and actions.

As ErfDemo has been designed for young players, it assumes a very simple emotional model containing only *joy* and *frustration* emotional states, which seems to be sufficient for an arcade game with a short gameplay when there is no time for the player to feel and express a wider spectrum of emotions. ErfDemo tries to use estimated information about the player's emotional state to maintain the game's difficulty at a suitable level.

ErfDemo continuously estimates the joy level of the player on the basis of the gameplay events. It takes into account several factors, such as: the elapsing gameplay time that steadily decreases the parameter (as it increases the player's frustration), each unsuccessful throw or dropping of a ball (that have the same effect), and each successful throw that increases the player's joy. This simple behav-

ioral approach can be easily extended by any other mechanism that allows for reasoning about the player's emotional state.

Although the emotion representation model chosen for ErfDemo game was very simplistic, it was sufficient to provide more enjoyment during gameplay. The model seems accurate for simple task games. If the game would be extended with automatic emotion recognition e.g. from facial analysis, mapping between the recognized model and the one used in game logic must be found.

4.2.2. Affective tutoring systems

The beginnings of affective computing are associated with computer-assisted learning. Therefore, a number of affective applications have been developed in e-learning domain. One of the examples is EMASPEL – affective e-learning framework. It gathers emotional information based on facial tracking. Emotional classification is based on the variation of distances from the neutral face and is based on the six basic universal emotions of Ekman [Ammar et al., 2010]. Another framework, proposed by Nkambou, uses different discrete model with five emotions: *satisfaction*, *confidence*, *surprise*, *confusion* and *frustration* [Brief et al., 1988].

Affective tutoring systems (ATS), which are tutoring applications that recognize learner's affective state and are able to react to it [Landowska, 2014], also use different emotion representation. For example Easy with Eve [Alexander et al., 2006], AutoTutor [D'Mello et al., 2008] and Fermat [Cabada et al., 2012] use their own discrete models, each one with different set of emotions. On the other hand Alepis et al. [Alepis, Virvou, 2011] and Hernández et al. [Hernández et al., 2009] used in their ATS OCC emotion model.

Gerda bot is a prototype of a conversational Intelligent Tutoring System developed under supervision of one of the paper authors at Gdansk University of Technology. Gerda questions students on operating systems and uses a metaphor of student-teacher conversation during oral examination. Gerda can be classified as an authoritative teacher [Landowska, 2013b].

There are some studies on how emotional states influence education and several general rules regarding the influence on learning processes can be derived from literature [Landowska, 2013b]. Emotional states of very high or very low arousal (both positive and negative valence) disturb learning processes and especially boredom is an emotional state that should be avoided in learning [Elliott et al., 1999]. Educational processes are supported by the states of engagement, concentration and flow [Picard, Klein, 2002; Baker, 2007] and emotional states with a higher dominance factor support the learning process (moderate anger is better than fear in the educational environment) [Ammar et al., 2010].

Taking the above into account, *Gerda* was designed to evoke high arousal emotions which are rather positive (or possibly slightly negative) and avoid boring the user. As textual analysis seems to make the main contribution to the emotional state recognition process in Gerda, selection of the emotion representation model was limited, as most of the sentiment analysis tools use dictionaries labeled with dimensional models.

PAD representation model was chosen for Gerda ITS and it was consequently used in emotion recognition, interpretation and control.

The GERDA emotional model development revealed two issues: missing algorithms for mapping between models and lack of representation of uncertainty for emotion recognition. From the three emotion recognition algorithms implemented in Gerda two used PAD as their output, while one provided a hypothesis with Ekman's six basic emotions model. An additional wrapper module was implemented that transformed the Ekman's output to a PAD point, however accuracy of the mapping was low, resulting in imprecise emotional state estimation and increased uncertainty.

Conclusion

Granularity of emotion representation can be defined as a level of detail considered in emotion model and can be measured with the number of distinguishable emotional state classes. Granularity and accuracy of emotion recognition are at least partially interchangeable and increase of both simultaneously is nowadays an open research issue. However, the challenge is not only to improve granularity and accuracy, but also to distinguish the emotional states, that are the most important from the perspective of the system application domain. Nowadays multiple diverse models are used and one observation might be made: the simpler the task, the simpler emotion model might be used. However, in some more complex applications, the chosen representation model might be a compromise between what is required and what is technically possible in the system context.

Both studies of games and affective tutoring systems revealed that there is a challenge of providing adequate mapping between models of emotion representation, especially if multiple recognition modalities are used in the same application. Another challenge concerning emotion models include representation of fuzziness and uncertainty related to recognized affective concept.

Practical implications for affect-aware applications include the following: one common representation model shall be used for affect recognition and interpretation, fuzziness of affective states should be also expressed and interpreted and

there is a need for an explicit definition of a distinguishable set of affective states that are under investigation for a specific tool or task.

Authors are aware of the fact, that this study is not free of some limitations, including: arbitrarily chosen case studies and subjectivity in their description. The purpose of the paper however was to depict diversity of emotion models, that are used in practice of affect-aware software development as well as identification of the main challenges in affect modeling.

References

1. Adams E. (2013), *Fundamentals of game design*, Pearson Education.
2. Alepis E., Virvou M. (2011), *Automatic generation of emotions in tutoring agents for affective e-learning in medical education*, "Expert Systems with Applications" Vol. 38.
3. Alexander S., Sarrafzadeh A., Hill S. (2006), *Easy with eve: A functional affective tutoring system*, in: Workshop on Motivational and Affective Issues in ITS. 8th International Conference on ITS.
4. Allen C.T., Machleit K.A., Marine S.S. (1988), *On assessing the emotionality of advertising via Izard's differential emotions scale*, "Advances in Consumer Research" Vol. 15.
5. Ammar M.B., Neji M., Alimi A.M., Gouardères G. (2010), *The affective tutoring system*, "Expert Systems with Applications" Vol. 37, Issue 4.
6. Baker R.S.J. (2007), *Modeling and understanding students' off-task behavior in intelligent tutoring systems*, in: Proceedings of the SIGCHI conference on Human factors in computing systems.
7. Bartneck C. (2002), *Integrating the occ model of emotions in embodied characters*, in: Workshop on Virtual Conversational Characters.
8. Brief A.P., Burke M.J., George J.M. (1988), *Should negative affectivity remain an unmeasured variable in the study of job stress?*, "Journal of Applied Psychology" Vol. 73.
9. Cabada R.Z., Estrada M.L.B., Beltr'n J.A. (2012), *Fermat: merging affective tutoring systems with learning social networks*, in: Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on.
10. D'Mello S., Jackson T., Craig S. (2008), *AutoTutor detects and responds to learners affective and cognitive states*, in: Workshop on Emotional and Cognitive Issues at the International Conference on Intelligent Tutoring Systems.
11. Ekman P., Friesen W.V. (1971), *Constants across cultures in the face and emotion*, "Journal of Personality and Social Psychology" Vol. 17.
12. Elliott C., Rickel J., Lester J. (1999), *Lifelike pedagogical agents and affective computing: An exploratory synthesis*, in: Artificial intelligence today. Springer.
13. Fisher C.D. (1998), *Mood and emotions while working-missing pieces of job satisfaction*, School of Business Discussion Papers 64.
14. Garcia-Rojas A., Vexo F., Thalmann D. (2006), *Emotional body expression parameters in virtual human ontology*, in: Proceedings of 1st Int. Workshop on Shapes and Semantics.

15. Hernández Y., Arroyo G., Sucar L.E. (2009), *Obtaining Teachers' Expertise to Refine an Affective Model in an Intelligent Tutor for Learning Robotics*, in: Artificial Intelligence, MICAI 2009. Eighth Mexican International Conference.
16. Izard C.E., Izard C.E. (1977), *Human emotions*, Plenum Press New York
17. Jarymowicz M., Imbir K. (2010), *Towards a Taxonomy of Human Emotions*, "Przegląd Psychologiczny", nr 53.
18. Landowska A. (2013a), *Affect-awareness framework for intelligent tutoring systems*, in: The 6th International Conference on Human System Interaction (HSI).
19. Landowska A. (2013b), *Affective computing and affective learning – methods, tools and prospects*, EduAction. Electronic education magazine.
20. Landowska A. (2014), *Affective Learning Manifesto – 10 Years Later*, In: Proceedings Of The 13th European Conference On E-Learning.
21. Landowska A., Wróbel M.R. (2015), *Affective reactions to playing digital games*, in: Proceedings of the 8th International Conference on Human System Interaction.
22. Mehrabian A. (1996), *Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament*, "Current Psychology" Vol. 14.
23. Mehrabian A., Russell J.A. (1974), *The basic emotional impact of environments*, "Perceptual and Motor Skills" Vol. 38.
24. Mihalinec I., Stevanovic R., Skala K. (2008), *Emotion recognition and its aesthetic interpretation*, in: ELMAR, 2008. 50th International Symposium.
25. Obaid M., Han C., Billinghurst M. (2008), *Feed the Fish: an affect-aware game*, in: Proceedings of the 5th Australasian Conference on Interactive Entertainment.
26. Ortony A. (1990), *The cognitive structure of emotions*, Cambridge University Press.
27. Picard R.W., Klein J. (2002), *Computers that recognise and respond to user emotion: theoretical and practical implications*, "Interacting with Computers" Vol. 14.
28. Plutchik R. (2001), *The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice*, "American Scientist" Vol. 89.
29. Russell J.A. (1980), *A circumplex model of affect*, "Journal of Personality and Social Psychology" Vol. 39.
30. Smith C.A., Ellsworth P.C. (1985), *Patterns of cognitive appraisal in emotion*, "Journal of Personality and Social Psychology" Vol. 48.
31. Steunebrink B.R., Dastani M., Meyer J-J.C. (2009), *The OCC model revisited*, in: Proc. of the 4th Workshop on Emotion and Computing.
32. Szwoch M. (2015), *Design Elements of Affect Aware Video Games*, In: Proceedings of the International Conference Multimedia, Interaction, Design and Innovation.
33. Watson D., Clark L.A. (1999), *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*, University of Iowa.
34. Watson D., Clark L.A., Tellegen A. (1988), *Development and validation of brief measures of positive and negative affect: the PANAS scales*, "Journal of Personality and Social Psychology" Vol. 54.
35. Whissell C. (2009), *Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language*, "Psychological Reports" Vol. 105.

Acknowledgements

This study is supported in part by the Polish-Norwegian Research Programme operated by the National Centre for Research and Development under the Norwegian Financial Mechanism 2009–2014 in the frame of Project Contract No Pol-Nor/210629/51/2013 and Project Contract No Pol-Nor/209260/108/2015, as well as DS Funds of ETI Faculty, Gdansk University of Technology.

Chapter 5

State of consumerization in Polish IT companies

Dariusz Kralewski

Introduction

Consumerization is one of the most important topics for employers and employees in recent times. On the one hand it is a derivative of a rapidly changing ICT ecosystem based on four pillars: mobility, business analytics / big data, cloud computing solutions and social networking applications. On the other consumerization is a phenomenon associated with a completely new IT user role – employee of the company.

Increasingly, employees want to use the corporate IT resources in a natural, intuitive way as they do when buying at online auctions, leading the bank account. Consumerization refers not only employees, but also the IT professionals who want to buy and deploy IT solutions in a way that reflects their experiences as consumers. The most visible today is the phenomenon BYOD-to-work (using their own devices for business purposes), which affects organizations across all sectors, both private and public. It has a massive impact for both users and IT professionals [Bless et al., 2010].

5.1. Review of background

Noticeable is the growing acceptance of BYOD strategy with the explosion of mobile devices and applications. Some organizations have accepted BYOD in order to pro-mote mobile access to new applications. In other cases, IT managers responded to the growing expectations of their devices, allowing users to access corporate re-sources. In addition, IT managers notice the indisputable trends: the growing number of mobile applications that serve as a tool to collect data, but also as a platform for their analysis, which will be associated with the growing demand for security solutions. There is a need to develop management strategies of mobile devices running on multiple platforms and sharing business applications. It is connected with the need to ensure adequate safety parameters. BYOD-related policies can not be created once and for all. They must be checked and changed with the changing patterns of use of devices and the associated risks.

5.1.1. Consumerization of IT

Smartphones and tablets changed the lives of consumers by providing access to in-formation and communication almost always and everywhere. Used solutions are easy to use, inexpensive, fast and easily accessible. In contrast, enterprise solutions are often slow, expensive, and only understandable for experts. Employees use corporate computers and applications every day, but they would like to work in the same way as on tablets. Personal experiences change expectations of employees in relation to enterprise solutions. IT managers expect that employees will be more "aware of the technology", while the users become "consumer's aware of the technology". They do not want to understand, to know the technology to use it. Available to consumers solutions based on mobility and the cloud are in many cases much more convenient than provided by IT departments [Blount, 2011]. Speed, good price and ease of use technologies based on consumer solutions is often a flashpoint in relations with IT departments. The combination of shorter (than in the case of corporate solutions) innovation lifetime (12 months compared to 36) with an increasing number of mobile devices becomes a major concern of IT departments. The problem will not diminish. IDC estimates that in 2015, nearly 55% of the devices used in enterprises will belong to employees [IDG Enterprise Consumerization of IT].

Both the volume and the variety of types of devices affects the need to develop a good practice to deal with their presence in the corporate infrastructures. Everything started from working on iOS and Android smartphones. Later developed tablet – originally conceived as a purely consumer device, quickly found its place in enterprises. The increase in the number of devices associated with the rapidly growing market of mobile applications, which are installed by thousands of users every day. Users expect that similarly easy to install and use applications will be offered by employers. This will create a strong link between the private mobile devices with corporate infra-structure. The growing number and diversity of devices is a challenge for IT departments, which should provide adequate support. IT sees the need to limit the types of devices and operating systems supported by the company to ensure consistency in the application and effective support. It should also be noted that the growing number of devices and applications is rising costs, the reduction of which will require exert pressure on suppliers of applications and telecoms, changes in business models and billings between companies. Consequently, some costs may be passed on to the users themselves [Kvanzant].

Virtualization increasingly applies to mobile devices and desktops, especially with the development of tools for managing virtualized environments and the increasing level of security that may be provided. What's more, mobile hypervisor can effectively separate professional from private areas by creating multiple in-

stances. In some operating systems appears the ability to run software that "changes" private device into company device with an appropriate level of security and remote management capabilities. You can limit corporate resources for selected applications and provide additional data security by their separation from the private user area. Many organizations, particularly sensitive to issues of data security would be interested in the possibility of complete management employee's device, including the possibility of their exclusion or clear the data. Employees also would like to have a sense of separation between their private data from corporate data, but they often worry too much interference in their own device [Stangarone, 2014].

A very important issue is the safety of BYOD. Many IT leaders claim that security issues for mobile devices are the same as for traditional infrastructure. Persons responsible for safety, seek to ensure full control over the entire circulation of information (end-to-end). However, wireless connectivity, mobile solutions, loss of equipment, installed consumer applications represent an additional source of risk. Despite the new developments, it is difficult to secure mobile infrastructure as effectively as a traditional model based on the desktops. New technologies and mobile solutions also make it difficult. For example, mobile payments, near field communications technology is a source of new threats, especially given the low level of utilization of mobile antimalware applications.

5.1.2. Management challenges

Support for employee's devices is a serious problem. Used technologies are often far from corporation's technology and are associated with the need of support a variety of users, usage models, devices and operating systems. An additional challenge is to manage the whole process of maintaining. Also, user expectations regarding the level of support offered are often too prohibitive [Golden].

Personal mobile phones are usually purchased by the customer and the support for devices is provided by the telephone company. With the development of applications installed on smartphones and tablets, IT increasingly becoming responsible for a complete solution and must take on the entire solution together with the relationship with the operator and software vendors. Relationship management with multiple suppliers becomes another difficult challenge [Moschella, 2011].

There is a need to develop new practices and tools for mobility management. At the same time the organizations that are trying to develop new models for the mobility management of IT area have to face the ever-growing expectations of users.

Despite the difficult situation, most IT leaders now understand the need for solutions related to the presence of employee's devices at company. Many companies understand that the prohibitions and attempts to prevent employees use their own devices does not make sense, since users will still be doing it, so it is better have some control over it. The key to success is to find a balance between continually changing expectations of employees and the necessary discipline when it comes to safety and efficient management of IT in the company.

5.1.3. BYOD policy

The great diversity of devices and applications with different responsibilities of employees (which is associated with varying access to data and systems organization) is a need to create a very inhomogeneous BYOD policy. In addition, the constant changes in the area of mobile solutions in the company means that such rules must be flexible enough so that they can be easily and continuously adapted to the needs of the company. The development of such a policy should include:

1. Overview of the current situation, the use of mobile devices and study expectations of employees in this area, including those who are not traditionally seen as mobile workers.
2. Development of user profiles which take into account: duties, the degree of mobility, requirements for communication and access to information, needs related to devices and applications necessary to work the necessary level of safety, the employee's current rules and regulations.
3. Matching profiles for different degrees of security and mobility solutions (equipment, applications, regulation, control, support).
4. Creating a set of rules pertaining to the use of mobile solutions and privacy. It is one of the most difficult parts, because on the one hand should take into ac-count the satisfaction wide range of users, on the other hand, refer to individual, pre-defined profiles.
5. Estimate the individual costs (purchase of equipment, maintenance, telecom services) and determine their accounting model based on user profile.

5.1.4. Secure development

As indicated above BYOD has many benefits associated with the need to liberalize the time and place of work and more efficient use of IT resources. However, companies often require higher level of security. Therefore, you should develop safety management policies for operating mobile devices and applications [Global Corporate IT Security Risks, 2013].

Mobile applications, including corporate, are often operated in different types of networks. They are often corporate networks, after hours home networks, sometimes used are public hotspots. In many cases, users do not have access to the secure net-work. Encryption of data transmitted by mobile devices should therefore become a good practice applications created [Thompson].

Mobility threatens to put many duplicates sensitive data on different devices in multiple locations. Information should be obtained from the corporate server only when they are necessary for the work-related operation. At the end of the work they should be automatically and immediately disappear from the user equipment. At the time when we need to store sensitive data on the client device it should be absolutely encrypted. At this point, we should develop a properly functioning keystore [Yu, 2013].

5.2. Method

A necessary condition for accurate and reliable research, next to correctly formulated problems and hypotheses is the choice of appropriate methods, techniques and re-search tools. Methods and research tools are determined by the research problem. In this study, a diagnostic survey method was used. Diagnostic survey is a "public opinion poll of collecting answers to the same questions asked specific group of people" [Dunaj, 1996]. According to T. Pilch "method of diagnostic survey is a way of gathering knowledge about the structural and functional attributes and dynamics of social phenomena, opinions and views of selected communities, escalating and directions of development of certain phenomena and any other not localized institutionally phenomena – have educational significance – on the basis of specially selected group representing the general population in which examined the phenomenon occurs" [Pilch, 1998].

The method determines and defines the selection of appropriate research techniques. For the purposes of this study were selected survey. W. Zaczyński defines a survey as a direct method of getting information through questions posed to selected individuals through a printed list of questions called the questionnaire [Zaczyński, 1995]. According to W. Okoń survey is study of mass phenomena based on properly prepared questionnaire [Okoń, 1998]. In this paper, a tool that was used was a questionnaire addressed to employees responsible for the IT department.

Work began by studying literature. On this basis the theoretical basis for research was created. The next step was the selection of methods, techniques and preparation of a research tool – a questionnaire. The research was conducted be-

tween December 2014 and February 2015. For the focus group, 56 employees of the companies responsible for the IT department were selected, all working in various Polish organizations. A mixture of private and public sector employees were chosen for wider variety of responses.

According to the literature, the survey should be short and straightforward, but questions should refer to each other. Best to include not more than ten questions and not spend more than three minutes [Kauf, Tłuczak, 2013]. Created survey contained 10 questions – 3 questions about the respondent (age, sex, type of ownership) and 7 research questions. The answer to the research questions required circumscribe yes / no. Based on the constructs that came out from the literature review, the research identified seven areas of investigation:

- Is known the concept of consumerization and BYOD?
- Have you noticed BYOD phenomenon in the company?
- Does the company developed BYOD management policy?
- In the current year are you planning to introduce management of BYOD?
- Do you plan to introduce security restrictions associated with the introduction of BYOD?
- Do you think safety restrictions of BYOD take effect?
- Are you going to implement MDM system?

Participant demographics are as below:

- The male respondents were more than females (95% versus 5%);
- 79% responses represented private sector in Poland; 21% represented public sectors;
- Majority of the respondents were older than 30 years (77% versus 23%).

5.3. Results

According to the forecasts resulting from the Cisco Visual Networking Index [Cisco Visual Networking Index], in 2017 half of the traffic network in the world will come precisely from mobile devices. This statistic is conducive to strengthening the BYOD trend which for the business it is no longer an alternative. BYOD has become a necessity, for which companies need to be prepared. Security in this communication model is facing new threats that companies must answer relevant activities. The responsibility of IT departments in companies is focused mainly on effective and stable management of terminal devices and their respective securing. Until recently, these were standardized computers and software. Advancing IT Consumerization and reaching for her BYOD trend meant that to technological infrastructure of companies began to penetrate different outfits. Bring Your Own

Device giving employees the freedom to decide on the selection of equipment used for professional purposes permanently changed the dimension of communication client – the service provider. On the one hand, the relationship in the new formula has become more mobile, unlimited place and time, actually permanently available. On the other, generating a lot of risks, of which, unfortunately, Polish companies are realizing only slightly.

At first, the basic question about the meaning of the term consumerization and BYOD only one of the respondents replied that he had never heard of it. Thus, up to 98% of the respondents are aware of the phenomenon of consumerization in the company. This one respondent was excluded from the rest of the questions.

This is the inevitable consequence of the increase in the number of mobile devices: smartphones, tablets, notebooks and process, which, as it must be assumed, will act simultaneously to the development of mobility. The efforts of IT departments should, therefore concentrate on choosing the correct procedures governing the use of BYOD in terms of security, so that the model could be the most beneficial for the business.

Therefore surprising was the answer to the third question relating to the management BYOD. On this question, only two companies (4%) responded that manage mobile devices in the company. Thus, despite awareness of consumerization and BYOD as much as 96% of Polish companies surveyed had not taken any steps to consciously manage applications and data on mobile devices.

The next question was – whether in the current, 2015 are you planning to introduce management of BYOD? The question is answered only respondents who had not introduced BYOD management. On this question, 34% of respondents answered positively. Thus, 68% of companies despite awareness BYOD year does not plan to manage it. This phenomenon is particularly worrying bearing in mind that enterprise mobility will continue to be one of the hottest topics in IT, and high on the list of priorities for all CIOs, according to Ovum's 2015 Trends to Watch: Enterprise Mobility report[Trends to Watch]. Our results are inconsistent with the results of a global survey

Preparing for security of companies to implement BYOD is insufficient. The study Global Corporate IT Security Risks in 2013 revealed that only one in eight companies confirms fully implemented mobile security policy for network. The statistics are even more alarming when you consider at the same time increasing cases of unauthorized interference with mobile phones and tablets in the absence in companies of mitigation procedures use BYOD. Mobile devices have become one of the critical data leakage channels. They cause now greater loss of confidential resources than the hosting attacks, employee fraud or corporate spying.

Despite the threat to the security of IT infrastructure posed by BYOD environment, most companies do not plan to introduce any restrictions associated with

it (78%), and a large part recognizes that such restrictions would not have served effect (82%).

Permanent limitation of companies in BYOD model is the lack of permission to full interference in the terminal employee. Consequently, this requires the application of relevant security to enterprise data stored on it and what is no less important – to protect the digital identity of the user. Despite the growing number of viruses and attacks aimed at the mobile device via the network, the most critical threat to the security of confidential data is the physical loss of the device as a result of theft or negligence of users. In the event of such incidents security policy should include appropriate control procedures of passwords, blocking its capabilities, data encryption, remote delete sensitive information. It is important to find systematic monitoring which corporate resources stored on those and constant replenishment of awareness of employees, how to manage access to them. Equally important it is whether the information stored on the device are available online or are also cached for permanent use. Be-yond doubt the priority is to provide the ability to remotely delete data if it is stolen or misplaced equipment. There are very simple, but still insufficiently used ways to in-crease the safe use of mobile devices, such as: the use of a password or PIN with maintaining their confidentiality and to install antivirus software on their devices.

Full range of mobile device management provide MDM systems - Mobile Device Management. They are responsible for activation on a mobile device suitable safety program, allowing among other things, restrict access to device with a password, to clear illegal software and install the appropriate, block or run selected functions, and finally remotely delete sensitive data. Among the companies which intend to implement management of BYOD as much as 88% want to implement MDM system. All companies that are planning to develop a policy for managing BYOD, plan to implement a MDM system.

5.4. Limitations on the research design and material

Quite a few limitations have been identified with this research.

1. The research has been conducted by a single researcher. Hence it limits the analysis to the perception of one individual and does not fulfil an 'investigator triangulation' [Denzin, 1970].
2. The research deals with limited data, collected from only 56 participants within a short timeframe of 3 months. Hence it does not fulfil a complete 'data-triangulation' as suggested by Denzin [1970].

Conclusion

BYOD shows the challenges faced by companies wishing to enforce rules for the use of their personal devices for business purposes. This trend also shows that the control by IT departments on where corporate data is stored and how it is obtained access to them is still very critical area. Indispensable for the maintenance of security BYOD implementation is also education of employees regarding the risks and the consequences resulting therefrom. The fear of weakening security certainly slows down the expansion of BYOD, however, can not withdraw it. Polish companies are poorly implement consumerization. A high percentage of knowledge about BYOD and visibility implementation of BYOD in companies does not move directly into new management BYOD procedures. Polish IT companies mostly do not intend to introduce restrictions on security to mobile devices because they think that it will not be efficient.

References

1. Bless E., Alanson M., Noble C. (2010), *Consumerization: What is in Store for IT?*, http://i.dell.com/sites/content/business/solutions/whitepapers/en/Documents/intel-imrconsumerization-wp.pdf.
2. Blount S. (2011), *The consumerization of IT: security challenges of the new world order*, http://www.ca.com/us/~/media/Files/TechnologyBriefs/Consumerization-of-IT-TechBrief.pdf.
3. Cisco Visual Networking Index, http://newsroom.cisco.com/release/1578507/Cisco-Visual-Networking-Index-VNI-Mobile-Forecast-Projects-Nearly-10-fold-Global-Mobile-Data-Traffic-Growth-Over-Next-Five-Years?utm_medium=rss.
4. Denzin N. (1970), *Sociological Methods: A Sourcebook. Aldine Transaction*: United States.
5. Dunaj B. (ed.) (1966), *Dictionary of Contemporary Polish language*, Warsaw.
6. Global Corporate IT Security Risks (2013), http://media.kaspersky.com/en/business-security/Kaspersky_Global_IT_Security_Risks_Survey_report_Eng_final.pdf.
7. Golden B., *What 'consumerization of IT' really means to CIOs*, http://www.itworldcanada.com/news/whatconsumerization-of-it-really-means-to-cios/143749.
8. IDG Enterprise Consumerization of IT in the Enterprise Study 2014.
9. Kauf S., Tłuczak A. (2013), *Methods and techniques of surveys on the example transportation behaviors.*
10. Kvanzant, *Consumerization of IT*, Part One, http://whiteboard.solarwinds.com/2009/10/29/consumerization-of-it-part-one.aspx.
11. Moschella D., *What the Consumerization of IT means to your business, ten messages for CXOs*, http://lef.csc.com/blog/post/2011/06/what-the-emconsumerization-of-item-means-toyour-business-ten-messages-for-cxos.
12. Okoń W. (1998), *New sociological dictionary*, Warsaw.

13. Pilch T. (1998), *The rules of sociological research*, Warsaw.
14. Stangarone J., *Six Key Enterprise Application Development Trends of 2014*, http://www.mrc-productivity.com/blog/2014/01/6-key-enterprise-application-development-trends-of-2014/.
15. Thompson H.H., *Consumerization and Security, Effective Security Practices Series*, http://download.microsoft.com/download/E/F/9/EF9F24B7-DB49-44D4-8F6AA49D5020B8B8/Consumerization_Final.pdf.
16. *Trends to Watch: Enterprise Mobility report*, http://www.ovum.com/press_releases/ovum-predicts-enterprise-mobility-to-be-top-of-the-cio-agenda-in-2015/.
17. Yu W.E. (2013), *BYOD Security Considerations of Full Mobility and Third-party Cloud Computing*, "ISACA Journal", Vol. 1.

Chapter 6

Big Data quality issues regarding unstructured data analysis. A case study

Jacek Maślankowski

Introduction

In recent years Big Data has been explored in many ways by scientists, giving opportunities to make a fast and reliable analysis of large data sets. However in several research papers we can find that the most valuable information is usually stored in unstructured datasets. The aim of this chapter is to show crucial points of Big Data analysis based on unstructured data sources. This involves assessment of data quality attributes such as relevance, coherence or accuracy, relating to the key aspects of Big Data dynamics: data acquisition, data processing and producing data results. The case study used in this chapter demonstrates that the data quality can be identified in the first stage of the analysis. To ensure that the quality of the data is enough, the recommendation is to assess Big Data quality in several phases, described in the third part of this chapter. The hypothesis used in this chapter is as follows: not all data quality attributes are relevant to unstructured data analysis using Big Data.

The chapter is divided into three parts. After introduction, in the first part of the chapter, Big Data is shown in several aspects of unstructured data analysis. In the second part the theoretical aspects of data quality relating to Big Data are briefly discussed. The third part shows the case study and identifies key attributes of Big Data quality and its wages proposed by author. Last part shows conclusions.

The goal of the chapter is to show practical aspects of data quality assessment. The hypothesis used in the chapter is as follows: Big Data quality can be mostly assessed automatically during first phase of processing the data source.

6.1. Big Data and unstructured data analysis

More recently Big Data have been used to describe the data sets and analytical techniques in applications that are so large and complex (including unstructured social media data) that they require advanced and individual data storage, management and analysis [Chen et al., 2012]. It is said that about 90% of information is stored in unstructured data sets [Gang-Hoon et al., 2014].

In this article Big Data is perceived as an alternative way of conducting business and marketing analysis based on unstructured data analysis. It is possible as emphasis on statistical significance is not so important for many people in such type of analysis [Kennedy et al., 2014]. Moreover, Big Data analysis are rather based on existing correlation between data than on statistical evidence of data [Mantelero, 2014]. Nowadays Big Data platforms such as Google, Amazon, Apple, Facebook, and others that accumulate ever increasing information on consumer behaviour, interests, and needs [Newman, 2014]. Very valuable data source can be microblog portal [Hua-Ping et al., 2014]. There are also an open data initiatives which makes data acquisition much easy and accessible. Open data is based on the idea that some kinds of data should exist beyond the limits of copyright, patents, censorship or any other parameters often placed around data [Bertot et al., 2014].

Before we can implement Big Data ecosystem in organisation we have to prepare it for Big Data Adoption, which can be shortly describe as fitting Big Data technology to the requirements of organisation [Kwon et al., 2014]. The typical ecosystem of Big Data analysis consists of the software necessary to process Big Data in traditional data feeding chain. This chain includes data acquisition, gathering and processing [Chen et al., 2014].

In literature there are several different terms used to describe data analysis using Big Data. For instance, some researchers introduced a new term: big social data for social data analysis using Big Data [Cote, 2014]. But for most of the types of data, Big Data analysis can be regarded as a data-intensive scientific discovery (DISD) [Chen, Zhang, 2014]. To describe text analysis using Big Data we can use a term Big Text Data [Kim et al., 2014]. A person who is responsible for Big Data analysis is called data scientist [Santaferraro, 2013].

6.2. The importance of data quality

The data quality in data processing is the key issue when trying to analyse unstructured data set. There are numerous articles regarding this topic. However not many of them are trying to identify key issues of data quality in Big Data terms.

The question is how we can improve data quality when using Big Data systems and unstructured data sets. The criteria used in the case study to identify Big Data quality has been written in the third part of this article.

In the article from [Wright, Ogbuehi, 2014] the main hypothesis is that completeness of answers will be equivalent across face-to-face interviews, paper and pencil questionnaires and electronic form data collection methodologies. How-

ever this does not work with unstructured data sets. In such datasets we try to define questions and then answer them based on the data that is burden, dirty and ambiguous. We have to strictly define that such kind of data quality does not work with other type of data such as machine generated data, which is mostly based on sensors [Mou et al., 2015]. It also includes data analytic functions (e.g., data science, predictive analytics, and big data) in order to enhance supply chain processes [Hazen et al., 2014].

Usually data quality is defined by its attributes which are: relevance, accuracy, coherence, comparability, timeliness-punctuality and accessibility-clarity. Typically we analyse data quality based on quality processing, quality organisation and output quality [Wernicke, 2014]. In Big Data the central data quality attributes are completeness and coherence.

We cannot forget about legal issues of Big Data processing especially when processing elementary data from social media [Bottles et al., 2014]. During publishing data analysis we have to keep in mind that there is no possibility to disseminate data in aggregates with less than 5 persons [Daries et al., 2014]. On the other hand eliminating the individuals from the data set should not affect the data results [Heffetz, Ligett, 2014].

6.3. A case study

The aim of the case study presented in this chapter was to prepare results based on job offer websites analysis. Those websites publish job offers in unstructured format, usually in hypertext tagged tables. The goal was to provide information on the demand on labour market based on data analysis. This case study allowed preparing a general framework of conducting such analysis. This framework was presented below and includes data quality issues. The list of Big Data quality attributes that were identified during the case study is presented in the third part of this chapter.

A typical Big Data analysis framework consists of three steps presented in Figure 6.1 by UML Activity Diagram.

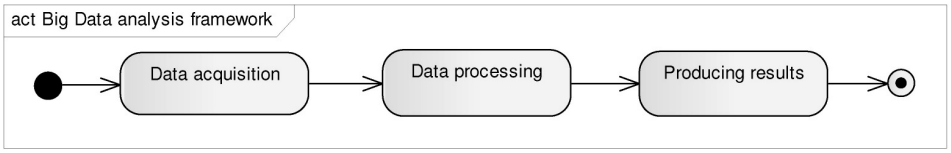

Figure 6.1. A typical Big Data analysis framework

Source: Own elaboration.

The first step is the data acquisition. Usually it is performed by downloading a set of documents from the Internet (such as websites) or prepare a dataset in the database. Based on the UNECE, the way the data is gathered depends on the types of data, which can be:

- human-sourced information (social networks and any other forms of human activity),
- process-mediated data (traditional business systems and enterprise websites),
- machine-generated data (automated systems such as sensors).

To process the data, several different platforms can be used. The most popular and easy to implement is Apache Hadoop ecosystem. To process the data, especially these unstructured, the MapReduce algorithms can be used. The last step of the typical Big Data analysis framework is producing results. It involves such activities as preparing result tables and charts. It also includes checking whether the results are reliable by giving cross referenced checks. It means that not only the individual value of the data must be checked but even its references to other data if possible. For instance if the data is referenced in several tables, each reference have to be checked individually.

To define the data quality in this chapter the Apache Hadoop ecosystem has been used. It especially includes MapReduce algorithms that were used to assess the data quality in two steps. Figure 6.2 shows UML Activity Diagram of suggested Big Data quality assessment process.

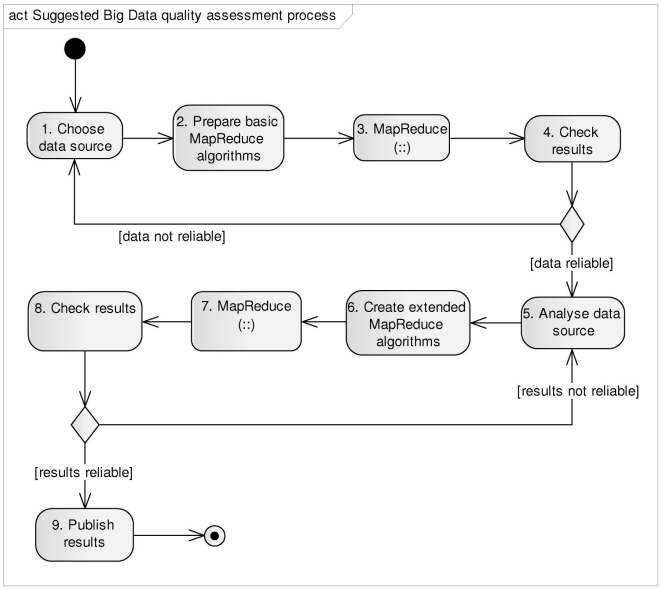

Figure 6.2. A suggested Big Data quality assessment process

Source: Own elaboration.

Although the case study is related to a platform that seek for job offers in numerous business portals, it can be used to process unstructured data from most of business related websites.

Firstly, after choosing the data source, it is necessary to check if the data is reliable. To check whether the data is reliable we suggest using the MapReduce algorithm prepared for WordCount and regular expressions.

Listing 1 shows the approach used in the case study to validate the data. In this listing the fundamental tools of Big Data were used, such as HDFS – Hadoop Distributed File System to store the data and MapReduce Java code used to identify the coherence of the data and its ambiguousness. The result of this analysis is the indicator that showed how the data set is distributed. Interpretation of this indicator leads to the decision whether the data source should be used for the next step of data quality assessment.

Listing 1. A suggested Big Data quality assessment process

```
# 1. prepare data source
hadoop fs -mkdir hdfs://localhost:8020/datasource
# 2. copy data
hadoop fs -copyFromLocal /home/dane/externaldata /datasource
# 3. checking quality of the data by word count: step 1
hadoop jar wc*.jar org.apache.hadoop.examples.WordCount /dataset
/preliminarydataresults
# 4. checking quality of the data by regular expressions: step 2
hadoop fs -cat /preliminarydataresults/part-r-00000 | findstr
"KEYWORD"
```

The listing above shows a typical way of identifying how many unambiguous keywords are presented in the dataset and how they are distributed. The more lines of the analysis result, the less reliable dataset is. This indicator is relative to the dataset, however an indicator with more than 50% of the size of its original file may be regarded as not reliable. Using such dataset leads to several problems with data quality. For instance the result is not accurate, may not be comparable and its coherence is very low. We can be sure that also the keywords are ambiguous.

The second phase of data quality assess is to provide first results. This can be done by searching for regular expressions on unstructured data source or by using specific MapReduce algorithms (point 4 in listing 1). Based on the observation, this process cannot be done automatically. This means that after preparing the first results of analysis, it has to be checked by data scientists to ensure that the data is comparable and reliable.

The case study leads to the conclusion that different aspects of Big Data quality are relevant, depending on the phase – acquisition, processing and producing outputs. In typical data quality framework, attributes are valid for the whole data

processing. We do not suggest such general approach. The next chapter shows the importance of each attribute in the Big Data analysis.

6.4. Suggested Big Data quality framework

The following quality attributes were identified during case study and previous surveys made by author of the chapter:

- relevance,
- comparability,
- timeliness,
- accuracy,
- coherence,
- accessibility,
- ambiguousness.

Although there are well known attributes of the data quality, we need to ensure to treat the quality attributes in different way. In this chapter below we can find how different attributes.

First attribute – relevance – refers to the first assessment of the data source. It may include such quality dimensions as completeness and representativeness. Comparability is mostly related to the issues of integrating different data sources. Usually in Big Data we need to integrate the data. If the particular data source is not comparable with others, the data source cannot be used in the integration process. Timeliness refers to the possibility of producing results in time series based on data sources used in the processing phase. Accuracy includes any aspects related to errors in the data or issues that make the data not being able to process correctly. Coherence means that the data must be consistent as well as valid for the purpose of the data processing. Accessibility is related to all the problems with privacy and security. In some cases data are not available to process them due to legal issues or in the future the data may not be accessible. These aspects should be checked and verified before the data source will be included in the data processing phase. Ambiguousness has been treated in this chapter as the separate attribute, however in numerous studies it was treated as part of completeness or accuracy. The goal of the ambiguousness phase is to find all the data that may be treated as ambiguous or not. The indicator used to identify this quality attribute is for example the number of records that is ambiguous.

It is suggested to assess each attribute individually. In the first phase the most important attributes are relevance, timeliness and accessibility. It includes both le-

gal issues of the data source as well as possibility to get it for further processing. It is presented in Figure 6.3.

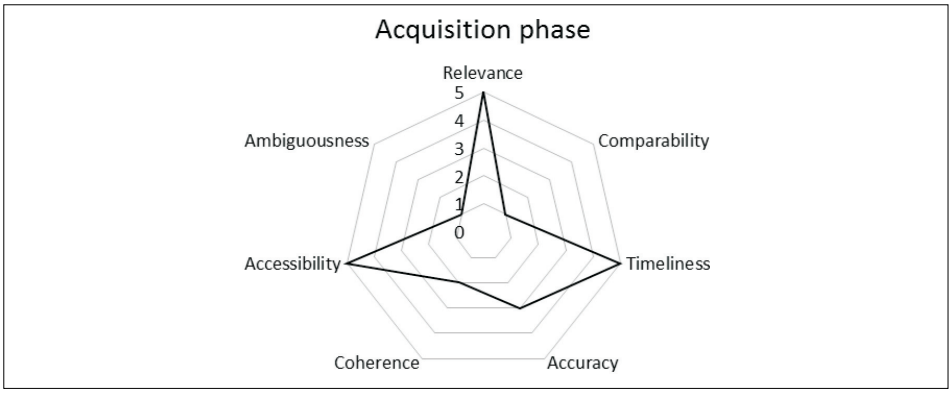

Figure 6.3. Weights of data quality attributes during acquisition phase
Source: Own elaboration.

As it was shown in the Figure 6.3, the most important issues that must be checked and verified during data gathering are accessibility, timeliness and relevance. If the metadata gathered on a specific data source is not enough to provide the data to indicate whether the data source is relevant, accessible now and in the future (if that issue is crucial to the analysis) and it is possible to make a timeline, we have to abandon this source and try to find substitute data source. However, practical examples of Big Data implementation show that if the data source has very low marks on the attributes written above, it can be forwarded to the next step of the data quality assessment.

In Figure 6.4 the following data attributes are the most crucial: ambiguousness and coherence. It especially involves problems in identifying the keywords in the context of the text written in the specific unstructured data source. Usually this problem cannot be omitted and we have to assess the data quality by eliminating the number of ambiguous sentences or keywords.

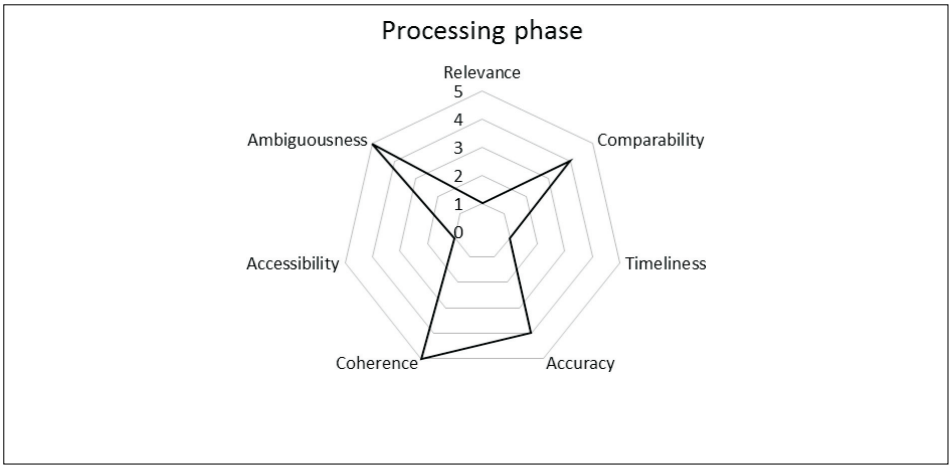

Figure 6.4. Weights of data quality attributes during processing phase
Source: Own elaboration.

In Figure 6.5 we can observe that comparability and accuracy are the most important data quality attributes during the output phase, which in fact is the most relevant phase.

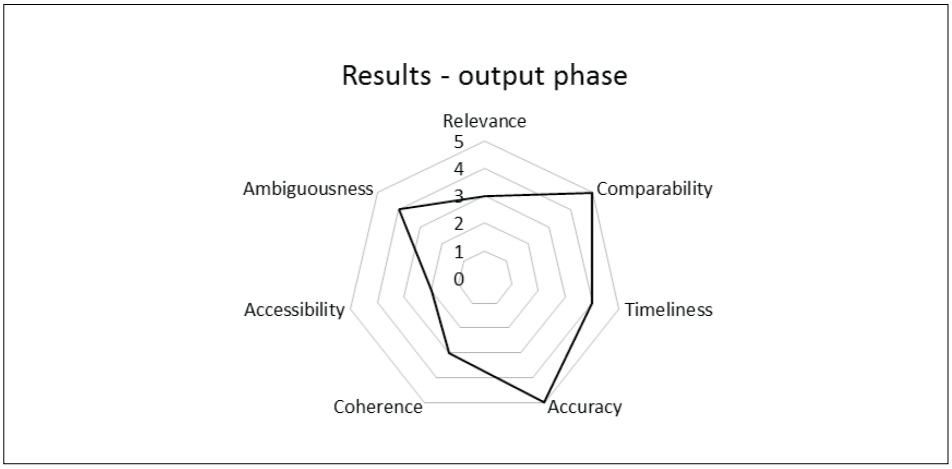

Figure 6.5. Weights of data quality attributes during output (producing results) phase
Source: Own elaboration.

The summary of the most crucial Big Data quality attributes is listed in Table 6.1.

Table 6.1. The relevance of Big Data quality attributes by different phase of analysis

Specification	Phase of Big Data analysis		
	acquisition	**processing**	**producing results**
Relevant attributes	relevance accessibility timeliness	ambiguousness coherence	comparability accuracy

Source: Own elaboration.

Based on the data in table 6.1 we can see that the attributes are distinct in every phase of Big Data processing. However this does not mean that attributes that are not highly relevant in each phase are not important. It means that those indicated with mark 5 in Figures 3–5 are the most relevant.

Conclusion

The case study presented in the chapter allowed identifying the potential problems with data quality regarding unstructured data analysis for Big Data. The hypothesis 'not all data quality attributes are relevant to unstructured data analysis using Big Data' has been confirmed. We can say that the most relevant issues of Big Data quality rely on seven attributes: relevance, comparability, timeliness, accuracy, coherence, accessibility and ambiguousness. Their roles are different depending on the phase. It means that we have to separate different phases of Big Data analysis which are data acquisition, data processing and producing results. It was shown in the third part of this chapter.

The chapter shows a suggested approach to Big Data quality issues by identifying business activities in business process. Moreover a crucial data quality in each phase of this process were chosen and provided as a part of its framework.

We can observe that the role of Big Data analysis for business purposes will increase and therefore the evaluation of Big Data quality issues is crucial. To extend the suggested model of Big Data analysis we decided to test this approach in different environments including integration of structured and unstructured data sources.

References

1. Bertot J., Gorham U., Jaeger P., Sarin L., Choi H. (2014), *Big data, open government and e-government: Issues, policies and recommendations*, "Information Polity: The International Journal Of Government & Democracy In The Information Age", Vol. 19, No. 1/2.
2. Bottles, K., Begoli, E., Worley, B. (2014), *Understanding the Pros and Cons of Big Data Analytics*, "Physician Executive", Vol. 40, No. 4.
3. Chen H., Chiang R., Storey V. (2012), *Business Intelligence and Analytics: From Big Data to Big* Impact, "MIS Quarterly", Vol. 36, Issue 4.
4. Chen M., Mao S., Liu Y. (2014), *Big Data: A Survey*, "Mobile Network Applications", No. 19.
5. Chen P., Zhang C.-Y. (2014), *Data-intensive applications, challenges, techniques and technologies: A survey on Big Data*, "Information Sciences", Vol. 275.
6. Cote M. (2014), *Data motility: The materiality of big social data*, "Cultural Studies Review", Vol. 20, No. 1.
7. Daries J., Reich J., Waldo J., Young E., Whittinghill J,. Dean Ho A., Seaton D., Chuang I. (2014), *Privacy, Anonymity, and Big Data in the Social Sciences*, "Communications Of The ACM", Vol. 57, No. 9.
8. Gang-Hoon K., Trimi S., Ji-Hyong C. (2014), *Big-Data Applications in the Government Sector*, "Communications Of The ACM" Vol. 57, No. 3.
9. Hazen B., Boone C., Ezell J., Jones-Farmer L.A. (2014), *Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and* applications, "International Journal of Production Economics", Vol. 154.
10. Heffetz O., Ligett K. (2014), *Privacy and Data-Based Research*, "Journal Of Economic Perspectives", Vol. 28, No. 2.
11. Hua-Ping Z., Rui-Qi Z., Yan-Ping Z., Bao-Jun M. (2014), *Big Data modeling and analysis of microblog* ecosystem, "International Journal of Automation and Computing" Vol.11, No. 2.
12. Kennedy R., Scriven J., Nenycz-Thiel M. (2014), *When 'significant' is not significant*, "International Journal Of Market Research", Vol. 56, No. 5.
13. Kim Y-I., Ji Y-K., Park S. (2014), *Big Text Data Clustering using Class Labels and Semantic Feature Based on Hadoop of Cloud Computing*, "International Journal of Software Engineering & Its Applications", Vol. 8, Issue 4.
14. Kwon O., Lee N., Shin B. (2014), *Data quality management, data usage experience and acquisition intention of big data analytics*, "International Journal of Information Management", Vol. 34, Issue 3.
15. Mantelero A. (2014), *Social Control, Transparency, and Participation in the Big Data World*, "Journal Of Internet Law", Vol. 17, No. 10.
16. Mou W., Liansheng T., Naixue X. (2015), *A Structure Fidelity Approach for Big Data Collection in Wireless Sensor Networks*, "Sensors", Vol. 15, No. 1.
17. Newman N. (2014), *How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and Other Vulnerable Sectors of the Population*, "Journal of Internet Law", Vol. 18, No. 6.

18. Santaferraro J. (2013), *Filling the Demand for Data Scientists: A Five-Point Plan*, "Business Intelligence Journal", Vol. 18, No. 4.
19. Wernicke I.H. (2014), *Quality of Official Statistics Data on the Economy*, "Journal Of Finance, Accounting & Management", Vol. 5, No. 2.
20. Wright B., Ogbuehi A. (2014), *Surveying Adolescents: The Impact of Data Collection Methodology on Response Quality*, "Electronic Journal Of Business Research Methods", Vol. 12, No. 1.

Chapter 7

Extending DEMO with an opportunity for simulation

Aleksey Shutov, Yulia Laryushina, Nikita Ponomarev, Oksana Radzinskaia

Introduction

DEMO [Dietz, 2006] that stands for Design and Engineering Methodology for Organizations was developed to represent and analyze business processes of organizations and to provide understanding of communication, information, action and organization. DEMO includes the PSI-theory, that explains functioning and structure of an organization, and organizational modelling methodology answering the question how to design the organizational model. The main contribution of this methodology is the provision of the essential model of an organization using several diagramming techniques that stem from the informational, documental and organizational realization [Dietz, 1999].

However, at present a sufficient software support for professional audience usage does not seem to exist. There are only the commercial editor Xemod [Xemod – product overview], a set of templates for Visio and browser editor prototype, but each of them has its disadvantages. Particularly, not only common interface issues but also common methodology-related drawbacks were found in these tools. Moreover, none of the above-mentioned editors takes into account the specific feature of the DEMO methodology: possibility of sharing several different types of diagrams with different sets of elements and different semantics. That is why it would be useful to develop a tool combining all common elements into one unit in order to avoid duplication and to simplify the process of modelling.

Moreover, the DEMO methodology is still not extended to support simulation experiments, and there are no known tools or executable environment where business-analysts could conduct model validation and investigate dynamic behavior of systems for better understanding of organization design.

Taking into account all of the above, obviously there is a necessity of development of an open platform for modelling, that will enable building and editing all types of DEMO-diagrams and will provide an opportunity for conducting simulation experiments on DEMO business processes models. The platform would expand the opportunities for modeling for independent developers and integrating modeling tools with other types of information systems. Thus, with the help of the

EMF (Eclipse Modeling Framework) technology [Budinsky et al., 2009], the authors developed the foundation for this open platform that supports not only the basic creation and editing of DEMO diagrams, but also provides ground for simulation in ProM 6 [ProM – Official tutorial] by converting PSD (Process Structure Diagramm) into a readable Petri Net in the pnml format. The method that enables building business process models based on formal semantics of Petri Net from DEMO business processes using DEMO Transaction Concept theory in order to verify whether business process models are syntactically correct and to provide a better understanding of possible future optimizations was earlier investigated by Joseph Barjis and Jan L.G. Dietz in their articles [Dietz, Barjis, 1999; Barjis, 2008; Barjis, 2007].

As a key concept at software development stage we will try to prove that the MDA (Model Driven Architecture) approach and a set of Eclipse EMF technologies will help to develop effective and efficient solutions by providing various levels of abstraction with each level emphasizing certain aspects or viewpoints of the system.

The present article includes four sections: review of the literature, describing of DEMOModeler and process of simulation, and conclusion.

7.1. Literature review

The following literature review will examine theoretical thinking and practical approach to the DEMO methodology – a methodology for the design, engineering, and implementation of organizations that focuses on explaining how and why people cooperate and communicate. Moreover, various approaches in simulation modelling using DEMO methodology will be described.

7.1.1. Model of an organization

Modern enterprises are mainly characterized by complexity and constantly changing nature which triggers the need for thorough understanding of organizational structure, interactions among divisions and transactional costs between them. That is why researchers try to develop a business domain model that satisfies the following requirement: it should make a clear distinction between the essential business actions and other processes, it should have the right (atomic, from the business point of view) level of detail and finally it should be complete (contain everything that is necessary and do not contain anything that is irrelevant) [Albani et al., 2006]. Such business domain mode is called enterprise ontology. Many various modelling languages for the business processes representations have been in-

troduced [Mendling et al., 2005; OMG, 2006; BPMN 2.0]. Nevertheless, despite being largely applied, the focus of all these languages is on graphical constructs rather than on formal semantics, so they are inappropriate for establishing common understanding and facilitating communication between people, because as people have different mental models they are likely to have different understanding. This problem can be solved by using the ontological approach. According to Gruber, an ontology is "a formal, explicit specification of a shared conceptualization" [Gruber, 1995], and all these languages cannot provide description of organization at the ontological level.

Some scholars tried to build an ontological business model framework that could provide users with easy to understand and analyse business model of their enterprises [De Nicola et al., 2007], [Osterwalder et al., 2004], or to develop an approach to modelling business processes at the semantic level, integrating knowledge about the organizational context, workflow activities and Semantic Web Services [Cabral et al., 2009]. Several authors also have used ontologies for describing and evaluating certain aspects of modelling languages [Wand, Weber, 2002; Weber, 1997].

7.1.2. Using the DEMO methodology for the enterprise engineering

Investigations provided by Jan Dietz have put forward new interpretations of enterprise modelling as a dynamic network of interrelated transaction processes [Dietz, 2006].

DEMO is constituted by the Enterprise Ontology that is defined by Dietz (the author of this methodology) [2006]. Enterprise Ontology is mostly based on the theory of Communicative Action and the Language-Action Perspective. According to Dietz [1999], DEMO is a methodology for designing, modelling and engineering organizations.

Numerous studies contain fundamentals of the methodology, including explanations of the basic definitions, and analysis of the four axioms of the Enterprise Ontology theory (distinction, production, transaction and composition), so these studies contain mostly theoretical research. The information about exact methods and diagrams of this theory was provided only by several researchers, such as Philip Huysmans, Kris Ven and Jan Verelst [Huysmans et al., 2010].

DEMO has a strong theoretical foundation; it provides clear and unambiguous definitions for the constructs used in various models and clear guidelines on how and why abstractions can be made. DEMO assists in understanding the process of modelling an organization by providing its high-level and abstract view. Moreover, DEMO studies the communication patterns between human actors, instead of the sequences in which activities are performed.

7.1.3. Simulation modeling

In order to conduct a better analysis and investigation of business processes, they have to be supported with simulation, because modelling itself cannot show all the details, bottlenecks and sufficient information needed for future optimizations [Barjis, 2011]. Analysis of business processes has attracted many researchers whose investigations in the sphere of business processes simulation gave us various perspectives of applications. They discuss all important aspects of a simulation study [Law, Kelton, 1991], evaluating alternative business process strategies using discrete-event simulation [Hlupic, Robinson, 1998], simulation modelling application in different spheres [Aguilar et al., 1999], efficacy of business process simulation (BPS) in designing and redesigning decisions [Giaglis et al., 1999], [Desel, Erwin, 2000] taking into account great amount of characteristics, such as resource availability, execution time [Van der Aalst et al., 2008].

Various approaches have been used to provide an efficient simulation technique for the DEMO models. Discrete-event simulation approach can be successfully based on the transaction-flow view [Hlupic, Robinson, 1998]. Using this method, we could present our system with discrete units dealing with limited resources, while moving through the system. Besides [Hlupic, Robinson, 1998], scholars point out that the key features of simulation is identifying bottlenecks and analyzing throughput times for better performance analysis. Thus, among available simulation modelling approaches using discrete-event dynamic systems, the Petri net is one of the most popular [Salimifard, Wright, 2001]. DEMO's Process Model Diagram contains all allowed steps in a business process as well as their relations, which is the most suitable diagram for creating a Petri Net simulation model [Vejrazkova, Meshkat, 2013].

In the article [Liu et al., 2003] the authors claim that "to study the organizational behaviour in the form of business processes, one needs an effective modelling method to capture dynamics of business processes". In particular, they stressed the importance of elaboration of the integrated modelling approach that involves organizational study from various angles. Despite the fact that the theoretical basis for the simulation of the dynamics has been carefully thought out (according to this article), when using existing modelling tools it would be virtually impossible to recreate the concept and add new dimensions to it.

7.1.4. Identifying drawbacks of DEMO

It is necessary to make an attempt to diminish the disadvantages of the practical application of the methodology lying in its unsuitability for visualization of bu-

siness processes dynamically using existing tools and adding new aspects to consider.

Despite the well-developed theoretical basis for the methodology and its active discussion, tools for its use (application in practice of all its features and ideas developed) are practically non existent, except for commercial products. This fact makes all the theoretical investigations meaningful only for researchers and useless for business modellers. The possible solution to this problem is proposed in the present study by integrating previously created graphical editor DEMOModeler with simulation modelling environment ProM 6 [ProM – Official tutorial].

7.2. Development of the object-oriented graphics editor for the DEMO methodology

7.2.1. Modelling framework

The DEMO methodology consists of four interrelated basic models that can be presented in the form of certain diagrams: Construction Model – Actor Transaction Diagram (ATD) and Organization Construction Diagram (OCD), Process Model – Process Structure Diagram (PSD), Action Model – Action Rule Specifications and State Model – Object Fact Diagram (OFD) [Dietz, 2006]. Construction Model determines the composition and structure of the organization (types of transactions, actors related to these transactions, information links, and links with data banks). Process Model describes a pattern (patterns of action) for each transaction, specified in the Construction Model. Action Model defines the rules of actions for each actor. State Model specifies entities and production facts of organization. All these models are closely related to each other, and each of them complements other.

When developing a DEMOModeler, the Model-Driven Architecture (MDA) approach for software design has been taken as the basis. Its main idea is to construct an abstract metamodel of management, exchange of metadata (models), and to establish the method for its transformation into a technology-supported programming (Java, XML, etc.). As a technology for applying the MDA approach, we have chosen the Eclipse EMF (Eclipse Modeling Framework) technology [Budinsky et al., 2009]. Eclipse EMF is a tool for creating models and generation code for building tools based on a structured data model from the model specification prescribed in XMI. Usage of the EMF technology facilitates expanding the project and integrating it with other projects.

7.2.2. Description of the implementation of the DEMOModeller

Following the logic mentioned before, we built the ecore-model for diagrams ATD, PSD, OFD and OCD. Example of the ATD ecore model is presented in Figure 7.1. This allows us to obtain an initial version of the editor, which contains all the basic elements of diagrams, gives an opportunity for building diagrams, but has a number of shortcomings: in particular, it displays a number of elements in each diagram incorrectly. That is why we had to correct the code for these items manually.

7.2.3. Automatically generated object in DEMOModeler

Through an analysis of the relationships between the models, three types of common elements were identified: common actors, common transactions and common boundaries. "Common" means that they can be generated automatically for all types of diagrams after being created in one. Under the current version, we have implemented an integration scheme ATD-> PSD, ATD-> OCD, ATD-> OFD. The full integration is further expected (related diagrams will be updated in base of any stored diagrams).

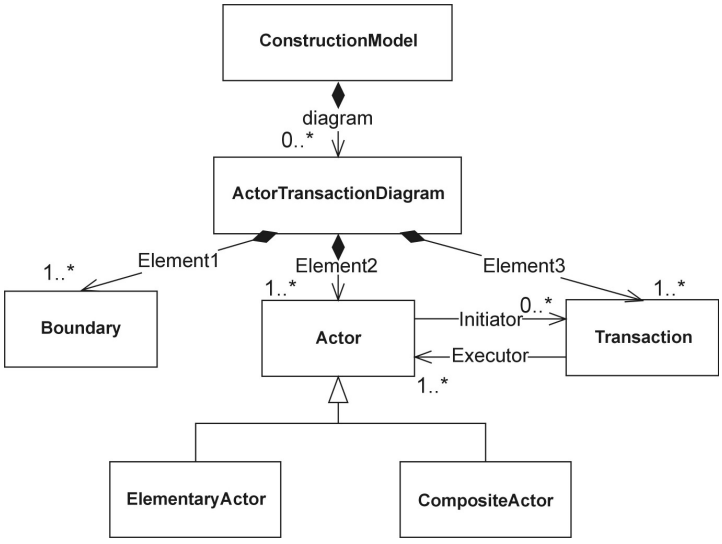

Figure 7.1. The model for ATD

Source: Own elaboration.

All the obtained schemes can be represented as follows (Figure 7.2). The scheme contains objects of the ATD diagram and "related" to them objects of other diagrams within the previously selected groups. In other words, this table

shows clearly, what objects in other diagrams are created after building a related ATD diagram. User interface of the DEMOModeler is presented in Figure 7.3.

As it has been mentioned earlier, the tools for the DEMO methodology have not been extended yet to support simulation experiments, and this was perceived as a weighty drawback. In terms of Enterprise Engineering, building a well-designed model is impossible without verification, validation and testing of the model [Vejrazkova, Meshkat, 2013]. But a high-level conceptual model like DEMO cannot be verified and validated automatically, and requires human factor.

In order to solve the problem without making any changes to the methodology itself, the method of translating models into Petri nets, proposed by Z.Vejrazkova and A.Meshkat [Vejrazkova, Meshkat, 2013], was applied. Particularly, the aim was to automate translation proposed and ensure the preservation of the obtained nets in a format that is received by one of the most widely supported tools – ProM 6 [ProM – Official tutorial].

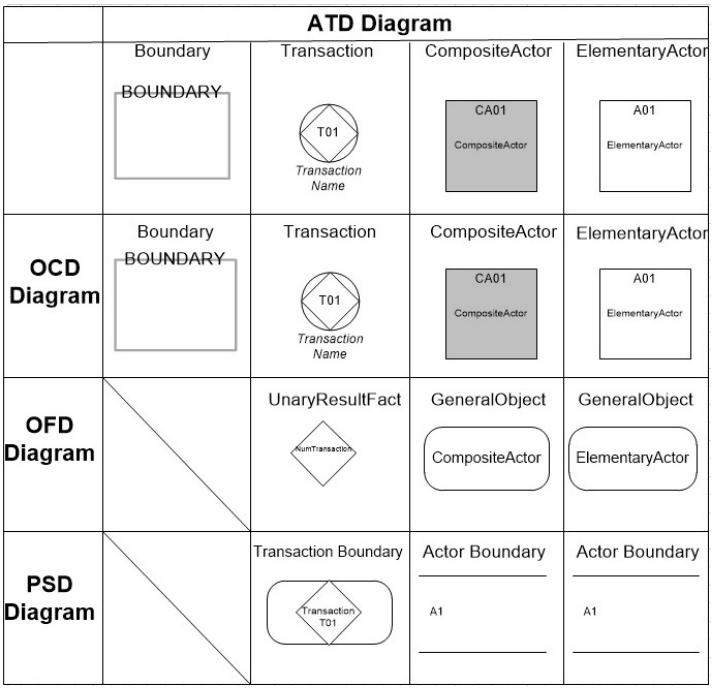

Figure 7.2. Integration scheme ATD-> PSD, OCD, OFD

Source: Own elaboration.

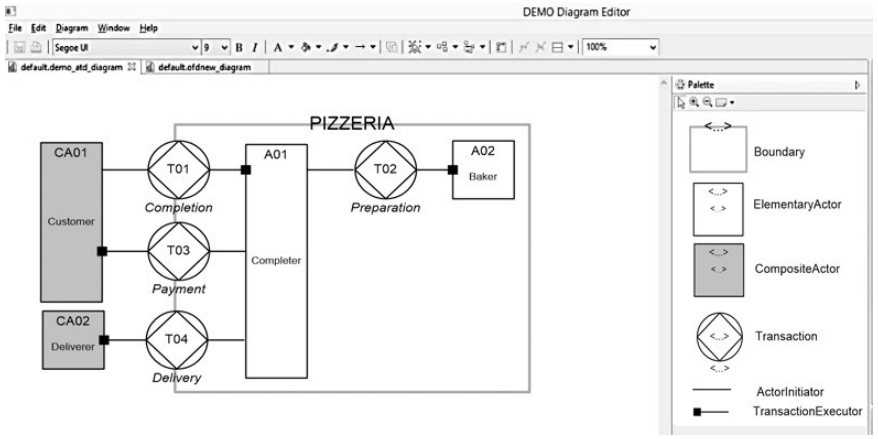

Figure 7.3. User Interface. ATD diagram

Source: Own elaboration.

The reasons why the authors have chosen the Petri Net based workflow system for modelling were formulated in [Van der Aalst at al., 2008], where the following three strengths are emphasized: 1) formal semantics despite the graphical nature; 2) state-based instead of event-based; and 3) abundance of analysis techniques. All these reasons can be reduced to one: proved suitability for simulation of Petri Nets, so the models can be easily used to examine the behavioural aspects of the modelled system during simulation [Desel, Erwin, 2000]. Moreover, Petri Nets also make it possible to represent organizations at the ontological level, which is a crucial factor when considering translation from conceptual DEMO models [Vidal et al., 2006].

Reasons for choosing ProM 6 lie in the broad functionality offered by this tool, which includes not only performing a simple simulation, but also more than two dozens of other different functions. It is also being actively developed, and independent developers constantly supplement it.

Therefore, in this section, the features of translation DEMO models, the necessary additions and the resulting format are described. The method of translating the DEMO model into a corresponding Petri Net proposed in [Vejrazkova, Meshkat, 2013] has been developed through examining different models in DEMO and creating the corresponding models in Petri Nets.

We will consider the following behaviour patterns proposed in the article [Vejrazkova, Meshkat, 2013]: basic transaction pattern, standard transaction pattern, self-initiated transaction, request & no wait, request & wait for completion of request, request & wait for promise.

The analysis of connection is based on the created XML files containing information about the objects drawn: IDs, labels references and so on. To use the trans-

lation, a non-mandatory for typical usage tag "parent" is required. The tag "parent" is intended to specify transaction of each Starting Process Step, Ending Process Step, and Execute Transaction object refers to and represents just an ID of "parent" transaction. Using this tag, the tool identifies exactly what pattern should be used analyzing relations between transactions. After performing the analysis, each transaction is translated into the corresponding Petri Net part according to the templates mentioned above. The resulting transaction represents a node of XML document, containing id, label, activity, appearance characteristics: color, position, dimensions. The construction of each pair of consecutive transitions is accompanied by the addition of a place that has the approximate form. Also, each link is also transformed into a node, and after generating all transitions, places and arcs, we add an end transition.

It is now possible to be loaded into ProM 6 or any other tool that supports the pnml format Petri Net, derived from Process Structure Diagram of DEMO. Actors are ignored in the current version, but their automatic conversion will be also developed.

Hereby, to support the claim of authors of the translating method, the DEMO model seems to contain all invariable information required to build a simulation model. Nevertheless, in fact, invariable information is not enough for proper simulation and experiments based only on such data are useless for model verification and testing (whereas validation does not require other types of data). For the last two indicated processes, performance measures are required [Van Der Aalst, 2010].

As simulation modelling is used to reduce the chances of failure to meet specifications, to eliminate unforeseen bottlenecks, to prevent under or over-utilization of resources, and to optimize system performance [Anu, 1997]; it is rational and not too different from performance measures obtained when conducting analysis.

In order to supplement DEMO with diverse performance measures attached to entities, optional attributes for each of them have been specified. In other words, within the created tool Process Structure Diagram of DEMO was supplemented by the rules from Action Rules Specifications. Each transaction can be supplemented by its execution and waiting time, wherein it is supposed that transaction is divided into phases (that way, the modeller specifies durations of separate phases within one transaction in order to handle nested transactions).

Moreover, decision rules for choice point were added as optional attributes for transactions. This was done in order to provide the proper simulation and model verification.

To sum up, the resulting Petri Net, either supplemented or not supplemented by the elements indicated above, can be used for ProM simulation experiments, for building other based on Petri Net models, provided by the tool, supplemented

by and analyzed with other proposed techniques (i.e. conformance checking or operational support), or may be translated into other notations (like BPMN) within ProM.

7.2.4. Example

In this section, we provide an example of automated translation of PSD diagram to a Petri Net of a well-known Pizzeria case in more detail. The PSD diagram for the case created in DEMOModeller (Figure 7.4) is presented below.

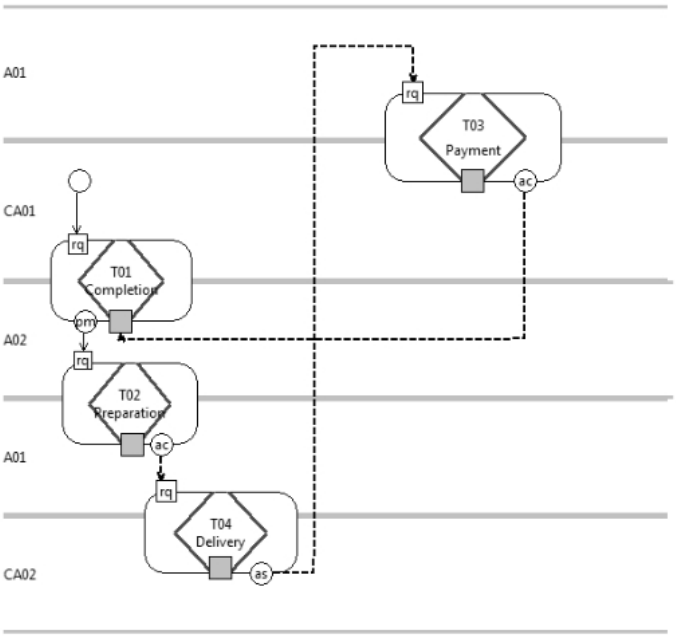

Figure 7.4. PSD Diagram created in DEMOModeller
Source: Own elaboration.

Firstly, we have to isolate transactions and analyze the connections between them. We also store their order for the subsequent generation of places and arcs. After analyzing connections and generating transaction nodes, we create a source place (it must be Initial Marking for ProM 6), and places for each two connected transactions. When places, transitions and stored connections have been obtained, we add an arc to each trinity "transition – place – transition". We also store actors and then use the information about them to transform our net and visualize it taking into account actor roles (currently represented only as determining the location of transitions factor). Below is the result of importing our pnml-net to ProM 6 (Figure 7.5).

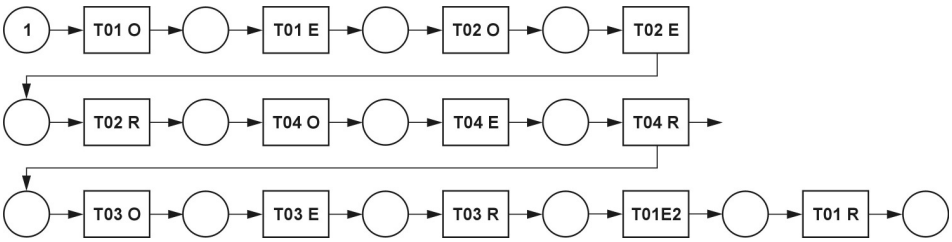

Figure 7.5. The result of importing pnml-net to ProM 6

Source: Own elaboration.

This is the view of our Petri net. Therefore, we can use all the features provided by this free tool. For instance, after uploading a pnml file generated by DEMOModeler to ProM the user can perform a simple simulation of a (stochastic) Petri Net. This kind of analysis provides user with an artificial log readable by the DEMOModeler. In other words, the DEMOModeler is able to parse, open and store compressed mxml files generated by ProM and other tools, and present them in the form of a regular table.

The automated translation of PSD diagram into the specific Petri Net Markup Language, supported by ProM (pnml), provides the user with a wide range of other opportunities. First of all, ProM 6 supports conversion of Petri Net to other notations (i.e. to a BPMN model) without supplementary time spent on (re)modelling. Secondly, many other tools, such as ePNK, PNML Framework, Petri Web, Coloane, etc., claim support of the PNML Standard, thus automatically generated in the DEMOModeler files can be easily used in an escalating number of tools.

Conclusion

There are several points that should be mentioned as advantages of the developed software: opportunity to create and edit all DEMO diagrams, integration of models in order to avoid duplication of entities, support for the latest version of the methodology (3.6c), user friendly interface, opportunity of simulation (availability of automated translation PSD diagrams in Petri Net for further analysis using tools supporting pnml format).

Among other advantages: extensibility of the editor due to the use of the EMF technology, opportunity of integration with other Eclipse projects, unified information model for all components.

In addition to the work done, we see the need to add a number of functionalities to the DEMOModeler and make some changes in order to provide a user-friendly interface, broaden opportunities for analysis with ProM and improve visualization. Thus, we are going to provide an opportunity to create optional transactions or to designate their multiple execution and, as a consequence, develop other automatic translation schemes proposed in [Vejrazkova, Meshkat, 2013] for further ProM analysis of generated Petri Nets. Moreover, adding probabilities and timestamps to transactions will also help to provide the basis for building a complete simulation model and it is also intended to develop a mechanism of transformation of PSD diagrams to executable simulation model of one of the most widely known languages – NetLogo or SIMAN using supplemented methods proposed in the article.

Thus, we have designed and developed a graphics editor for creating and editing DEMO models that is appropriate for professional usage and contains essential characteristics currently ignored by other tools supporting the DEMO methodology. Moreover, one of the disadvantages of discussed methodology was diminished by providing an opportunity for conducting simulation experiments.

References

1. Aguilar M., Rautert T., Pater A. (1999), *Business process simulation: a fundamental step supporting process centered management*, Winter Simulation Conference Proceedings. 2, ACM, New York.
2. Albani A., Dietz J., Zaha J. (2006), *Identifying business components on the basis of an enterprise ontology*, in: Konstantas D., Bourrières J., Léonard M., Boudjlida N. (eds.), *Interoperability of Enterprise Software and Applications*, Springer, London.
3. Anu M. (1997), *Introduction to modeling and simulation*, Proceeding WSC '97 Proceedings of the 29th conference on Winter simulation, IEEE Computer Society, Washington DC.
4. Barjis J. (2007), *Developing executable models of business systems*, ICEIS 2007 – 9th International Conference on Enterprise Information Systems, Proceedings ISAS, 3.
5. Barjis J. (2008), *A Business Process Modeling and Simulation Method Using DEMO*, in: Filipe J., Cordeiro J., Cardoso J. (eds.), Enterprise Information Systems, Lecture Notes in Business Information Processing, 9th International Conference, ICEIS 2007, Funchal, Madeira, June 12–16, 2007, Vol. 12, Springer, Berlin Heidelberg.
6. Barjis J. (2011), *Enterprise Modeling and Simulation Within Enterprise Engineering*, "Journal of Enterprise Transformation", Vol. 1.
7. BPMN *2.0 – Documents Associated With Business Process Model And Notation*, Object Management Group: http://www.omg.org/spec/BPMN/2.0, accessed 05.12.2014.
8. Budinsky F., Steinberg D., Paternostro M., Merks E. (2009), *EMF: Eclipse Modeling Framework*, Addison-Wesley Professional.

9. Cabral L., Norton B., Domingue J. (2009), *The Business Process Modelling Ontology*, SBPM '09 Proceedings of the 4th International Workshop on Semantic Business Process Management, Crete, Greece: ACM New York, NY, USA.
10. De Nicola A., Lezoche M., Missikoff M. (2007), *An ontological approach to business process modeling*, Proceedings of the 3rd Indian International Conference on Artificial Intelligence (IICAI).
11. Desel J., Erwin T. (2000), *Modeling, Simulation and Analysis of Business Processes*, in: Aalst W., Desel J., Oberweis A. (eds.), *Business Process Management*, LNCS, 1806, Springer, Heidelberg.
12. Dietz J. (1999), *Understanding and Modelling Business Processes with DEMO*, Proc. 18th International Conference on Conceptual Modeling (ER'99), Paris.
13. Dietz J. (2006), *Enterprise Ontology: Theory and Methodology*, SpringerVerlag, Berlin Heidelberg.
14. Dietz J., Barjis J. (1999), *Supporting the DEMO Methodology with a Business Oriented Petri Net*, Proceedings of the Fourth CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design (EMMSAD'99), Heidelberg, Germany.
15. Giaglis G., Paul R., Hlupic V. (1999), *Integrating simulation in organizational design studies*, "International Journal of Information Management", Vol. 19, No. 3.
16. Gruber T. (1995), *Toward principles for the design of ontologies used for knowledge sharing?*, "International Journal of Human-Computer Studies", Vol. 5–6.
17. Hlupic V., Robinson S. (1998), *Business process modelling and analysis using discrete-event simulation*, Proceedings of the 30th Conference on Winter Simulation, IEEE Computer Society Press, Washington, D.C.
18. Huysmans P., Ven K., Verelst J. (2010), *Using the DEMO methodology for modeling open source software development processes*, "Information and Software Technology", Vol. 56, No. 6.
19. Law A., Kelton D. (1991), *Simulation modeling & analysis*, McGraw-Hill, United States of America.
20. Liu K., Sun L., Barjis J., Dietz J. (2003), *Modelling dynamic behaviour of business organisations-extension of DEMO from a semiotic perspective*, "Knowledge-Based Systems", Vol. 16, No. 2.
21. Mendling J., Neumann G., Nüttgens M. (2005), *Yet another event-driven process chain*, in: Van der Aalst W., Benatallah B., Casati F., Curbera F. (eds.), 3rd International Conference, BPM 2005, Lecture Notes in Computer Science, 3649, Nancy, France, Springer Berlin Heidelberg.
22. OMG (2006), *Unified Modeling Language: Superstructure, Version 2.0, formal/05-07-04*, Object Management Group, Needham.
23. Osterwalder A., Parent C., Pigneur Y. (2004), *Setting up an ontology of business model*, EMOI workshop at Conference on Advanced Information Systems Engineering (CAISE), Riga.
24. *ProM Official tutorial*, http://www.promtools.org/prom6/downloads/prom-6.0-tutorial.pdf, accessed 15.01.2015.
25. Salimifard K., Wright M. (2001), *Petri net-based modelling of workflow systems: An overview*, "European Journal of Operational Research", Vol. 134, No. 3.

26. Van Der Aalst W. (2010), *Business Process Simulation Revisited*, in: Barjis J. (eds.), Enterprise and Organizational Modeling and Simulation. 6th International Workshop, EOMAS 2010, held at CAiSE 2010, LNBIP. 63, Hammamet, Tunisia: Springer, Berlin Heidelberg.
27. Van der Aalst W., Nakatumba J., Rozinat A., Russell N. (2008), *Business process simulation: How to get it right?*, BPM Center Report BPM-08-07.
28. Vejrazkova Z., Meshkat A. (2013), *Translating DEMO Models into Petri Net*, in: Barjis J., Gupta A., Meshkat A. (eds.), Enterprise and Organizational Modeling and Simulation.9th International Workshop, EOMAS 2013, Held at CAiSE 2013, Valencia, Spain, June 17, 2013, Selected Papers, LNBIP, 153, Hammamet, Tunisia, Springer, Berlin Heidelberg.
29. Vidal J., Lama M., Bugarín A. (2006), *A high-level Petri net ontology compatible with PNML*, "Petri Net Newsletter", Vol. 71.
30. Wand Y., Weber R. (2002), *Research Commentary: Information Systems and Conceptual Modeling – A Research Agenda*, "Information Systems Research", Vol. 13, No. 4.
31. Weber R. (1997), *Ontological foundations of information systems.* Coopers & Lybrand and the Accounting Association of Australia and New Zealand.
32. *Xemod – product overview*, Xemod, http://www.mprise.eu/xemod-product-overview.aspx, accessed 15.01.2015.

Summary

As it was shown in the monograph, there are still several different challenges for researchers that are strictly related to information systems development in the area of social media, system designing, systems security or Big Data.

First of all, increasing popularity of social media (SM), including social networks sites - SNS, microblogs and blogs, allows researchers to conduct surveys and make deep analysis of this phenomenon in several aspects of their appliances. Chapter one shows that social media is very efficient as a tool to support recruitment processes.

The second very important conclusion from the research described in this monograph is that creating efficient software for monitoring the productivity can be only performed by designing the software that will cover all the aspects of organization in one model. In current days, creating the software by designing separated modules without any integration is not a way to provide high quality software that will increase the productivity of the company.

Concerning security issues of software, there is still a lack in knowledge of people using new technologies. As we know, the importance of security issues is increasing, especially in the world in which most of us produce lots of information every day. It includes e-mails, weblogs or comments on websites. The conclusion of lack of security in todays enterprises is very important for numerous researchers that are making surveys in that matter.

In the world of information, modeling emotions is very relevant issue especially for several analytics that are investigating sentiments in unstructured datasets. Differences between emotional states may be a key to understand the information and provide a conclusion.

The way we are using electronic devices today created a new term in IT which is BYOD that stands for Bring Your Own Device. This monograph shows the challenges faced by companies wishing to enforce rules for introducing BYOD. The main conclusion from this research is that a high percentage of knowledge about BYOD and visibility implementation of BYOD in companies does not move directly into new management BYOD procedures.

Big Data term that has been introduced few years ago must be enhanced by the knowledge of its quality. Several possible indicators has been shown in this monograph. We can see that the role of Big Data analysis for business purposes will increase and therefore the evaluation of Big Data quality issues is crucial. The conclusion is that quality issues must be treated separately for each phase of Big Data lifecycle.

In the last chapter we can find the possibilities and advantages of the developed software based on DEMO Diagrams. This leads to the conclusion that such solution can be expanded and extended for better development of the software and providing a high quality tools. It includes also simulation of a stochastic Petri Net. As Authors mentioned in the chapter: the automated translation of PSD diagram into the specific Petri Net Markup Language, supported by ProM (pnml), provides the user with a wide range of other opportunities.

I hope that this monograph will provide valuable information for numerous researchers. As we can read in chapters, lots of work are in progress, which means that in the future we can expect a new high quality scientific papers showing further researches and deep analysis of the topics discovered in this monograph.

About Authors

Dorota Buchnowska, Ph.D., Department of Business Informatics, Faculty of Management, University of Gdańsk, e-mail: dbuchnowska@gmail.com

Michał Firek, Ph.D. student, Department of Business Informatics, Faculty of Management, University of Gdańsk, e-mail: michal.firek@gmail.com

Przemysław Jatkiewicz, Ph.D. graduate, Department of Business Informatics, Faculty of Management, University of Gdańsk, e-mail: jatkiewicz@ip-info.pl

Agata Kołakowska, Ph.D., Department of Intelligent Interactive Systems, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, e-mail: agatakol@eti.pg.gda.pl

Dariusz Kralewski, Ph.D., Department of Business Informatics, Faculty of Management, University of Gdańsk, e-mail: darek@univ.gda.pl

Agnieszka Landowska, Ph.D., Department of Software Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, e-mail: nailie@eti.pg.gda.pl

Yulia Laryushina, Ph.D., National Research University Higher School of Economics, Nizhny Novgorod, e-mail: yudlaryushina@edu.hse.ru

Jacek Maślankowski, Ph.D., Department of Business Informatics, Faculty of Management, University of Gdańsk, e-mail: jacek.maslankowski@ug.edu.pl

Nikita Ponomarev, Ph.D., National Research University Higher School of Economics, Nizhny Novgorod, e-mail: noponomarev@edu.hse.ru

Oksana Radzinskaia, Ph.D., National Research University Higher School of Economics, Nizhny Novgorod, e-mail: ovradzinskaya@edu.hse.ru

Aleksey Shutov, Ph.D., National Research University Higher School of Economics, Nizhny Novgorod, e-mail: ashutov@hse.ru

Mariusz Szwoch, Ph.D., Department of Intelligent Interactive Systems, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, e-mail: szwoch@eti.pg.gda.pl

Wioleta Szwoch, Ph.D., Department of Intelligent Interactive Systems, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, e-mail: awszwoch@eti.pg.gda.pl

Michał R. Wróbel, Ph.D., Department of Software Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, e-mail: wrobel@eti.pg.gda.pl

Stanisław Wrycza, Prof., Ph.D. habil., Department of Business Informatics, Faculty of Management, University of Gdańsk, e-mail: swrycza@ug.edu.pl