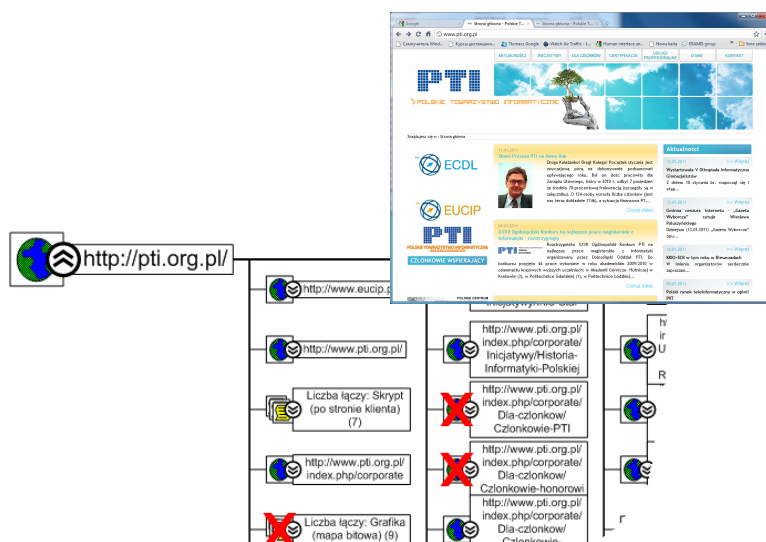
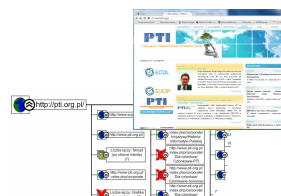


Edited by Marek Miłosz

ICT

- from Theory to Practice



ICT
- from Theory to
Practice

ICT - from Theory to Practice

Edited by MAREK MIŁOSZ

Polish Information Processing Society

Lublin 2010

Edited by
Marek Miłosz, Ph.D.

Reviewed by
Prof. Zbigniew Pastuszak, D.Sc., Ph.D.

Copyright © by Polish Information Processing Society
Lublin 2010

ISBN 978-83-931710-1-9

Table of Contents

Preface	7
<i>(Marek Miłosz)</i>	
1. Management of IT Projects in Public Administration Sector	9
<i>(Anna Kaczorowska, Anna Pamuła)</i>	
2. ICT e-Infrastructure for IT Platform of Managing Research Processes at Lublin University of Technology.....	33
<i>(Bogdan Wit)</i>	
3. Universal Structural Model of a Standard Economic Cluster	51
<i>(Rifkat Sirazetdinov, Ariana Brazhkina)</i>	
4. Using Google Search Engine to Get Unauthorized Access to Private Data	63
<i>(Maciej Laskowski)</i>	
5. Speech Technologies on the Internet	77
<i>(Mariusz Dzieńkowski)</i>	
6. ICT Applied in Management of Stables - Common Practice and the Future.....	99
<i>(Radosław Marciniak)</i>	
7. Electronic Contacts between the Applicant and Office Amendment of the Act on Informatization of Activities Carried out by the Public Task Performing Entities	129
<i>(Anna Kaczorowska)</i>	
8. Features of Software Development for Natural Language Processing	149
<i>(Ualsher Tukeyev, Zhandos Zhumanov, Diana Rakhimova)</i>	
9. Application of MALDI-TOF Mass Spectra Processing and Classification	175
<i>(Małgorzata Plechawska-Wójcik)</i>	

10. Application of Principal Component Analysis and Projection to Latent Structures for Modeling Processes of Aging and Destruction of Polymeric Materials	193
<i>(Alfred M. Suleymanov)</i>	
11. Creation of the User-Oriented Requirements Specification of the Data Intensive Web Application Basing on WebML	203
<i>(Kamil Żyła, Jacek Kęsik)</i>	
12. Experience in Building the Information Society in Poland	219
<i>(Marta Juszczuk, Elżbieta Miłosz)</i>	
Author Index	233

Preface

This book is the result of international cooperation between Lublin University of Technology (Poland) and universities from Lublin and Łódź (Poland), Kazan (Russian Federation) and Almaty (Kazakhstan). The main area of this cooperation is the Information and Communication Technology (ICT). The international cooperation has been developed since 2006 within the framework of the TEMPUS project.

Some topics presented in this book were reported on the conferences:

- Intelligent Information and Electronic Technology (IIET 2009, Lublin, Poland)
- Information Technology at Enterprises (2010, Kazimierz Dolny, Poland)
- XVII Microcomputer Workshop of the Lublin University of Technology (2010, Zwierzyniec, Poland).

Due to the relatively large variety of themes, the book is addressed to various groups of readers. A student, a researcher and a computer specialist from industry can find here something interesting for him/herself. A special recipient of the book is a Computer Science student, who wants to select the topic of the master thesis. The

book will allow him/her to find potential issues of the thesis and to recognize the degree of their complexity.

The book was published due to high commitment of the chapters' authors and the reviewer. I want to express my thanks to everyone who has contributed to the creation of this book.

Marek Miłosz

The Editor

Management of IT Projects in Public Administration Sector

In 2005 the Polish public administration launched the informatization process with the aim to provide services electronically. At that time a new form of management, i.e. through establishing projects, appeared in this sector. Therefore, since that time our administration's activities towards the State informatization should not be taken ad hoc, to achieve short-term purposes as fast as possible, but in compliance with the project management art they first should be planned and for each launched project some assumptions should be prepared and Terms of Reference should be determined. However, information projects undertaken by public administration contain many barriers which the authors pointed out in the conclusion. They list in it recommendations to improve implementation of such undertakings in the subsequent financial perspective of the State Informatization Plan for 2010 – 2013.

1.1. MILESTONES ON THE WAY TO MANAGEMENT THROUGH PUBLIC ADMINISTRATION PROJECTS

1.1.1. THE ACT ON INFORMATIZATION OF ACTIVITIES CARRIED OUT BY THE PUBLIC TASKS PERFORMING ENTITIES

E-activity of Polish public administration is possible owing to the Act on informatization of certain entities performing public tasks, of 17 February 2005 (Journal of Laws No 64, sec. 565 with amendments; hereinafter referred to as the UINF).

The UINF defined precisely the principles of establishing the State Informatization Plan and information projects of public use and their inspection. Since 2005 art. 3 § 6 has contained the following definition of information project of public use: „a group of organizational and technical activities aimed at the construction, extension or refurbishing of the teleinformation system used for accomplishment of public tasks, providing services relating to maintenance of this system or development of the procedures of electronic accomplishment of public tasks.”. Paragraphs 7 and 8 of the same article point out, among information projects, the supra-sector and sector projects respectively. The supra-sector project is such an information project whose subject range refers to issues belonging to the properties of more than one division of governmental administration.

In UINF defined certain terms, thereby enabling their use in offices.

The act contains entries which were used to work out in June 2006 the „State informatization Plan for 2006” (hereinafter referred to as – PIP2006).

1.1.2. THE STATE INFORMATIZATION PLAN FOR 2006

The 1st of August 2006, the Council of Ministers adopted Regulation ((Journal of Laws of 18 August 2006) where § 1 states that „Be established the State informatization Plan for 2006”. PIP2006 was the annex to that regulation. In the introduction was given that the main objective of the plan is to create conditions for the existence,

consistent not only nationally but also in Europe, of system of on-line services based on the collaboration of the teleinformation system to meet the needs of residents and businesses. It was also noted that for effective implementation of public services should be established center, with a strong position to coordinate the process of the State informatization.

PIP2006 was a tool for the coordination of actions to prevent irregularities that occurred in the documents preceding the UINF, such as "Strategy for the informatization of the Republic of Polish - e-Poland 2004-2006". This strategy is "limited to the presentation of the activities planned by individual ministries and the only way to monitor its realization was to give the Council of Ministers by the Minister responsible for the informatization of the periodic reports on its realization " [RMR06].

In the PIP2006 specifies two priorities for the development of ICT systems (the first - the rationalization of public expenditure in the field of informatization and the second - the creation of the modern state friendly to the citizens) and the establishment of the program of actions for the development of information society. Due to the fact that the range of electronically supplied public services has been assessed as limited and below the European average level, the first objective, under the second priority, was defined as: The introduction and dissemination of electronically supplied government's services in order to help residents and businesses conducting the official affairs without having to appear in person at the office."

The fourth section of PIP2006 contains a summary of the sector and supra-sector IT projects (tab. 1.1).

PIP2006 presented the state of works aimed at Poland's informatization, three the supra-sector and three sector IT projects were singled out, due to new form of managing was implemented –t by setting up projects to the public administration sector. Thus, only since the 2006 management by projects, implemented here from the sphere of business, is considered as the most efficient in the process of the state informatization.

Tab. 1.1. The summary of the supra-sector and sector IT projects.

No.	Project name	Project category	Time of realization	Estimated cost or the source of funding
1.	E-PUAP – Electronic Platform of Public Administration Services	supra-sector	2006-2008	PLN 35 mln
2.	STAP – Teleinformation Network of Public Sector	supra-sector	2006-2007	Budgetary reserve
3.	E-STATEMENTS – tax returns for businesses	sector	IV quarter of 2005 – I quarter 2008	PLN 152 mln
4.	PESEL2 – rebuilding of the state register	sector	Was given that that it will be determined after the elaboration of the assumptions for the project	PLN 200 mln
5.	CEPIK – Central Register of Vehicles and Drivers	supra-sector	2006-2007	PLN 52 mln
6.	E-PORTAL – access to the national court register	sector	2006-2007	PLN 6,2 mln
7.	TOTAL COSTS			> PLN 445,2 mln

Source: Own work based on [RMR06]

That first plan of Poland's informatization was the foundation (a „leaven”) basing on whose a subsequent document was preparing, i.e. the “State Informatization Plan for 2007-2010” (on 21 April 2007 a decree on that plan, issued by the Ministry Cabinet, became effective (hereinafter referred to as PIP2007-2010).

1.1.3. THE STATE INFORMATIZATION PLAN FOR 2007-2010

On the basis of the regulation of the Council of Ministers of 28 March 2007 was established the second plan in the history of public administration informatization through establishing projects The State Informatization Plan for 2007-2010 (being the annex to that regulation; Journal of Laws No 61, sec. 415).

PIP2007-2010 maintains the priorities and objectives of the state informatization set out in the PIP2006 and will continue its tasks.

In the second part of PIP2007-2010 are listed the 23 sector and 5 the supra-sector IT projects set for implementation of identified priorities and services together with the description of projects which includes information on estimated cost, possible sources of founding and entities responsible for implementing them.

The summary list of sector and supra-sector ICT projects includes the following undertakings:

1. E-PUAP – supra-sector project, time of realization - 2006-2008, estimated costs – 35 mln of zloty,
2. E-PUAP2 – supra-sector project, time of realization - 2008-2013, estimated costs – 70 mln of zloty in years 2008-2010 and 184 in years 2008-2013,
3. STAP - supra-sector project, time of realization - 2006-2008, costs resulting from attaching to the STAP infrastructure and the use of it, in adequate proportion, are charged to the budgets of the particular users (budgetary reserve),
4. PI.ID – Polish ID card - supra-sector project, time of realization - 2008-2013, estimated costs – 70 mln of zloty in years 2008-2010 and 400 mln in years 2008-2013,
5. The realization of the central node of Polish component of SIS II and VIS - supra-sector project, time of realization - 2007-2008, estimated costs –54 mln of zloty,
6. CEPIK – Central Register of Vehicles and Drivers – sector project, time of realization - 2006-2007, estimated costs – 52 mln of zloty,
7. E-Statements I – support for electronic tax declarations of entrepreneurs - sector, time of realization – the 1st quarter of 2006 – the 11nd quarter of 2008, estimated costs – 75 mln of zloty,
8. E- Statements II – support for electronic tax declarations of entrepreneurs - sector, time of realization - 2007-2009, estimated costs – 100 mln of zloty,
9. PESEL2 – reconstruction of the state registers - sector, time of realization - 2006-2008, estimated costs – 166 mln of zloty,
10. The informatization of land registers - sector, time of realization - 2007-2010, estimated costs – 116 mln of zloty,

11. Building of domain electronic platform of the Ministry of Justice, and within it a run of services for entrepreneurs on access-information portal, providing electronic access into economic departments, registration and others - sector, time of realization - 2009-2010, estimated costs – 24 mln of zloty,
12. The information portal for notaries, lawyers, bailiffs, lawyers building - sector, time of realization - 2009-2011, estimated costs – 26 mln of zloty in years 2009-2010; 39 mln of zloty in years 2009-2011,
13. Integrated Management System of the State budget task force - sector, time of realization - 2007-2010, estimated costs – 44 mln of zloty,
14. Dostosowanie systemów informatycznych Resortu Finansów do ustawy o swobodzie działalności gospodarczej oraz zmianie niektórych innych ustaw - sector, time of realization - 2008-2009, estimated costs – 53 mln of zloty,
15. Sharing platform of on-line services for entrepreneurs and digital resources of medical registers – sector, time of realization – 2007-2013, estimated costs – 31 mln of zloty in years 2007-2010, 66 mln of zloty in years 2007-2013,
16. Electronic platform for collecting, analyzing and sharing digital resources for medical events - sector, time of realization – 2007-2013, estimated costs – 658 mln of zloty in years 2007-2010, 877 mln of zloty in years 2007-2013,
17. Building a communication platform for monitoring of disability and people from the social support (disabled persons benefiting from social assistance and family benefits system and maintenance support) - sector, time of realization – 2007-2013, estimated costs – 34 mln of zloty in years 2007-2010, 94 mln of zloty in years 2007-2013,
18. SI SYRIUSZ integrated service system of labour market and social security - sector, time of realization - 2007-2009, estimated cost – 110 mln of zloty,
19. SPPP supply and demand forecasting system - sector, time of realization - 2007-2010, estimated cost – 44 mln of zloty,
20. E-Taxes - sector, time of realization – 2008-2012, estimated costs – 88 mln of zloty in years 2008-2010, 197 mln of zloty in years 2008-2012,

21. IKONKA (consists in starting points of public access, so. Internet reading rooms, to Internet in public libraries of all communities of the province - sector, time of realization - 2007, estimated cost – 2 mln of zloty,
22. Construction of the cadastral system - sector, time of realization – 2007-2010, estimated cost – 190 mln of zloty,
23. TERYT2 national official registries of territorial division - sector, time of realization – 2007-2010, estimated cost – 45 mln of zloty,
24. Construction of a geo-referenced database of topographic objects - sector, time of realization – 2007-2011, estimated costs – 160 mln of zloty in years 2007-2010, 190 mln of zloty in years 2007-2011,
25. Official statistics information system - sector, time of realization – 2007-2013, estimated costs – 65 mln of zloty in years 2007-2010, 160 mln of zloty in years 2007-2013,
26. Consolidation and centralization of the customs and tax systems - sector, time of realization, estimated costs – 150 mln of zloty in years 2007-2010, 165 mln of zloty in years 2007-2013,
27. Central information on business activity - sector, time of realization - 2007-2008, estimated cost – 32 mln of zloty,
28. Presentation and sharing of audio and video archive - sector, time of realization - 2007-2010, estimated cost – 97 mln of zloty.

The element of that plan of informatization is also "Action Programme for the Information Society" (the 3rd part of PIP2007-2010), consistent with the "i2010 European Information Society for growth and employment" (adopted by the European Commission on the 1st of June 2006). The program is intended to the ranges of realization of priorities for the development of ICT systems.

Part four of PIP2007-2010 defines public tasks that should be provided electronically. These are priority services for individual citizens and businesses, which will be provided by electronic means, or as an e-services.

In PIP2007-2010, it is assumed that another State Informatization Plan would cover the period 2010-2013. This is to ensure the correlation between the State Informatization Plans and EU operational programs of structural funds for 2007-2013

and interact with the "Directional strategy of development of Polish informatization" (adopted by the Council of Ministers on 29th of June 2005).

1.1.4. DEVELOPMENTAL STRATEGY OF THE INFORMATION SOCIETY IN POLAND UP TO 2013

In October 2008 the Polish government prepared another document – the „Developmental strategy of information society in Poland up to 2013” (hereinafter referred to as the Strategy). It summed up the present condition of the development of information society in Poland and said: „The level of 20 basic services of public administration, recommended by the European Commission, which should be fully accessible on-line, is in Poland twice lower than the Union’s average value (in 2007 it reached 25%, with the Union’s average value 59%). Only four services reached the transaction level, enabling a complete interaction with offices through the use of electronic channels of access”[MSW10]. It was noted that 56% entrepreneurs send the completed forms electronically. Such result is surely affected by obligatory use of certain services (statistical data, ZUS) electronically. In the Strategy, while calculating the percentage index for Poland, assumed as a reference value was the average of indexes for three EU countries which were acknowledged as the best in a given category.

1.1.5. AMENDMENT OF THE ACT ON INFORMATIZATION OF ACTIVITIES CARRIED OUT BY THE PUBLIC TASK PERFORMING ENTITIES

On 12 February 2010 the Sejm enacted an Act on amendment of the act on informatization of activities carried out by the public task performing entities and certain other acts¹. The most important amendments were those introduced after 5 years to UINF².

¹ Journal of Laws No 40, sec. 230

² Journal of Laws No 64, sec. 565

The act determines the principles of accomplishment of tasks resulting from the developmental Strategy of the information society in the Republic of Poland. Article 12a added to UINF says precisely that responsible for working out and implementation of the Strategy is the minister competent in informatization (i.e. presently the Minister of the Interior and Administration) and that it must not be changed more often than once a year. The Strategy is to be used to create the State Informatization Plan³, the Plan being an executive act to the act on informatization of activities carried out by the public task performing entities. On the other hand, for implementation of the State Informatization Plan, the information projects are established. Revoked were articles 8, 9, 10, 11 related to supra-sector and sector information projects. Preparing applications for supporting the projects we should be aware that subsidies may be granted when subsidized undertakings are used for achievement of purposes listed in the Plan and Strategy. The added art. 3a of the act determines which requirements should be met by the documentation which has to be prepared by the applicant for subsidizing.

One of the basic principles of implementation of projects supported by EU funds is the obligation to maintain the investment for five years after it is completed. The authors of changes in the UINF provisions, administrative proceedings code, also those preparing a new act on electronic signatures had to include this condition of maintaining the already implemented or being still implemented information projects.

Other newly added articles deal with appointing external experts and evaluating their expertises by the Informatization Council.

The amended UINF combines PIP and SRSI.

1.2. IT PROJECTS ON RANKING LISTS

In 2007 the list of key projects related to the development of electronic administration, made up by the government, consisted of 48 items. After verification by the Ministry of the Interior and Administration of the new government, only 15 projects remained, 33

³ At present the Council of Ministers regulation is in force the State Informatization Plan for 2007-2010

of them were eliminated because either they were planned wrong or their completion date coincided with the tenure of the consecutive government which wanted to introduce its new rather than continue the predecessors' undertakings. The present ranking of the projects, prepared by the Ministry of the Interior and Administration (according to the advancement of implementation), on the basic list contains 18 items. Detailed information is placed only for 8 of them. No wonder, under these circumstances, observers of the State informatization and recipients of new e-services have the right to fear if their expectations can be satisfied without the continuity of activities of consecutive governments.

The Institute of e-Government at the Weseda University in Japan in cooperation with the e-Government Studies Centre in Asia Pacific Economic Cooperation (APEC), worked out a report – The 2008 Weseda University World e-Government Ranking according to 4-year monitoring and evaluation of availability of public administration services provided electronically by 34 countries. Poland was not evaluated in the Weseda University ranking, but it contained several important recommendations for the e-government developing countries which we can use in terms of management through projects. These include, among others, the need to establish new measurable purposes and introduce “accelerators”, so that more systems providing on-line services in administration arise, and the projects' precise defining specific control results and deadlines containing simple and concrete formulations, as e.g.: “the project has to be completed by 31 December 2010.”

“The quality of e-administration services offer has been investigated for many years by the Brookings Institution from the United States. Last year that organization investigated 1667 public administration web sites from 198 countries. Poland came 110th, a year before we were placed 57th. What happened within the year? The explanation is simple: our e-administration was not changed at that time, neither was the evaluation by points: both in 2007 and 2008 Poland obtained 32.7 points of 100 feasible. In the same period, unexpected competitors jumped out of the information backwardness: Ghana came 13th from 75th place, Togo 19th from 189th, Columbia 23rd from 59th, Togo 42nd from 165th.”

1.3. IT PROJECT IN BUSINESS AND GOVERNMENT UNITS

Each private company runs the business on its own way. In case of a private business, freedom of action plays the most important role, company can take up all that is not forbidden by law and has to perform all duties according to parliament's acts. In a public sector situation is different – according to the Republic of Poland Constitution – no freedom of action is allowed - public authorities perform their activity based on and within Polish law with an obligation of equal treatment for all partners.

In case of public administration according to the Constitution of Poland, there is no freedom of action- „public authorities operate on the basis of and within the law and there is an obligation to treat all people equally.”

Managing this type of organization is different and big differences between IT projects in the public sector and the business sector appear.

IT Project in business is strongly related to the strategy of an organization. The role of IT departments changes now from typical services to real business partner playing an important role in creating business value and company growth. Business and IT are connected and focused on serving business unit's needs. For enterprises applying IT governance the main objective of IT business projects is to obtain benefits or achieve better market place position.

Choosing between different IT projects investments and trying to achieve the project with best opportunities, organizations use tools like ROI, NPV and risk assessment for each project. Those projects are open to risk taking focused on financial and service outcomes. In most cases projects that show strategic importance to the enterprise are chosen although some projects can be run without ROI analysis – those are the projects that are driven by regulatory demands compliance or are basic business cost like web presence. In most cases the projects while delivery are even unknown for the shareholders and clients, so if they are cancelled or fail, hardly any information in media can be found.

IT projects in public bodies are mostly driven by and must follow demands of national government an EU legislation. The public administration body has no need to determine its position in the market. Outcomes associated with the ITC investments

may rather be important for a particular political party because the group policy or position of the political market is not fixed and a fine IT success may be a great value in next elections. It is important to notice that the objectives set for the public sector are rather of a qualitative nature rather than quantitative. Efficient services for residents and state authorities are extremely important to the interests of citizens, but difficult to quantify. In most cases IT systems are linked to departments and design to deliver better service or improve efficiency so they are not in the competition with other projects but they should interact properly. They have multiple aims that make it difficult to find measurable factors of success. According to POST Report 200 July 2003 Government IT projects Analysis of the problem, those projects highly attract media and public attention. The need for public accountability can lead to a risk averse culture in government⁴. Very often public authorities present new IT projects at the very first stage as powerful tools helping to solve administrative procedure and after the project delivery, it turns out that public expectations are not met. The policy and legislation process is another important factor causing changes in IT projects making that managing very difficult and more open to failure. The government body has no capacity to variance, everything is subject to a rigid top-down, statutory regulations - requirements to control how public money is spent. In procurement, through negotiations, you can negotiate only a strictly defined set of factors deals.

To assess and run the IT projects business organizations acquire well trained professionalizes with a big experience.

Government departments and units usually include some IT staff, but most often they are not very well familiar with the newest IT developments and solutions and have no experience with project management. Project managers are selected not on ability (skills, experience, leadership) but by grade. Public authorities' IT teams still need to discover the role of a better leadership, good relations with suppliers, management of risk and user involvement in IT projects. Public bodies need trained or new group of IT managers familiar with up to date technology and process of IT implementation to actively manage all IT suppliers.

⁴ POST Report 200 July 2003 Government IT projects Analysis of the problem
<http://www.parliament.uk/documents/post/pr200.pdf>

1.4. IMPLEMENTATION OF INFORMATION SOCIETY IN EUROPE - E-SERVICES RECOMMENDED BY EUROPEAN COMMISSION

In 1999 A Green Paper on public sector information in the Information Society presented a topology of Electronic Government services introducing 3 main types [GPE98] and possible services for everyday life, tele-administration and political participation:

- Information services – browsing, sorting and finding an information on demand;
- Communication services – using email, chats, bulletin and other for interactive forms;
- Transaction services – creating interaction between different stakeholders like on line reservation and shopping, electronic submission of forms and electronic vote while elections.

To measure the implementation of Information Society in eEurope targets some quantitative and qualitative indicators for benchmarking had to be established. In the year 2001 “eGovernment indicators for benchmarking eEurope” was presented with a set of 20 main public services (previous list consisted of 25 services) divided in two groups for citizens (12) and businesses (8) and 4 indicators for measuring them. For the eGovernment indicators, the following model was proposed:

- Stage 1 Information: online info about public services,
- Stage 2 Interaction: downloading of forms,
- Stage 3 Two-way interaction: processing of forms, incl. authentication,
- Stage 4 Transaction: case handling; decision and delivery (payment).

The methodology proposed for measuring the degree to which a service is available online is based on the method developed by the Dutch government. More about the methodology used can be find in [EGO02]. For benchmark tests in measurements in 2007 and 2009 new 5th stage was added- Targetisation/Automation - customer oriented services.

The list of basic public services with the Polish result in 2007 and 2009 for citizens covers following services [EGO02]:

1. Income taxes: declaration, notification of assessment
(Poland in 2007 classified as 4/5, 2009 5/5)
2. Job search services by labour offices (Poland in 2007 classified as 4/4, 2009 5/5)
3. Social security contributions (3 out of the following 4)
(Poland in 2007 classified as 1-2/5, 2009 3/5):
 - a. Unemployment benefits
 - b. Child allowances
 - c. Medical costs (reimbursement or direct settlement)
 - d. Student grants
4. Personal documents (passport and driver's license)
(Poland in 2007 classified as 1-2/4, 2009 2/4)
5. Car registration (new, used and imported cars)
(Poland in 2007 classified as 1-2/4, 2009 2/4)
6. Application for building permission
(Poland in 2007 classified as 1-2/4, 2009 3/4)
7. Declaration to the police (e.g. in case of theft)
(Poland in 2007 classified as 1/3; 2009 1/3)
8. Public libraries (availability of catalogues, search tools)
(Poland in 2007 classified as 3/; 2009 5 4/5)
9. Certificates (birth and marriage): request and delivery
(Poland in 2007 classified as 2/4 2009 2/4)
10. Enrolment in higher education / university
(Poland in 2007 classified as 2-3/4 2009 – not relevant)
11. Announcement of moving (change of address)
(Poland in 2007 classified as 1-2/4 2009 2/4)
12. Health related services (interactive advice on the availability of services in different hospitals; appointments for hospitals)
(Poland in 2007 classified as 0-1/4 2009 – 2/4)

The list of Public Services for Businesses covers following services

1. Social contribution for employees (Poland in 2007 classified as 4/4, 2009 4/4)

2. Corporation tax: declaration, notification (Poland in 2007 classified as 2/4, 2009 4/4)
3. VAT: declaration, notification (Poland in 2007 classified as 2/4, 2009 4/4)
4. Registration of a new company (Poland in 2007 classified as 2/4, 2009 4/4)
5. Submission of data to statistical offices (Poland in 2007 classified as 3/5, 2009 3/5)
6. Customs declarations (Poland in 2007 classified as 4/4, 2009 4/4)
7. Environment-related permits (incl. reporting) (Poland in 2007 classified as 0-1/5, 2009 1/5)
8. Public procurement (Poland in 2007 classified as 3/4 2009 4/4)

1.4.1. EPUAP A STRATEGIC OF POLISH E-GOVERNMENT PROJECT

At the beginning of a XXI century the information systems implemented in different departments of Polish public government were managed separately that caused a situation where each department had its own quite good but closed to others IT infrastructure. A rapid growth of more advanced (web based) systems, their integration and opening to the e-society was necessary and became a big challenge.

To fulfill this challenge and according to The Strategy for the Development of the Information Society in Poland until 2013 assumptions a project called ePUAP (eElectronic Platform of Public Administration Services) was established. Basic assumptions of the project were created in December 2002 as a part of Wrota Polski concept. At the beginning of 2005 a full program to achieve high availability and sophistication of public services for citizens and businesses was prepared and started. The project was run by CPI (Information Projects Centre Ministry of Interior and Administration) in 2005-2008. ePUAP portal is part of the platform ePUAP. The task of the portal is to provide information on public services, carried out via the platform. It is available now at www.e-puap.mswia.gov.pl.

Main eservices offered by the portal were:

1. Electronic communication between all users (citizens/business/government departments).

2. Catalogue of services to simplify finding with the relevant public resources.
3. Service to help citizens and businesses to coordinate the efficient implementation of complex public services.
4. Registration broker – service that allows public bodies to obtain information regardless of their location and current form.

The project was finished in 2008 and in order to continue the process providing access to more simplifying administrative procedures and changing them in to an electronic way a project called ePUAP 2 was established.

The amount of the grant for the project from the 7th priority axis of the Innovative Economy Operational Programme is 140 million pln (85% of eligible expenses comes from the European Regional Development Fund, 15% of the national co-financing)⁵.

Extending the EPUAP was in 2009 one of the strategic eGovernment priorities.

Registry box was a service opened to public in 2008. ePUAP does not offer any special public service, eServices provided by the platform must be provided by other institutions. In 2010 65 from 100 services are available. Next should be accessible soon among them: business application, an entry in the register of hotels, consent to the felling of trees, the march of time agreeing on a pilgrimage.

The use of those services still demands the electronic signature verified by a qualified certification. Announced in January 2010 trusted profile is still not working.

For proper functioning of the platform municipalities and other local government units must make services accessible. Some municipalities have already spent a lot of money on IT investments on well functioning web platforms for local services. The process to integrate ePUAP, which is the only trusted place to store profile and the local government services is necessary. A good interface and economical reason (maintain own websites and services on ePUAP duplication is too expensive) can make some municipalities to transfer their local services to the platform.

Trusted profile should be tested by future users. Therefore detailed instruction for citizens and Trusted Profile confirmation ePUAP points, on the ePUAP portal are

⁵ <http://cpi.mswia.gov.pl/portal/cpi/46/193/ePUAP2.html>

available now. The guides for interface to external IS seems to be the most important for integration with local government bodies IS [KOS10].

The legal regulations of ePUAP functioning are still under construction. For a legal wide usage of ePUAP services Ministry of Interior and Administration prepares draft regulation on the scope and terms of use of an Platform of Public Administration Services, that should effect mainly those public authorities who wish to use the platform to provide services electronically⁶. The regulation defines:

- setting up an account for ePUAP (that is a necessary condition to use a service),
- the way to run eService catalogue,
- terms of information exchange between ePUAP and other IT systems bodies performing public tasks.

The scope and terms of ePUAP were divided in regulation into three groups: members, the minister responsible for information technology and service providers Registered ePUAP user will be able to act on its own, as well as on behalf of various entities. ePUAP functioned since 2008, but there were no regulations governing the scope of its functioning. Draft regulation implies that registered users and operators keep their accounts established so far.

Development of electronic services in public administration now depends largely on increasing confidence in the platform ePUAP, widespread use of a trusted profile and change of the officials' mentality.

In case of public sector projects, the success depends on local governments and Polish government as these bodies decide on the matter that can be settled electronically and that requires a personal appearance in a specific office.

ePUAP is one of the first steps towards building a modern integrated public administration, whose central point is a citizen. The main threat to the project are not technical or economical problems, but a change in the approach of the institution from different ministries and their habits.

⁶ www.mswia.gov.pl

1.4.2. POLAND IN E-SERVICES BENCHMARKING TEST

Research conducted by the Economist Intelligence Unit commissioned by Oracle that the global economic crisis has not stopped the Polish government spending to invest in is ICT. According to the assumptions of officials in 2009, government spending on ICT in Poland (funding provided by the state, the EU and the private sector) increased compared to the year 2008 by 5-6 times and amounted to about 2 - 2.5 billion euro⁷.

The 8th measurement of eServices across Europe from November 2009 shows Poland as one of the EU27+ countries that comparing to 2007 made a big progress in sophistication of online basic 20 services availability – 20,6% growth (together with Latvia classified to the group of 15%-25% growth). This is a significant effort but still did not bring Poland to the average (76% in 2007 and 83% in 2009). In a 8th Benchmark Measurement two new indicators (eProcurement Availability and eProcurement process Benchmark) measuring the availability and visibility for the suppliers in main process phases were established. In eProcurement Availability Poland scored 61% (56% average for EU27+).

According to Eurostat data in 2009 the percentage of online availability of 20 basic public services that (indicator shows the percentage of the 20 basic services which are fully available online i.e. for which it is possible to carry out full electronic case handling) is equal to 53. It is much better than 25% in 2007 but still far from average for Europe (74% for 27 countries, 76% for 25 EU countries and 81% for 15 EU countries).

The Eurostat data shows that percentage of individuals using the Internet for interaction with public authorities in 2009 was equal to 18% comparing to 16% in 2008. Average for 27 European Countries 30%, 25 EU 32%, 15 33%. Some of activates in this area are presented in the table 1.2.

⁷ <http://www.oracle.com/global/pl/corporate/pressroom/2009/24112009.html>

Tab. 1.2. Citizens using the Internet for interaction with public authorities in 2009

Obtaining information	16.3 %
Downloading forms	11.5 %
Returning filled forms	5.4 %

Source: Based on http://epp.eurostat.ec.europa.eu/portal/page/portal/information_society/data/main_tables

Percentage of enterprises using the Internet for interaction with public authorities in 2009 was equal to 61% comparing to 68% in 2008. Average for 27 European Countries 72%, 25 EU 74%, 15 74%. Some of activities in this area is presented in the table 1.3.

Tab. 1.3. Enterprises using the Internet for interaction with public authorities in 2009

Obtaining information	53 %
Downloading forms	56 %
Returning filled forms	57 %

Source: Based on http://epp.eurostat.ec.europa.eu/portal/page/portal/information_society/data/main_tables

The analysis of the data in Electronic Business in Poland⁸ shows a very small growth of the Internet users searching an Internet for the information from the public administration web pages 29,4 % comparing 28,6% w 2008) - for downloading official forms, 9.7% to return filled forms.

In the 8th measurement of eServices across Europe emphasized that the high evaluation of Polish results from the latest strategy in the ICT area under the name "Digital Poland", published in December 2008 by the Ministry of Internal Affairs⁹. But the report also pointed to problems with Internet access of Polish households that means lack of opportunities to use public e-services.

⁸ M.Kraska Electronic Business In Poland Report 2009, Biblioteka Logistyka Poznań 2009

⁹ <http://www.crn.pl/artykuly/raporty-i-analizy/rzadowe-wydatki-na-e-government/#axzz11Z3n9E7f>

According to Information Society Indicators¹⁰ in 2009 the number of citizens and enterprises having an access to an Internet in Poland is growing. Detailed indicators are presented in the table 1.4.

Tab. 1.4. Internet access in Poland

Percentage of households with Internet access	59 %
Percentage of enterprises with Internet access	91 %
Percentage of individuals using the Internet at least once a week	52 %
Percentage of households with a broadband connection	58 %
Percentage of enterprises with a broadband connection	51 %
Percentage of individuals having purchased/ordered online in the last three months	18 %
Percentage of enterprises having received orders online within the previous year	5 %

Source: Based on http://epp.eurostat.ec.europa.eu/portal/page/portal/information_society/data/main_tables

“Digital Poland” strategy assumes the provision of broadband throughout the country by 2012. In case of a successful implementation of the project should change the negative indicators. As was mentioned at the beginning, Poland is investing heavily in Information Society growth. Except ePUAP among the biggest projects are: the e-Health, a digitization of all public health data (from 877 million fund pln), and pl.ID - an electronic identity card, which enable citizens to access public services online (370 million fund pln).

1.5. BARRIERS TO SUCCESSFUL IT PROJECT IN E-GOVERNMENT

The number of successful IT Project in public sector is growing in Poland, but some of them still fail. In many cases Government's IT projects faces the same barriers, which slow down its development in business. Those barriers can be defined as potential problems for the development and implementation of ICT systems. There are many types of barriers. They can be classified according to their origin into regulatory,

¹⁰ http://epp.eurostat.ec.europa.eu/portal/page/portal/information_society/data/main_tables

technical, economic, social, nature of IT technology and other groups. In some cases a barrier can belong to more than one group.

Obvious technical barriers include the lacking communication infrastructure (ex. problems with internet access of Polish households) particularly in small societies and the poor maturity of ICT tools (or client application to a wider systems) which are needed by local communities. Lack of coherent concept of IT for whole public sector is another important barrier. Technology barriers come exactly from the nature of IT and its rapid growth.

Regulatory barriers are related to not completed legislation process (ex.-lack of directives).

Social change in public service is one of the cornerstones foundation of e-government. The new way of e-government administration demands client as a main source of interest. The basic role in a concept of e-government administration should play well defined procedures triggered by e-customer question or demand. This demands a wide reengineering process" (and changing the way civil servants treat the clients from a superior "power-king" to a subordinate "common servant which is difficult, as a people quickly get used to be important chain of the decision process and do not want to loose the power). From the IT point of view it demand a good front-office (save and easy to navigate IS systems and portals) and back-office (fast and effective internal and external (between different levels and units) procedures and hardware to service the client demand. All of his requires a lot of IT and processes experts work and a high level budget what classifies this barrier to economical and technical group.

The main role in all e-services plays the system of forms schema that are supposed be stored in Central Documents Schemes Repository – Repository of e-government forms and services and when it is necessary in local repositories.

The electronic scheme of the document should include:

- XSD – definition of document structure,
- XSL- definition of document presentation,
- XML - definition of document scheme.

The forms are created by the software applications according to document scheme. Many applications can create different forms for the same scheme, but because of the created definitions it will be possible to interchange data even after a long time. The proper electronic schemes is a basic condition later IT systems and eServices functioning. A lot of schemes are now available at ePUAP platform.

1.6. CONCLUSIONS

Evaluation of information projects implementation in the public administration projects should be carried with a set of the best ITIL practices, the recent Cobit versions - standard of IT management order in organizations, norms ISO/IEC 20000 (the only official norm in the information services management area) and 27001 (international standard determining the requirements for information safety management systems) and methods of management of PRINCE2 projects in view of its worldwide origin in the public sector.

The project feasibility studies which are enclosed to the application for subsidizing should clearly indicate the significance of the objective of conducting a given undertaking with reference to the beneficiary's strategy (in PRINCE2 the reference to the strategy of the entity submitting the project is a condition to acknowledge the objective as significant). However, only those can refer to the strategy who have it. The applicants' feasibility studies should also clearly show whether or not they manage the risk. It is not enough to identify the risks, one should also know how to use a wide variety of instruments available: techniques, methods and software, providing quantitative and qualitative services of identified risks.

Polish administration has already implemented the management through projects in its sector, having first heard about the effectiveness of this form of introducing changes in business area. Numerous projects have also been set up with the aim to efficiently introduce new e-services for applicants. Now the administration should get prepared for changes in the policy and organization of its functioning, financing training courses connected with project management and particularly coordination,

because a situation arose where many interrelated undertakings have to be managed simultaneously.

Polish administration still learns how to manage informatization through projects. Consecutive informatization plans should not be created to satisfy governmental departments' ambitions, but they should contain reasonable, measurable activities without unspecific formulations like "better" or "more effectively".

Aiming at working out professional project standards we should eliminate the cardinal sin in accomplishment of undertakings, i.e. the lack of strict association of objectives with cost-effectiveness analysis, which causes most errors in it.

Knowing that the skills and experience of those who conduct major IT projects significantly affect their success, we should carefully select the management staff. For in the administration sector the greatest threat for implementation of big and long-lasting projects (longer than the consecutive election tenure) is susceptibility to staff changes of political nature.

The need to discuss the undertaken activities and evaluations in numerous groups creates the feeling of the decision anonymity and thereby increases the susceptibility to changes, because in consecutive state governmental teams there may be no people who want to continue the projects launched by their predecessors.

Amending every 5 years the act which initiated the e-administration in Poland is insufficient and does not fit in with the Public Procurement Law (act of 29 January 2004; Official Journal of Laws 06.164.1163; hereinafter: PZP). The PZP act which enables the „fulfillment” of the administration's information requirements, does not harmonize with the specificity of information products. This must be true, as the state does not want to use the competitive dialogue form to clarify the description of the procurement object. In effect, tenders drag on endlessly.

The factor in management of fundamental projects are people. Our administration needs first of all the ability to set up stable project teams managed by competent project managers from the ranking list of professional project managers or those holding high competent positions in their entities. Another important factor is also the ability to continue and coordinate many undertakings in compliance with the *raison d'état*.

REFERENCES

- [BEN09] *8th Benchmark Measurement* November 2009
- [CHM10] Chmielarz W.: *Studium rozwoju systemów e-administracji w Polsce*, <http://web.ae.katowice.pl/stanley/konferencja/pdf/Chmielarz.pdf>, 6.10.2010
- [EGO02] *eGovernment indicators for benchmarking eEurope*, http://ec.europa.eu/information_society/eeurope/2002/action_plan/pdf/egovindicators.pdf
- [GOV10a] *Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne*
- [GOV10b] *Ustawa o zmianie ustawy o informatyzacji działalności podmiotów realizujących zadania publiczne i niektórych innych ustaw*
- [GPE98] *Green paper on public sector information In the information society*, ftp://ftp.cordis.europa.eu/pub/econtent/docs/gp_en.pdf
- [KNI08] Książ A.: *Szyldy zamiast projektów. Informatyzacja administracji*, W: Elektroniczna administracja, listopad/grudzień 2008
- [KOS10] Kosieliński S., Tomkiewicz M.: *Na zaufanie trzeba sobie zasłużyć*, Computerworld, 6 lipca 2010
- [MSW10] *Strategia rozwoju społeczeństwa informacyjnego w Polsce do roku 2013*
- [PAP08] Papińska-Kacperek J. redakcja: *Spółeczeństwo informacyjne*, PWN 2008
- [RMR06] Rozporządzenie Rady Ministrów w sprawie Planu Informatyzacji Państwa na rok 2006
- [RMR07] Rozporządzenie Rady Ministrów w sprawie Planu Informatyzacji Państwa na lata 2007-2010

ICT e-Infrastructure for IT Platform of Managing Research Processes at Lublin University of Technology

The development of information and communicational technology (ICT) decreases the distance of geographical space between men of science who, while being at their places of work, are able to work and cooperate in solving specific research issues. The research potential of various foreign and domestic centers is connected by telecommunication and information technology infrastructure that consists of:

- Human resources (number of employees, structure and their knowledge, experience and skills) together with their connections (relations with other people)
- Equipment resources – computers of various type: personal (PDA, handheld PC, smartphones, palmtops), mobile (laptops, notebooks, netbooks, ultra-mobile PC), desktop PC, working drives, central computers (mainframe), supercomputers and network accessories.
- Program resources – software of various type enabling interaction between human and digital machine in a form of realization of services (operating systems, tool programs, specialist programs and general use programs).
- Institutional character expressed in features and abilities of a specified institution as a whole to face challenges adequately to conducting research processes in the 21st century. Institutional character is understood as a specific organizational resource that results from experiences of a definite

institution that is incorporated into the geographical, social, economical, informational and regional context. Institutional character is a tradition and reputation of an institution that is able to fulfill statute goals in an organized way. The individual organizational units are able to act in an effective and synergetic way and therefore contribute to achieving desired goals.

The use of the aforementioned resources should be oriented to solving real problems which, when transforming into research problems, should contribute to effective solution of issues during an implementation process and therefore efficient achievement of predefined goals.

Lublin University of Technology (LUT) is one of the many higher education institutions located in the east part of Poland. Certain research subjects result from its tradition and character of its activities. LUT also solves engineering research problems. In order to measure up to all types of challenges, the higher education institution should possess definite institutional as well as program and equipment resources that would allow it to conduct scientific research at the European level, especially during the realization of team-oriented works spatially spread in Poland and abroad alike.

In connection to conducting various engineering research works, the personnel of the Faculty of Management (FM) should be provided with favorable scientific conditions understood as elimination of barriers in access to modern telecommunication and information technology infrastructure and assuring functioning of network services during realization of research. Therefore, in April 2010, the personnel of the Faculty of Management prepared and submitted in the Ministry of Science and Higher Education the application project entitled: "ICT e-Infrastructure for Information Technology Platform of Managing Research Processes at LUT" within the frames of priority axis 2 B+R sphere infrastructure for activity 2.3 Investments connected with the development of information technology infrastructure of the Operational Program Innovative Economy 2007-2013 science (OPIE) [INS10] The name of this platform has been defined as the Information Technology Platform of Managing Research Processes (ITPMRP) enabling conducting research works. It concerns: creation of information technology infrastructure, creation of information

technology resources, use of advanced applications and network services in conducting research works.

What contributed to the creation of the Project were insufficient equipment and program resources that have been in possession of the Faculty of Management and which have made impossible to conduct specialist research by scientific personnel as well as to cooperate with external scientific centers of Poland and the European Union.

2.1. EQUIPMENT AND PROGRAM NEEDS FOR THE PURPOSES OF CONDUCTING RESEARCH WORKS AT THE FACULTY OF MANAGEMENT

The Employees of the Faculty have been conducting various research works for many years. Various necessary research devices have been purchased within the frames of these works, including computer equipment and software. For the sake of the variety of the financial means and lack of their synchronization in time, the necessary equipment and program resources have been purchased at request, ad hoc. The lack of coordination of purchase activities, variety of different kinds of equipment, exploitation materials, different time of their exploitation and different needs of the employees have resulted in spreading out calculative power in the whole Faculty building and therefore an increase in disproportion between employees' devices and devices that operate employees, i.e. servers. The tightness of solutions and lack of investments in servers and network infrastructure resulted in moral (economic) and most importantly technical wear. The purchase of servers and network accessories from the research projects could not be justified enough due to the subject matter and amount of necessary expenses. Therefore, today, the employees do not have an access to advanced network services and cannot create and share databases, conduct team work and most importantly they cannot create neither backup copies nor manage archives. The employees lose all data in case of equipment failure or program instability

The diagnosis of barrier in access to modern telecommunication and information technology infrastructure along with assurance of functioning of network services are

regarded as a part of a second priority PO IG “*B+R Sphere Infrastructure*” [SZO09, NSR09, NSR09A]. The submitted project is a chance for the Faculty of Management of LUT to create unified standard in equipment and program accessories as well as standard in access to network services for conducting research works. The improvement in quality of equipment and programs as well as the use of network services accomplished thanks to implementation of the project should ensure favorable scientific conditions for every scientific employee of the Faculty of Management and for scientific employees who cooperate with LUT and with employees from the research environment alike.

The equipment and program resources submitted in the application should overcome the technological barrier and cause interest in new scientific problems which, for the sake of the state of present equipment and program resources, could not be conducted at the Faculty of Management. The new research directions and their results will be published via purchased information technology equipment and necessary software.

On the other hand, the purchased e-infrastructure should create openness of the higher education institution to employees from other higher education institutions, including the foreign ones, as well as openness to conducting research using external research centers.

2.2. THE SCOPE OF THE PROJECT

The Project concerns establishing modern information technology infrastructure in IP ver. 6 technology at the Faculty of Management as well as realization of network services for conducting research at the European quality level. The e-infrastructure will enable creation of multipurpose, modern platform that will allow to carry out research and manage research processes at the Faculty of Management at LUT.

The established Information Technology Platform of Managing Research Processes will be used for conducting research works by all employees in various fields. It is a part of a Priority no. 2 PO IG and it concerns. *Establishing information technology infrastructure* (sub-activity 2.3.1 PO IG), *Establishing information*

technology resources (sub-activity 2.3.2), *Using advanced applications and network services* (sub-activity 2.3.3) [REG10]. The purpose of the platform is to assure proper conditions for development of the scientific personnel using permanent and safe access to advanced information technology infrastructure as well as enabling conducting modern research and assuring connectivity with international scientific information technology networks. Because of the fact that there are many research works conducted at the Faculty of Management with the use of various information technology equipment, the new information technology platform will allow to manage personnel resources of the Faculty effectively and control research processes.

The project encompasses the following activities:

- a) Creating a central server room together with providing it with proper equipment and assuring physical, electronic and program safety.
- b) Creating an independent local area network (LAN) with dedicated electric network to ensure permanent connectivity and unfailing work.
- c) Purchasing equipment and software for the final users in order to create and use network services
- d) Implementing ICT services using connections in client-server architecture via dedicated intranet service among others.
- e) Providing employees with professional training in the field of development, exploitation and managing IT infrastructure.
- f) Promoting project in a form of Public Relations activities, creating a public access Internet service and promoting results of B+R works gained thanks to subsidized infrastructure.

2.3. TECHNICAL ANALYSIS OF RESOURCES, KEY ISSUE, AIMS OF THE PROJECT

The technical analysis of available resources of the Faculty of Management indicates that:

1. Because of the fact that the current central server room has been created by assigning a part of the hall for that purpose, it does not meet requirements of the

room that could be called a central server room with running servers. When the room was established, only two computer-servers were meant to be created there (1995). One didactic server with UPS support and another one for research and didactic employees. With the passing of time, various departments have purchased different computers and adapted them as servers and for the purposes of constant work in a room. The equipment has also been purchased from different means each and every time and therefore the network has been created ad hoc. To sum up, it is very difficult to use current room for further expansion as it is small, it is impossible to install air-conditioning there and it is located very far from the switching station.

2. The analysis of equipment resources indicates that the current equipment is technologically obsolete. The sum of all of the servers' discs is 544 GB (PAM) and approximately 10 GB of operating memory. Nowadays, a mobile computer worth 6.000 PLN has a hard disc with 500 GB of space (the same capacity as all the installed servers) and 4 GB of operating memory (two times fewer than all the installed servers). The available resources prove that calculative powers are relatively small in relation to the needs of employees of the Faculty of Management concerning conducting research works.
3. The analysis of program resources in the field of operating systems indicates that those are various operating systems that are being constantly updated. However, because of the insufficient equipment resources, they operate with high delay in relation to the expectations of the users.
4. The level of the current use of program and equipment resources is very high and the exploitation of the devices is enormous due to limitations of operating memory capacity and low capacity of hard discs. The lack of free output capacity limits the possibilities of conducting scientific research and development works. On the other hand, the level of exploitation of ICT services is relatively low and concerns only the use of e-mail and remote work on a server with Windows operating system.

To sum up, due to the high failure frequency of the existing equipment, the scientific workers encounter barriers in access to reliable equipment and program resources

and opportunities of cooperation with scientific employees from their country and from the European Union. The lack of stability of the equipment results in software instability and therefore causes high risk of loss of important data and makes research realization impossible. The lack of means of archiving resources as well as lack of possibilities of storage of research works results causes significant spreading out data among various personal computers in the building. The aforementioned phenomenon makes effective teamwork impossible (practically, there is no teamwork because of the lack of proper software). The level of the use of the equipment and program resource is very high and meets maximum level of exploitation of the machines. The low calculative powers make research time-consuming and vulnerable to the loss. The Faculty cannot realize new research works due to the lack of sufficient output capacities of the computer equipment.

Barriers in penetrating into the domestic and European market of scientific research are the key problem defined in the Project. They are significant obstacle in realization of the projects that are realized within the frames of the domestic and international market. Furthermore, they cause measurable financial loss (in a sense of category of alternative costs) in gaining European means for scientific research. This causes the increase in distance between scientific employees of the Faculty of Management and scientific environment. The aim of the project is to create ICT e-infrastructure for the functioning of the Information Technology Platform of Managing Research Processes at LUT that will allow to solve research problems of the employees of LUT. The main aim of the project consists of the group of goals:

- overcoming technological barriers in conducting scientific research
- decreasing input barrier of penetrating into the domestic and European scientific research market thanks to increase of research and scientific potential of the institution using ICT e-infrastructure corresponding to the level of world's standard.
- preparing a higher education institution to be an active competitor on the international market in the field of creation of competitive economy.
- establishing and developing connections between the field of research and enterprises.

The general aim is to enable conducting modern research using ICT e-infrastructure while simultaneously providing connectivity with international research teams via telecommunication and information technology networks.

2.4. ORGANIZATIONAL STRUCTURE OF THE PROJECT IMPLEMENTATION

The project will be realized mainly by the employees of the Faculty of Management and two additional persons hired from Technical University of Silesia or Technical University of Radom.

The structure of the project is organizationally isolated. There are 25 persons who will be responsible for the implementation of the project. They will be experts assembled in 5 teams. The organizational scheme of the project is presented in the Fig. 2.1 and the functional structure of the experts in the Fig. 2.2.

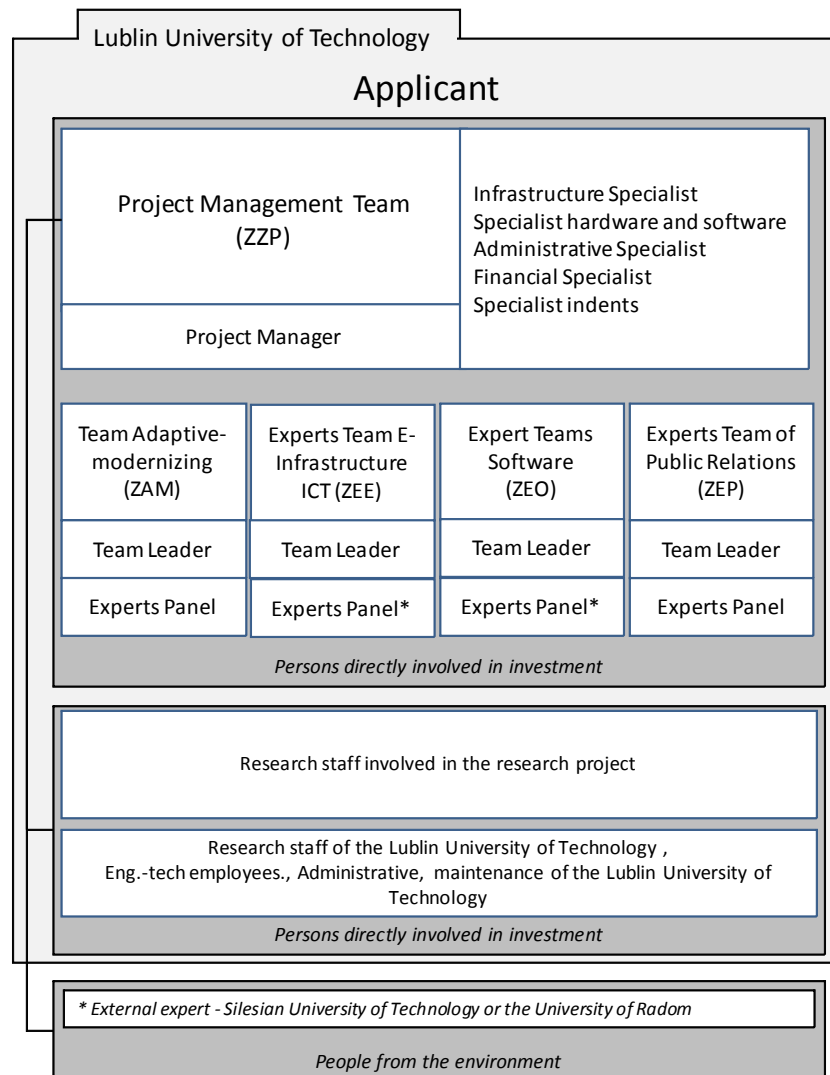


Fig. 2.1. Organizational scheme of the teams responsible for the implementation of the project along with advisors using ICT infrastructure

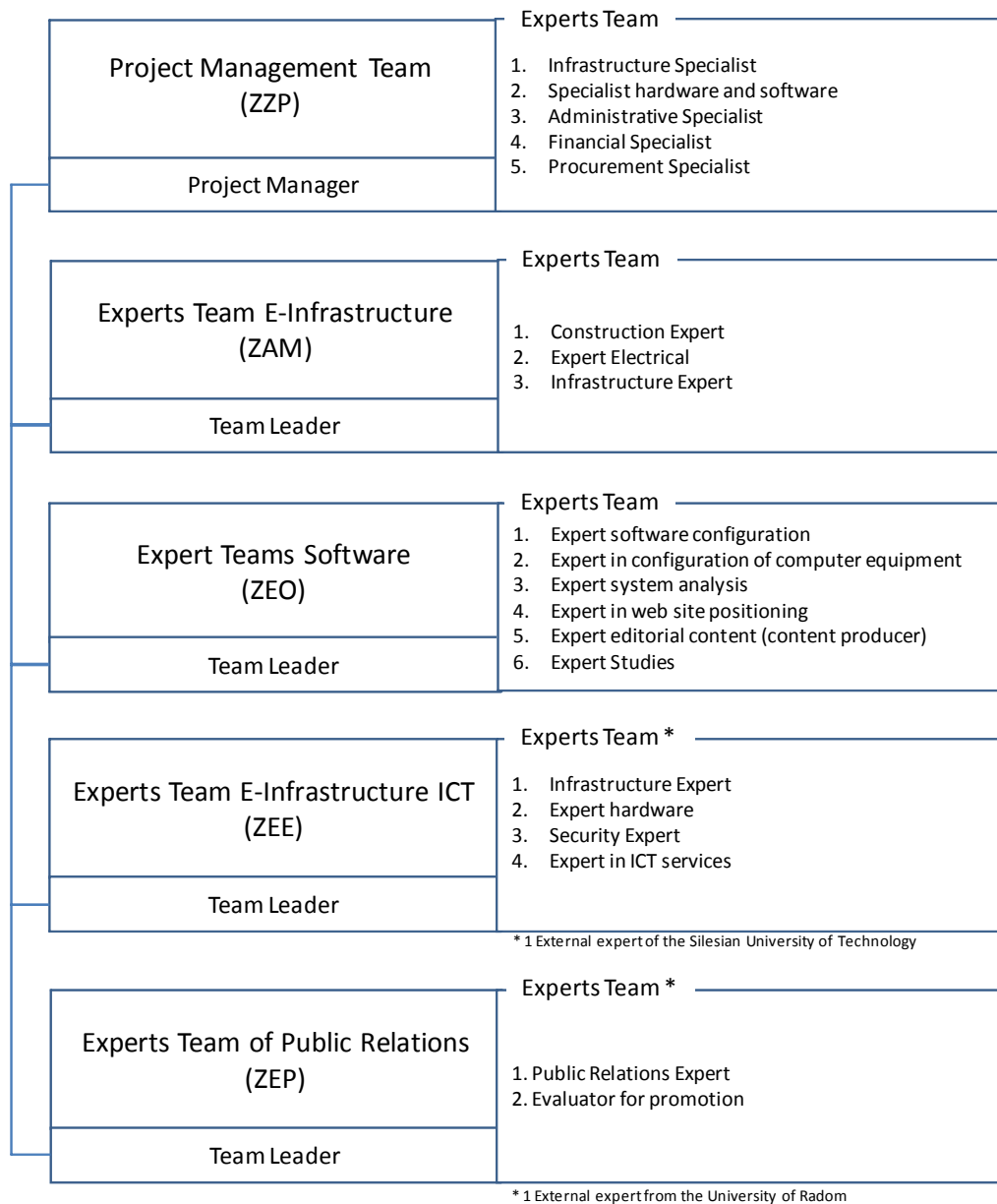


Fig. 2.2. Functional structure of the teams responsible for the implementation of the project

The team responsible for the Project management is the Team of Project Management (TPM) which will be called into being in the structures of the Mover. The TPM structure results from the character and range of the project that is to be realized. The names of the positions are defined functionally and are logical in terms of the future character of activities of persons holding those positions. The experts from a given team may also become experts in other teams (the position contests are open).

The project will be realized in accordance with regulations of the statute on Public Order Law and the policy of equal chances of women and men, prevention of any form of discrimination, providing access for the disabled [RRE06]. Due to that fact, the personal potential of the project should be adequate to the structure of tasks, range of activities and responsibility.

2.5. SCHEDULE OF THE PROJECT REALIZATION

The schedule of the project realization has been prepared in the MS Project 2003 and will also be used during the realization of the investment (Fig. 2.3-2.4).

0	▣ e-Infrastructure for ICT for Informatics Platform Process Management research at the Technical U
1	▣ Work Packages 1. Project Initiation
1.1	Preparation of tender documents for personal bonded clusters (ZZP, ZAM, ZEE, Zeo, PER)
1.2	The order of the tender to the Department of Procurement PL, invitation to tender, selection of teams
1.3	Appointment of Project Management Team - the Order of the Rector of the Technical University of Lublin
1.4	Develop principles of the project, scope of activities needed Teams
2	▣ Work Packages 2. Project Management Team (ZZP)
2.1	The Processes of project management (reporting, documentation management, change management)
2.2	Appointment of Team Adaptive-modernizing (ZAM)
2.3	Appointment of the Panel of Experts E-ICT infrastructure (ZEE)
2.4	Appointment of the Panel of Experts Software (ZEO)
2.5	Appointment of the Panel of Experts Public Relations (PER)
2.6	Formalization of expert teams
2.7	Internal monitoring and evaluation of the project, preparation of periodic reports and accounts
2.8	Request for tender for the audit of the Department of Procurement
2.9	Conducting an external audit
2.10	Preparation of final report
2.11	Closing financial, organizational, technical project
3	▣ Work Packages 3. Adapting of rooms for network
3.1	Preparation of tender documents for the work of modernization and adaptation
3.2	▣ The contract tender adaptation of premises for the Department of Procurement LUT
3.2.1	Order for the implementation of the works - Works modernization and adaptation
3.2.2	Order to the work of structured cabling and install a separate power supply
3.2.3	Order for the implementation of the server room air conditioning work
3.2.4	Order to the work space server management and security servers
3.3	Construction work of modernization and adaptation
3.4	Implementation of structured cabling and wiring
3.5	Metering and collection of structural cabling and wiring
3.6	Implementation of the air conditioning system
3.7	Implementation of the server room equipment and security servers
3.8	Reception room server
3.9	Solution Team ZAM
4	▣ Work Packages 4. Purchase of hardware and software for network
4.1	▣ Preparation of tender documents server hardware and software
4.1.1	Hardware servers
4.1.2	Specification of server software
4.2	▣ The order of the tender to the Department of Procurement LUT, invitation to tender, the selection of contractors
4.2.1	Order hardware servers
4.2.2	Order the software server
4.3	Order processing hardware
4.4	Implementation of orders for server software
4.5	Receive substantive orders
4.6	The creation of the demilitarized zone (DMZ)
4.7	Installation, configuration of hardware and software, IP 6, testing, V&V
4.8	Acceptance tests, the transfer of hardware and software for network
4.9	Drawing up a list of as-built documentation software configuration
4.10	Establishment of operational software built documentation
4.11	Receive hardware and software for network

Fig. 2.3. Schedule of the project realization

5	▣ Work Packages 5. Implementation of the installation of the LAN and power supply
5.1	Preparation of tender documents for project implementation and LAN outlet
5.2	The order of the tender to the Department of Procurement LUT, invitation to tender, the selection of contractors
5.3	Implementation Network Installation
5.4	Metering and collection of structural cabling and wiring
5.5	Establishment of technical documentation; LAN network installation
5.6	Establishment of technical documentation; electrical installations
5.7	Transmission Network Installation
6	▣ Work Packages 6. Purchase of hardware and software for users
6.1	Preparation of tender documents for equipment and software (Stage I)
6.2	The order of the tender to the Department of Procurement PL, invitation to tender, selection of contractors (stage)
6.3	Installation, configuration of hardware and software, testing (stage)
6.4	Transmission of hardware and software, acceptance testing (stage)
6.5	Preparation of tender documents hardware and software (Stage II)
6.6	The order of the tender to the Department of Procurement PL, invitation to tender, selection of contractors (second stage)
6.7	Installation, configuration of hardware and software testing (phase II)
6.8	Transmission of hardware and software, acceptance testing (phase II)
6.9	Establishment of as-built documentation of hardware and software users
7	▣ Work Packages 7. Implementation of ICT services
7.1	Implementation of ICT services, software testing client-server, Acceptance testing
7.2	Preparation of documentation intranet
7.3	The order of the tender to the Department of Procurement PL, invitation to tender, selection of contractors (stage)
7.4	The production process of software implementation of an intranet
7.5	V & V, internal tests, acceptance tests intranet
7.6	The acceptance of ICT services
8	▣ Work Packages 8. Training for staff in the development, operation and management ICT
8.1	Preparing for recruitment and training schedule of ICT
8.2	The order of the tender to the Department of Procurement EN
8.3	Implementation of training
8.4	Evaluation of the effectiveness of training
9	▣ Work Packages 9. Promotion Project (Public Relations)
9.1	Preparation of media plan
9.2	Preparation of documentation web site for public access
9.3	The order of the tender to the Department of Procurement PL, invitation to tender, selection of contractors (stage)
9.4	The production process of software implementation of web site
9.5	Implement web site for public access
9.6	V & V, internal tests, acceptance tests
9.7	Seo search engine terms of service - action SEO (search engine optimization)
9.8	Public Relations Activities promote the project and the results of the BR
9.9	Internal Promotion Project Internal Public Relations
9.10	Establishment of media-built documentation plan

Fig. 2.3. Schedule of the project realization (contd.)

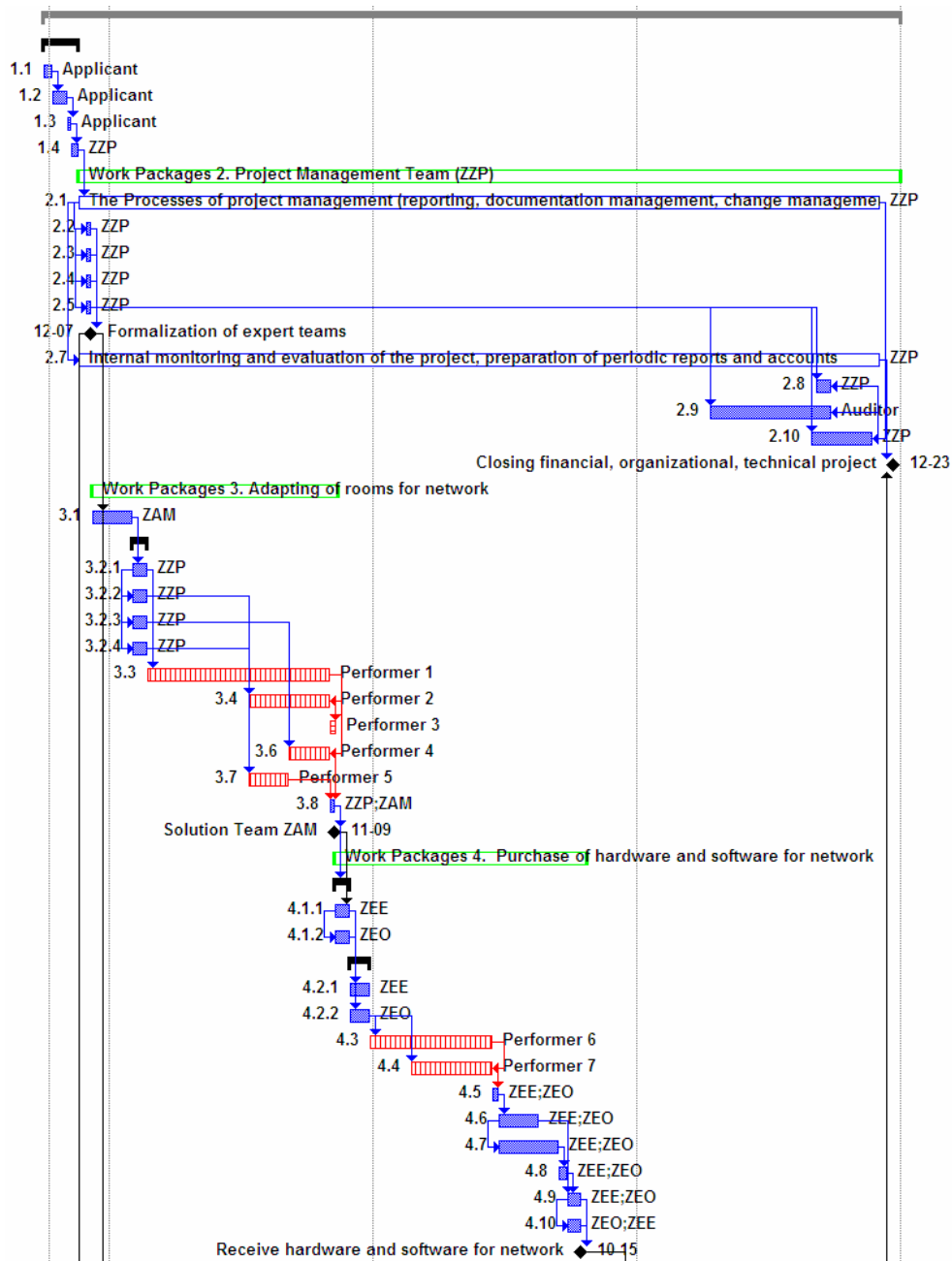


Fig. 2.4. The Gantt graph of the project realization

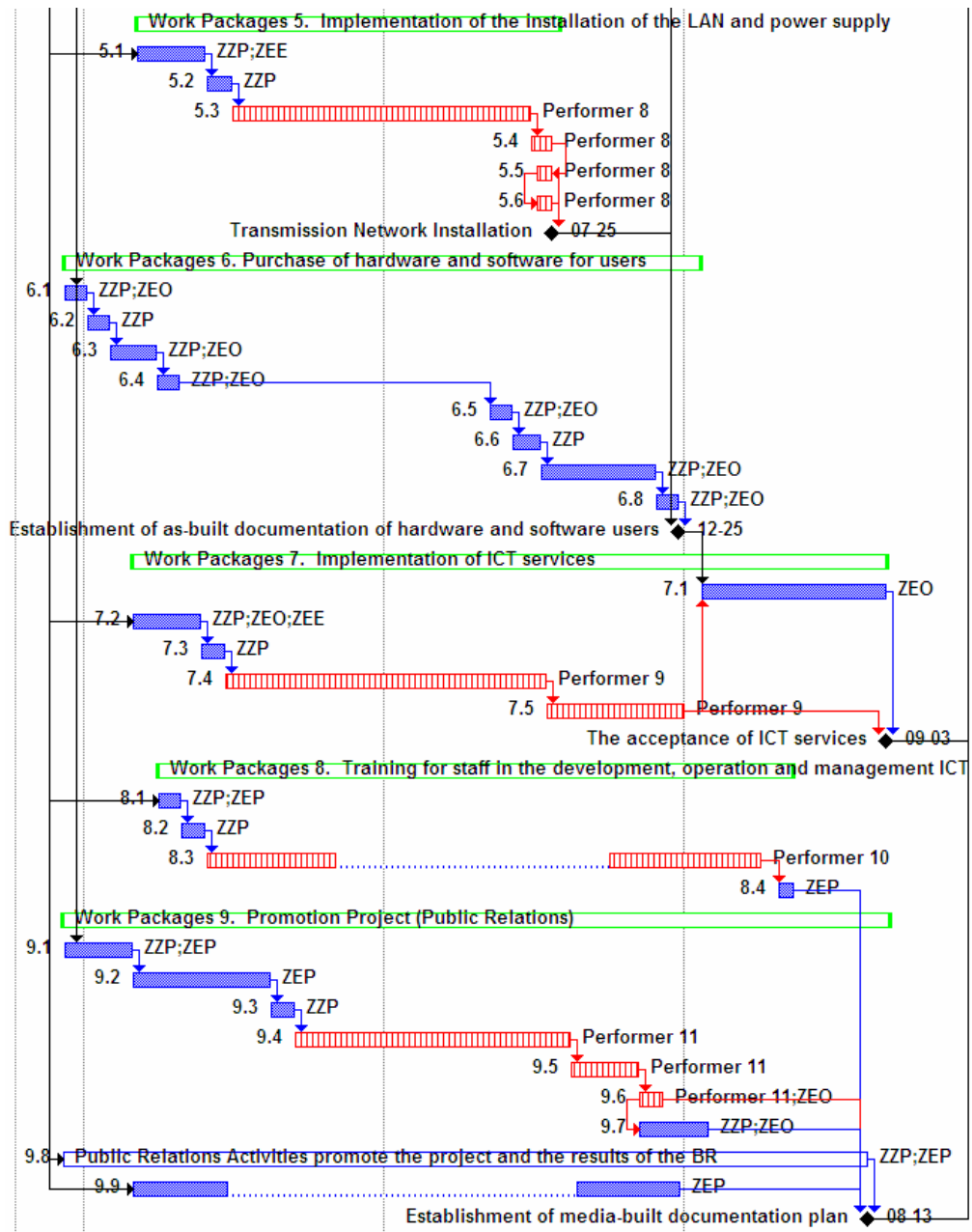


Fig. 2.4. The Gantt graph of the project realization (contd.)

2.6. SUMMARY

Because of the social and economic conditions and agricultural character of the region, the employees of the Faculty of Management find it hard to overcome obstacles of penetrating into the research communities in Poland and in the European Union. The realization of the project will not equalize the gap between a poor region and the rest of the country, but it will help the province of Lublin to make that gap smaller and, most importantly, it will allow to overcome the barrier of penetrating into the research community.

Thanks to the aforementioned infrastructure the employees of the LUT will develop their knowledge and increase professional qualifications. With the ICT services implemented, team work with the best employees in Poland will be a distinguishment of the employees of LUT to the state level. An increase in the number of research publications achieved thanks to the program and equipment resources will allow employees to match their research level with the state level.

The equipment and program resources submitted in the project should overcome the technological barrier and cause growing interest in new research problems, which could not be conducted due to the state of current resource. New directions of research and the results of their works may be published through purchased information technology equipment and necessary software.

The project assumes purchasing new equipment and program resources and placing them in newly renovated room which can be defined as a research central server room. The selected computer network should provide employees with continuous access to calculative equipment.

ICT e-infrastructure will enable creation of the platform of managing research processes in the field of balancing water resources of Central and Eastern Europe, especially:

- it will provide research employees cooperating with the higher education institution with constant and safe access to advanced information technology infrastructure
- it will allow to take advantage of services and information technology applications that support research processes

- it will give an opportunity to conduct research on the common research subject thanks to remote use of calculative powers of the e-infrastructure as well as exchange of information while preserving confidentiality, safety and interoperation
- it will allow research employees who are engaged in the project to use data repositories.

Achieving goals of the project will allow to synergize the intellectual potential of the research employees who are spatially spread among various research centers in Poland and the European Union and who work on common research subjects. Additionally, the e-infrastructure will allow the Faculty of Management to undertake new research activities.

REFERENCES

- [DRW07] Directive of the European Commonwealth Council dated 27th June 1985 *on assessment of the impact on the Operation Program Innovative Economy environment by some public and private endeavor, 2007-2013*, National Strategic Frames of Reference (Narodowe Strategiczne Ramy Odniesienia) 2007-2013 dated 1st October 2007.
- [INS10] *Manual on filling in an application form on subsidizing the realization of the project within the frames of the Operation Program Innovative Economy*. Version PO_IG/2.3/10/w01.
- [NSR09] National Strategic Frames of Reference (Narodowe Strategiczne Ramy Odniesienia) 2007-2013, *Guidelines in the field of selected issues connected with the preparation of investment projects, including projects that generate income* dated 15th January 2009.
- [NSR09a] National Strategic Frames of Reference (Narodowe Strategiczne Ramy Odniesienia) 2007-2013, *Guidelines in the field of qualification of expenses within the frames of the Operation Program Innovative Economy*, 2007-2013, 9th April 2009.

-
- [PZP07] Statue dated 29th January 2004 on *Public Order Law* (Dz.U. [2007], No. 223, item 1655)
- [REG10] The rules on conducting a contest within the frames of the Operation Program Innovative Economy 2007-2013, Priority II "B+R Sphere Infrastructure" Activity 2.3: *Investments connected with the development of the information technology research infrastructure*, 26th February 2010.
- [RRE06] The Regulations of the European Commonwealth Council, No. 1083/2006, 11th July 2006, item 16. *Equality of women and men and non-discrimination*. General regulations on European Regional Development Fund, European Social Fund and Coherence Fund repealing regulation of the European Commonwealth Council No. 1260/1999.
- [SZO09] *The detailed description of the priorities of the Operation Program Innovative Economy*, 2007-2013, National Strategic Frames of Reference (Narodowe Strategiczne Ramy Odniesienia) for the years 2007-2013, 30th December 2009.

Universal Structural Model of Standard Economic Cluster

Nowadays so-called economic clustering is a widely adopted method in economic researches. This method is based on conception about economic clusters which was introduced by Michael Porter [POR90, POR98] and defined as 'geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions (e.g., universities, standards agencies, trade associations) in a particular field that compete but also cooperate' that inevitably creates one interaction process. M.Porter came to this conception after he did competitive ability analysis of countries' economic sectors in the world market. He proved that different interrelated and interconnected groups of companies, which close to each other geographically and do not compete sever, can represent a strong competitive power in the global economy. Clustering method for analysis of economic issues develops and expands. New and more accurate definitions are constructed. Conception of economic clusters now used both in regional economy and in federal economy. For example, in our country more attention is drawn to regional educational clusters, innovative clusters and etc.

Often as an example in literature you can find structures of the following economic clusters: the California wine cluster [POR98], which's structure is illustrated in Fig. 3.1, the forest products cluster in Sweden (Fig. 3.2), the Cambridge bloodstock cluster (Fig. 3.3) and etc. Here components of the clusters, geographical features,

competitive advantages, and reasons why they cooperate with each other in one cluster were well explained. More particularly in books you will find only descriptions of economic clusters, which are researched by classic economical tools and, however, do not make a dynamic economic-mathematical modeling of these clusters possible.

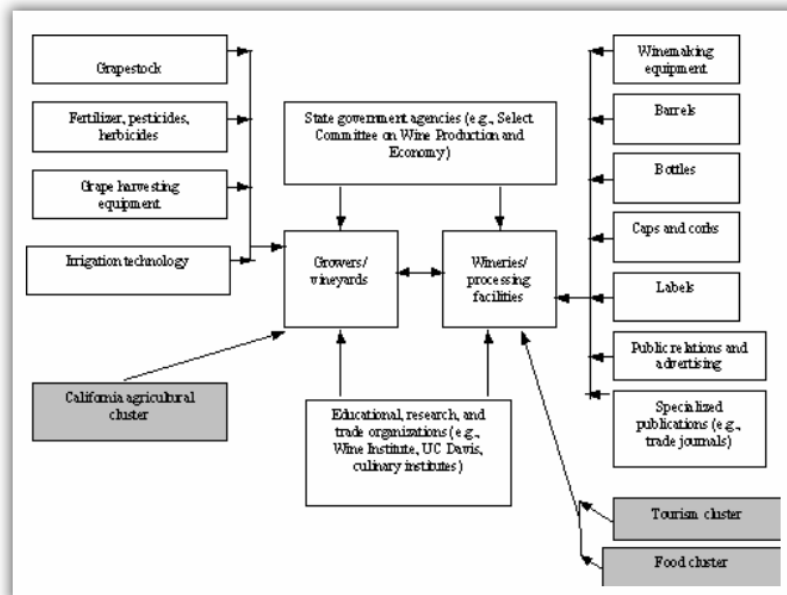


Fig. 3.1. The California Wine Cluster

Therefore, we dedicated our work to contrive a universal structural model of standard economic cluster. Our work relies on the exploration of fundamental attributes and distinctive feature of economic clusters. This work was done in Tupolev Kazan State University's Research Center and based on methods of dynamic economic-mathematical modeling of economic entities [SIR05, SIR10, SIR96].

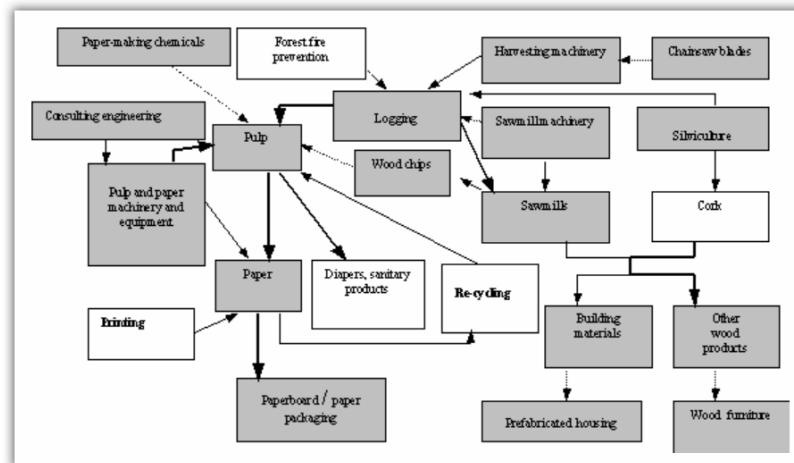


Fig. 3.2. The Forest Cluster in Sweden

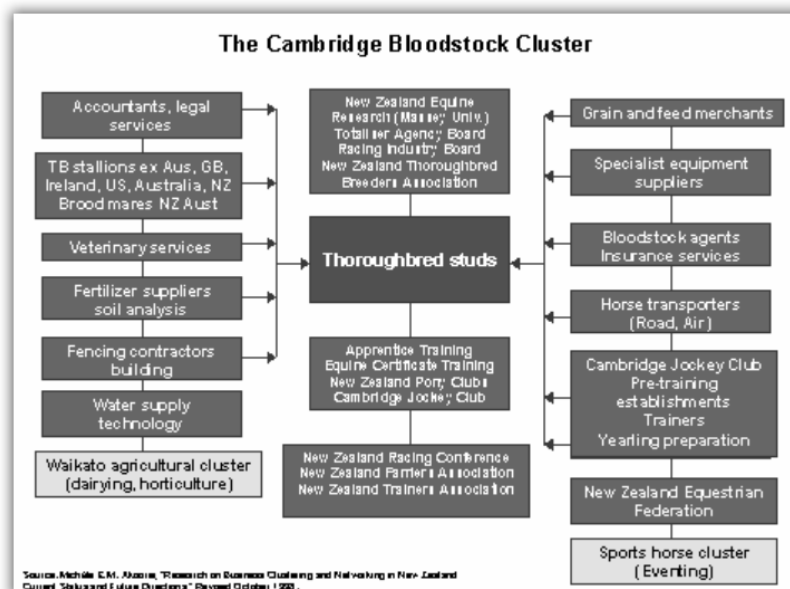


Fig. 3.3. The Cambridge Bloodstock Cluster

3.1. MAIN PROCESSES IN THE UNIVERSAL STRUCTURAL MODEL OF STANDARD ECONOMIC CLUSTER

General view of suggested structural model of standard economic cluster is shown in Fig. 3.4. The development of the structural model relies on so-called process approach where primary importance is assigned not to functional unit or an object or any subsystem but to a process that flows in the system itself.

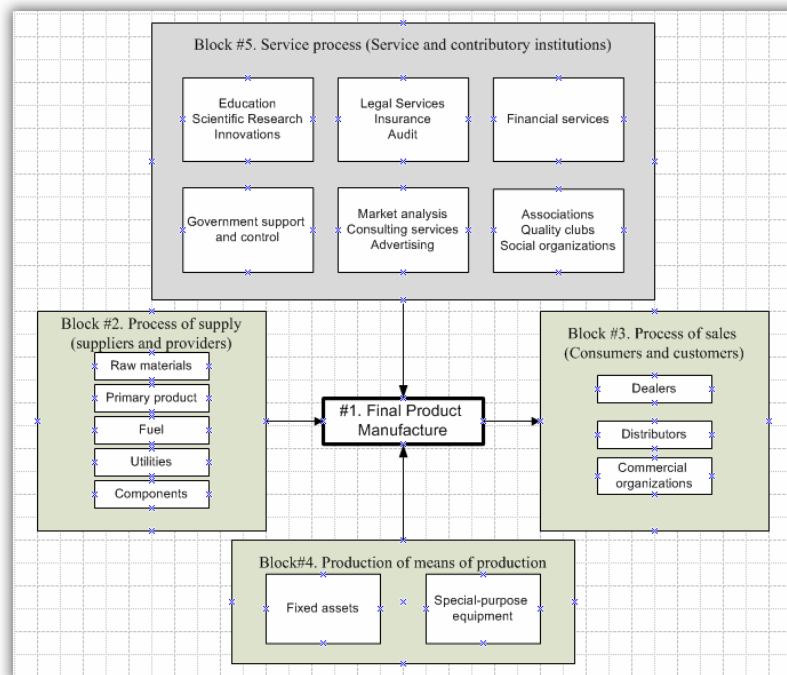


Fig. 3.4. Universal structural model of standard economic cluster

Geographic proximity of cluster's components is one of the distinctive features of a standard economic cluster. This proximity determines by the distance of data exchange, commodities exchange, and other transactions between companies within the cluster. Therefore, let us examine an economic cluster within the bound of some region, i.e. territorial district.

Fundamental attribute of any economic cluster we take its orientation on some specified final product. Therefore, a central block in the structural model of some economic cluster becomes a process of its final product manufacture. For example, the California wine cluster's final product manufacture is a wine production; for the Cambridge bloodstock cluster it is bloodstock breeding. By the way, in forest cluster in Sweden (Fig. 3.2) is not focused on one final product manufacture but it involves several wood productions such as carving wood manufacture, cellulose manufacture, and paper manufacture. Therefore, a structural model of the whole forest cluster in Sweden consists of three interconnected universal structural models in coordination with these three manufacture processes.

Thus the fundamental process of any economic cluster is its manufacture of some specified final product (Block#1, Fig. 3.4). Mechanisms of this block are all companies that manufacture the final product in the region under review. For example, for the California wine cluster there are all companies that produce wine in the geographic region under review: for the Cambridge bloodstock cluster there are all stud farm in the region.

The next distinctive feature of the economic cluster is an existence of the production chain, which provides stability of the cluster as a system. In our work we suggest to focus on two closest chain links only, i.e. on the one side – raw material suppliers and fuel providers and so forth,; and on the other side – consumers and customers. The process of circulating assets supply and delivery becomes an input process in the universal structural model of a standard economic cluster. Here all suppliers and providers appear as corresponding mechanisms (Block #2, Fig. 3.4). For example, in case of the California wine cluster this block includes all vineyards, which supply raw materials for the wine producers that belong to the wine cluster.

The final product is an output of the cluster's main process and it comes in the market for further realization. Therefore, we include the process of sales in to the universal structural model as an output process with respect to the main process of final product manufacture (Block #3, Fig. 3.4). All companies/customers that consume the final product as well as all commercial units that organize the distribution of the final product belong to this economic cluster.

The quality of means of production impact on the competitiveness of the product a lot. The quality of means of production depends on availability of modern equipment, quick re-equipment, and updating. The production of special-purpose equipment for the concrete final product plays the most important role. Therefore, the process of production of means of production is also included to the universal structural model of standard economic cluster where attention should be paid to the production of the special-purpose equipment (Block #4, Fig. 3.4). For example, at the wine cluster it is the manufacture of special wine barrels. Here at first sight it seems that it would be enough to consider the delivery only and not to consider the production, i.e. to stop at raw material market description. However, the purpose of the cluster's structural modeling is an in-depth study of cluster's competitiveness, its sustainability, and its economic security in the region. It can be achieved by stability, sustainability, and security of the suppliers as well. Therefore, into our universal structural model we include direct producers of products and materials needed for the final product manufacturing process.

One of the fundamental attributes of the economic cluster is an availability of companies, institutions, and organizations, which support and maintain the main cluster's process that is the final product manufacture. So here we include educational institutions and universities that provide skilled personnel for the cluster; we also include research centers that invent technological innovations. Furthermore, legal services and consulting companies also belong to this group as well as different social organizations and associations that may impact on the main process of the economic cluster including their impact on product's quality or its popularity on the market. The government institutions are also included to this block because they both support and control the regulation of the main process of the cluster.

Thus we add a block of service processes (service and contributory institutions) into the universal structural model of standard economic cluster, which mechanisms are companies and institutions described above (Block #5, Fig. 3.4). This block makes a distinction between a cluster and an industry as well as it makes the structural model complete as a system.

In conclusion, the universal structural model of standard economic cluster as in the figure 4 consists of the following five fundamental processes:

1. Final product manufacture
2. Process of supply
3. Process of sales
4. Production of means of production
5. Service process

3.2. ECONOMIC CLUSTER'S STRUCTURAL MODEL MAKING PROCEDURE

Generalized procedure of standard economic cluster's universal structural model making as follows:

- 1) Choose some final product that's manufacture can be admitted as an economic cluster in a particular region. The manufacture of this product should pay an important role in regional economy.
- 2) Find out all companies in the region involved to the final product's manufacture process. Examine their production facilities, work load, and level of technological development.
- 3) Determine principal suppliers of circulating assets (raw materials, components, fuel, energy and etc.) required for final product manufacture. Examine reliability of the most important suppliers especially those who allocated in the region under research.
- 4) Determine the final product's core customers. Examine market capacity inside and outside the cluster. Estimate reliability and capacity of the customers in the proper perspective.
- 5) Determine core suppliers of means of product and of special-purpose equipment used for the cluster's final product manufacture. Estimate reliability of the main suppliers and their products' quality in the world market. Draw your attention to suppliers from abroad.

- 6) Explore and reveal those companies and firms that promote product on the market or support and service the process of the final product's manufacture.
- 7) Arrange all participants of the economic cluster in due blocks of the universal structural model of standard economic cluster. This is how we build a block scheme structural model of any interested economic cluster.

3.3. STRUCTURAL MODEL EXAMPLES OF DIFFERENT ECONOMIC CLUSTERS

In this paragraph you will find structural models of economic clusters mentioned above. They all were transformed to the universal view using the above economic cluster's structural model making procedure. So a structural model of the California wine cluster is in the Fig. 3.5 and a structural model of the Cambridge bloodstock cluster is in the Fig. 3.6.

As it is mentioned earlier the forest cluster in Sweden involves several different wood productions. Therefore, a structural model of the whole forest cluster in Sweden consists of three interconnected universal structural models such as structural model of carving wood manufacture (Fig. 3.7), structural model of cellulose manufacture (Fig. 3.8), and a structural model of paper manufacture (Fig. 3.9).

By the way, sawmills produce both carving wood and wood chips and deliver them to pulp mills. By using these wood chips the pulp mills produce cellulose and deliver it to paper manufactures. By using the cellulose the paper manufactures produce paper and deliver it as a final product for printing, paperboard, and paper packaging. Used paper comes back to the pulp mills as a paper for recycling. Thus we see how the forest cluster in Sweden actually consists of three interconnected fundamental clusters.

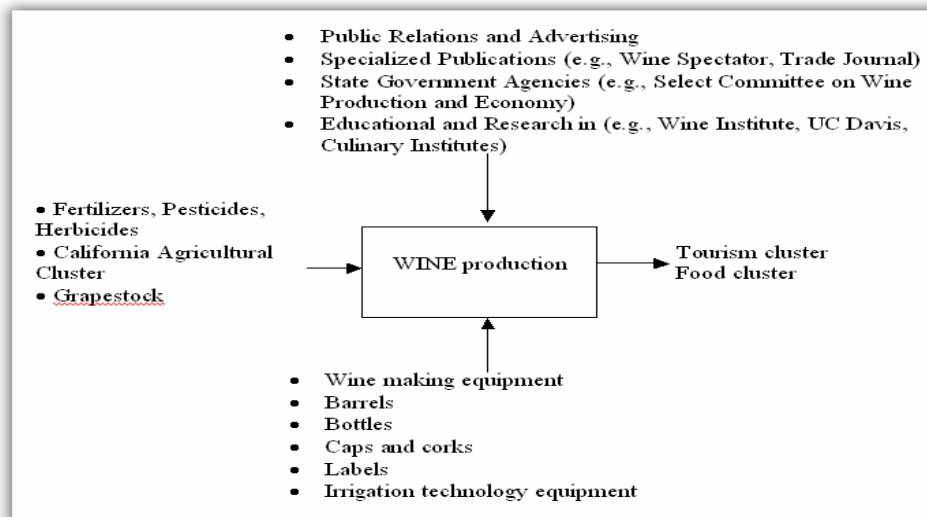


Fig. 3.5. The California Wine Cluster's Structural Model

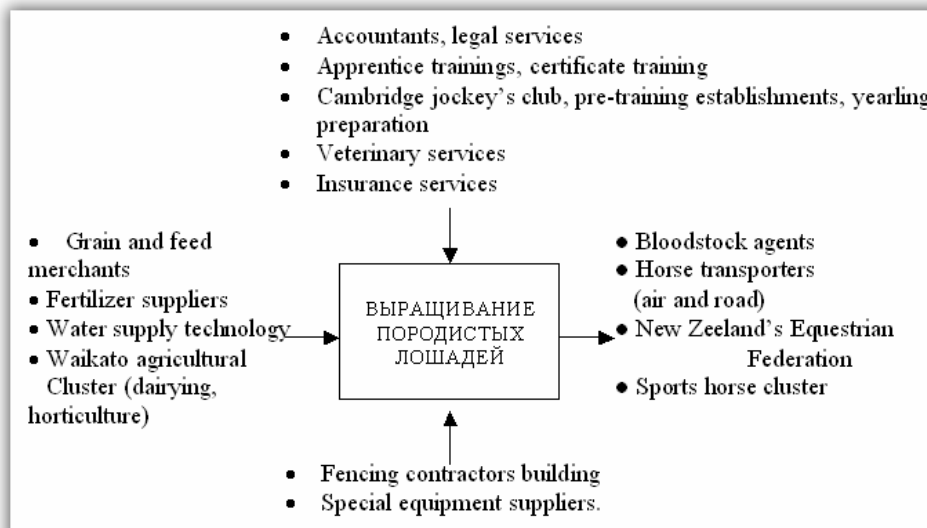


Fig. 3.6. The Cambridge Bloodstock Cluster's Structural Model

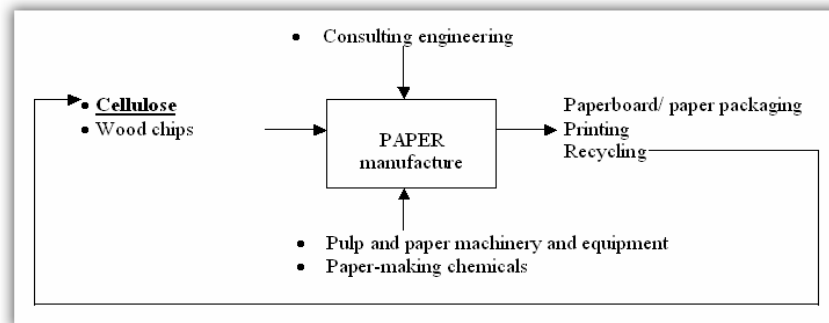


Fig. 3.7. The Paper Cluster's Structural Model

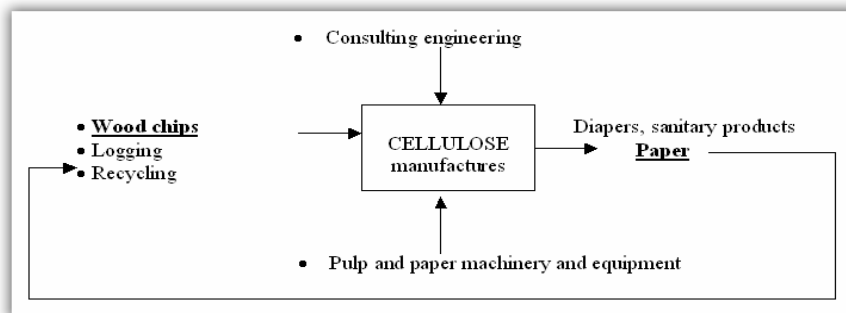


Fig. 3.8. The Cellulose Cluster's Structural Model

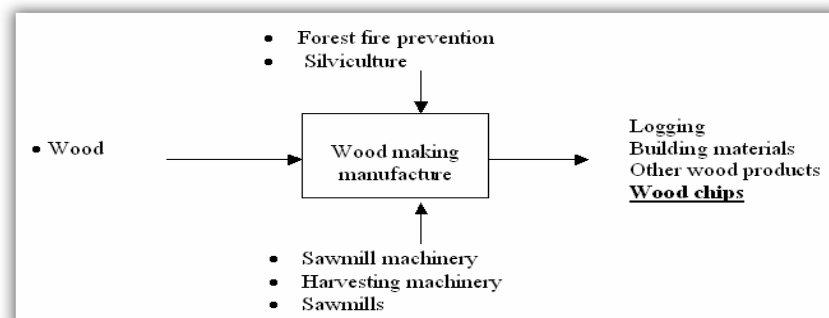


Fig. 3.9. The Wood Making Cluster's Structural Model

3.4. CONCLUSION

In conclusion, in this work we contrived the universal structural model of standard economic cluster. Also here we suggested the cluster's structural model making procedure that makes it possible to identify and research economic clusters in a particular region or in the whole country.

This approach makes it possible to consider the economy of the country (or the economy of the region) as one system of interconnected economic clusters. The universal structural model of the standard economic cluster allows solving different objectives in the region such as cluster's economic sustainability in the region, its economic development, investment allocation and so on.

The contrived universal structural model of standard economic cluster is a basis for information model making of economic clusters or of the region as a whole that could be a database with information about the cluster's companies including their interconnections and links to outside cluster partners. Using the information model dynamic mathematic-economic models could be made that in turn would allow solving different prognostic objectives, strategy planning, and economic management objective.

REFERENCES

- [POR90] Porter M.: *The competitive advantage of nations*. New York: Free Press. 1990
- [POR98] Porter M.: *On Competition*. Harvard Business School Press, Boston 1998
- [SIR05] Sirazetdinov T.K., Rodionov V.V., Sirazetdinov R.T.: *Dynamic modeling of regional economy*. "Fen", Publ. Science academy Tatarstan Republic, Kazan 2005
- [SIR10] Sirazetdinov T.K., Brahkina A.A.: *The Universal structural model of standard economic cluster – Large-scale system management. Collected articles*. Volume 29, Russian Science Academy, Moscow 2010
- [SIR96] Sirazetdinov T.K.: *Dynamic modeling of economic objects*. "Fen", Kazan 1996

Using Google Search Engine to Get Unauthorized Access to Private Data

Nowadays Google's search engine, which can be found at google.com, is considered as the best, the fastest and the most reliable free search engines by majority of Internet users [PIO05].

It also (as of July 2010) attracts over 70% of the searches, followed by Yahoo (almost 17% of searches) and Bing (almost 10%) [SEO10]. In some countries Google domination is even more evident – in Poland, for example, Google has almost 95% of web search engine market share [FIR10]. Most of the top search properties worldwide experienced significant growth in search query volume each year since 2005 [COM10].

Special software, known widely as searchbots (in case of Google called Googlebots), crawl the Web constantly, gathering information to provide better searching service.

Basically Google (like any other web search engine) was designed just to allow even the most uninitiated web surfer to find what he/she was looking for. But despite many positive sides of its features - as web, images, newsgroups, catalogue and news searches, including even language and document translation, and more – most of them offer far more nefarious possibilities to malicious Internet users (like hacker/cracker community, identity thieves or even terrorists).

How could that be possible? In order to know that, the methods of how the Googlebots work should be analyzed.

But first, it should be emphasized, that the majority of the techniques presented in this article could be successfully used on search engines different than Google. Author chose this particular search engine due to two reasons: its leading role in the market and the widest working spectrum.

4.1. HOW DO THE GOOGLEBOTS WORK?

Googlebot is defined as 'Google's web-crawling robot. It collects documents from the web to build a searchable index for the Google search engine' [BOT10].

The creators of Google came across a very complicated problem – there is no easy way to index all of the webpages. Which kind of basis should be chosen? At the time Google was developed, the majority of web search engines were based on webpage address being added manually to their databases by the its webmaster. This narrowed down the range of websites to be searched through, as many webmasters submitted their URLs just to few search engines (usually the major ones).

One of the methods to avoid such a narrowing was to follow conception of information being filtrated by Internet users themselves – let's imagine such a situation: somebody puts on his website a link, directing to another one - this may mean that he considers the other one as interesting in some kind of way and (as the result) this page should be indexed in the database of a search engine. This is the main idea behind Google web search – the Googlebots crawl the network in two types of crawl – fresh crawl (run every day, searching for new data) and main crawl (which is being run about once a month) following each link they come across (except for pages disabled for crawling by using robots.txt file, meta tag `<meta name="Googlebot" content="nofollow">` in the `<head>` section of the page or disabling a single outgoing link by referring to it using tag `<a>` with an option `rel`, e.g. ``I can't vouch for this link, better don't touch this``).

If the pages or links are not disabled in one of the ways described above, they'll

be indexed, which means they'll be reachable through a query via google.com (or any other Google domain).

Basically this situation is completely correct and acceptable - but imagine configuration, password or private data files being accessible to everyone – this is a huge hole in security system of a web server. Using only specified query attacker can gain unauthorized access to private or proprietary data – all by using a popular search engine.

4.2. BASIC GOOGLE SEARCH TECHNIQUES

Mainly to perform a quick search using Google, user should enter the search phrase into textfield on Google search main page.

But entering for example the term john smith will return not only pages referring to Mr. John Smith, but also pages containing the words john and smith (e.g. pages about actors John Wayne and Will Smith). Using the plus sign (+) (with no space following e.g. +john +smith) forces Google to perform a search for an overly common word, while using the minus sign (-) excludes a term from a search. Searching for the exact phrase requires entering it between double quotes (e.g. "john smith"). A period (.) serves as a single-character wildcard, while the asterisk (*) can act as any word – but not as a multiple-character wildcard. [FIR10]

Google also accepts Boolean logic – AND operator is automatically added between each word in a query. But if OR is used – the results will show sites containing either one word or the other, while using NOT will result in pages containing all the queried words except the one (or the ones excluded).

4.3. ADVANCED SEARCH TECHNIQUES

After being introduced to the basics, we should take a look at more advanced search options. A query can contain some predefined advanced operators, used with syntax like this: operator: search_term [LON10]. It is important to follow this syntax, especially the fact of no space between the operator, the colon and our search term.

The most popular operators are: `site:`, `filetype:`, `link:`, `cache:`, `intitle:`, `inurl:`.

The `site:` operator forces Google to narrow down the query to the specific website or domain name. This is used mainly in a situation where Google search engine is implemented on certain website.

The `filetype:` operator results in searching within the particular type of files, e.g. pdf. It is important not to include a period (.) before the file extension.

The `link:` operator instructs Google to search within hyperlinks for a search term, while `cache:` displays the version of a page as it appeared during last crawl of a Googlebot (if it is only possible) [LON10].

There are two main Google occurrence operators – `intitle:` and `inurl:`. The first operator instructs Google to perform the query within the title of the documents, while the latter instructs to search only within the URL of the document.

Of course there are more operators, just like `date:`, `safesearch:` or `related:`, but they seem to be insignificant in terms of acquiring unauthorized access to private data using web search.

4.4. FINDING PRIVATE/PROPRIETARY DATA WITH GOOGLE

Having the knowledge of using Google's operators combined with basic search techniques every Internet user can perform a specific query – either to gain access to personal/proprietary data or to find some system vulnerabilities. This technique is widely known as "Google hacking".

Google hacking is most frequently used to find private or proprietary data. For example, a user in Lublin is searching for some music by Polish singer Kayah. He's especially interested in servers located in Lublin (as he assumes the download should be at least a little faster). Web search query *kayah mp3 inurl:lublin.pl* returns three matches. One of them is just a text-based website, but the other two are fully indexed

directories of proprietary mp3s, unprotected in any way, ready to download (see Fig. 4.1 for details).



Fig. 4.1. An example of finding proprietary files using specific query (source: own work)

Google also allows people to search for keys to different proprietary software. Actually publishing cd-keys on a website may be an intentional behaviour, but still this data is considered as private (and proprietary, too). If the attacker searches for unspecified cd-key using `allinurl:cdkey.txt`, the query will result in almost 300 hits. Specifying the software name, e.g. by using `allinurl:cdkey.txt office`, will narrow down the results number. Even if the indexes websites have been removed from Internet, almost all of the data needed is shown in search result window, as it is shown in Fig. 4.2.



cdkey: FM9FY TMF7Q KCKCT [redacted]
 cdkey: FM9FY TMF7Q KCKCT V9T [redacted]
 lms.e-studies.com/soft/OFFICEXP/CDKEY.TXT - 1k - Supplemental Result -

Serial Key F8BFX-2B8M6-X373J-[redacted] K3C76 DY6WQ-D3F
 Serial Key F8BFX-2B8M6-X373J-[redacted] K3C76 DY6WQ-D3FYG-V89BY-8k
 www.personal.psu.edu/users/r/n/rnc120/Laptop%20Backup/Microsoft%20Vis

CDKEY:J2MV9--JYYQ6--JM44K--[redacted]
 CDKEY:J2MV9--JYYQ6--JM44K--QMYTH [redacted]
 61.241.81.4:8801/software/office2000/cdkey.txt - 1k - Supplemental Result -

1421-33965-2155 1977-56082-[redacted]-68725-0041 0186-'
 1421-33965-2155 1977-56082-[redacted] 68725-0041 0186-16608-7129 2344
 4336-43569-9588 3049-15217-[redacted] 84637-1932 2575-50891-6775 5555
 www.huangwei.net/public/cs-cdkey.txt - 2k - Supplemental Result - Cached

Fig. 4.2. Google's result window for allinurl:cd-key.txt search query (source: own work)

Web search engine 'hacking' (author decide to use quote marks, as it is rather a form of open source intelligence than a form of hacking) is much more serious matter than finding proprietary data on the Internet. Many websites contain some scripts written in php to manage the content easily – not a full CMS systems, but for example gallery or news script. To use it, the user must authenticate himself.

Usually users list and password are stored in config.php [LIU10]. If the script is written in a sloppy way, the configuration file can be easily indexed by searchbots, like Googlebots. If the passwords are encrypted, the potential break in the security is less possible (although they still can be cracked, e.g. by using the RAINBOW tables [OEC10]), but it does not deny the fact that such a situation is unacceptable.

Imagine the attacker wants to substitute some information on a website which uses such a script. If the aggressor queries just for *inurl:config.php*, he'll get over 51500 result, including the how-to documents, which contain words config.php in their URL, like www.karakas-online.de/EN-Book/config-php-file.html. But after modifying the query to *intitle:"Index of" config.php* – (assuming that file via indexed directory will be easier accessible) – the search returns about 15 thousand hits. Some of them are not valid, in some cases the PHP code won't return any HTML code on screen, but in

some cases (e.g. server misconfiguration) after filtering and downloading the attacker can gain access to confident data like this:

```
nobody@void ~$ cat config.php
[...]
$dbhost = 'localhost'; // database server
$dbuser = 'michael'; // db username
$dbpass = 'nimda'; // db password
$dbname = 'datenator'; // db name
[...]
```

According to information posted on [JOH10] some forums have vulnerabilities that would allow an attacker not only to read config.php, but, for instance, some versions of one of the most popular [JOH10,PIO05] Open Source web forum engine - phpBB - have a vulnerability that allow an attacker to run Linux commands on the server hosting the web site. All the attacker needs is login and password, which can be gained in a way described above.

Some users block access to their websites by using .htpasswd files. This is a very good way to disabling unauthorized access, as these files are executed server-side, but under some circumstances (especially non-standard filename and wrong set of file permissions) they can be indexed and accessible via Google. Moreover, it should be noticed that Google itself introduced a special file type called htpasswd. This illustrates the scale of the problem.

Imagine the situation when the attacker queries for filetype:htpasswd htpasswd. Google returns over 1500 results, with logins and encrypted passwords (which in some cases can be decrypted [LIU10]), e.g.:

```
dave:1eAziU9yTB.16
udel:f2JnN2GXS1CXI
saic:gRLfE/bakYDyQ
coventor:.4F5PcUhEK.n.
arl:/wp29T4JXojEE
pitt:h.Da7bCn8hcVs
```

Another method of gaining access to .htpasswd files is to perform a query on

intitle:"index of" ".htpasswd" "htgroup" -intitle:"dist" -apache -htpasswd.c . This search returns with significantly less number of hits (around 60), often resulting in error 403 (Forbidden. You don't have permission to access /.htpasswd on this server). As it can be noticed, this method is less successful then previous one.

Searching for password.log files is also a popular method of gaining access to private data. The most interesting thing is the fact, those passwords are stored as a clear text. Searching Google for *filetype:log inurl:"password.log"* results in over 600 possible hits. Typical password.log file contains three pieces of information: login, password and URL, where the user should log in, e.g.

```
name: = "test";  
password: = "test";  
URL: = "passdemol.html";
```

The aggressor can easily use this information to gain unauthorized access to the website.

Google can also help attacker to steal identity of another Internet user. Many people publish a number of documents online, which include some very private data (as Social Security Number, address, ID card number, etc.). By using this data attacker is able, for example, to apply for a credit card in some countries. All has to be done is to search for *"phone * * *" "address *" "e-mail" intitle:"curriculum vitae"* (almost 41 thousand hits), or more specified, *"social security number" "phone * * *" "address *" "e-mail *" intitle:"curriculum vitae"*, which returns only about 150 hits. In just a couple of minutes the attacker can gain access to very private and confident data.

Google can also be used to build a spam list very quickly. For example, some people export their address boxes as .csv files and make them accessible from Internet. Searching for *e-mail address filetype:csv csv* results in over 600 hits and thousands of e-mail addresses and often more personal information, like phone numbers and postal addresses.

Nowadays, in business, some documents are crucial and shall not be accessible to anyone except for the interested parties. By searching for *"not for distribution" confidential*, the attacker instantly may be given access to over 12 thousand of confidential documents, which shouldn't be distributed worldwide.

4.5. FINDING SYSTEM VULNERABILITIES WITH GOOGLE

Google hacking means the attacker being able to find system vulnerabilities. The most popular method is to specify the exact version of the web server software running on a server. It is information the aggressor needs before launching a successful attack against that server – many older versions of web server software had a number of vulnerabilities detected.

Analyzing HTTP headers that are sent by server while user connects to it can provide this information. But the connection probably will be logged – and it may lead the server administrator to the attacker. However, it is possible to retrieve similar information from Google's cache without ever connecting to the target server under investigation [LON10]. All the attacker has to do is just to perform a search for *intitle:index.of server.at*. Of course, some system administrators may fake this information, or implement *security by obscurity* policy by hiding it. But in most cases server version is present and it is legitimate. If the attacker is interested in particular machine, he probably will narrow down the search by *intitle:index.of server.at site:target_site*.

Content Management Systems (CMS) are quite popular even among common Internet users. But some part of them run such a system without securing it properly – for example leaving the configuration of MySQL database with root privileges without a password. Searching for *intitle:phpMyAdmin "Welcome to phpMyAdmin ****" "running on * as root@*"* returns almost 7000 hits, most of them being still active.

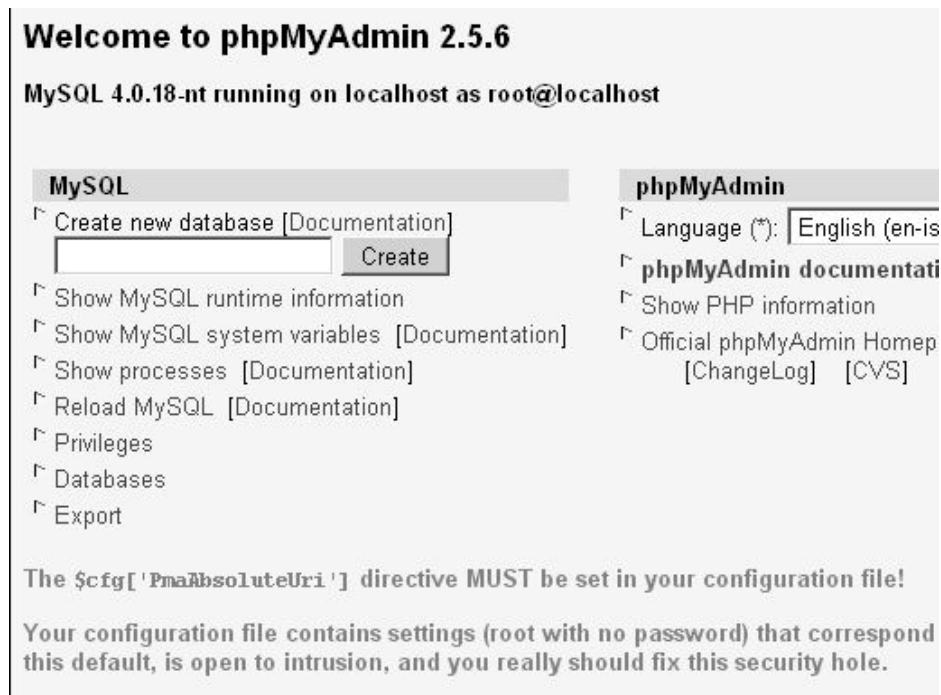


Fig. 4.3. Welcome screen of unprotected phpMyAdmin 2.5.6

(source: own work)

Fig. 4.3 and 4.4 show unauthorized modifying of database structure.

If the vulnerability of web server software is discovered, the attacker can find quickly potential victim before he's going to patch it. Searching for characteristic text strings or filenames is considered as the most popular method.



Fig. 4.4. Modifying the database structure by adding database 'hacked'
(source: own work)

4.6. BASIC METHODS OF PROTECTING DATA FROM UNAUTHORIZED ACCESS VIA WEB SEARCH ENGINES

First of all, users should not put sensitive data on the web, even temporally, especially without securing it with .htaccess file. Moreover, web server administrator should turn off listing the files in the directories in case the index file (index.html, index.php, etc.) is not present. If it is not possible (e.g. in case of having unprivileged user account on server), webmaster should place a blank index file in each directory (even those seeming insignificant) or a .htaccess file with rule redirecting the user to the main page of the website.

The second basic rule is to use tools provided by web search engines themselves, like robots.txt file or meta tag `<meta name="Googlebot" content="nofollow">` in the `<head>` section of the page or disabling a single outgoing link by referring to it using tag `<a>` with an option `rel`. It should be noticed, that both

robots.txt file and <a> tag are considered as a respected standard and the crawling software of major search engines do recognize robots.txt file and its contents. Of course, even robots.txt file can be used by attacker to gain knowledge about the structure and protected directories and files, but this issue seems to have no solution at the moment in author's opinion, as this file has to be accessible by everyone (file permissions set at least to 644).

Scanning a website for finding any vulnerabilities (for example ones described in this article) is also quite a useful method. Moreover, the source code of website should be cleared off any unnecessary links. This should result not only in fixing some of the problems described in this paper, but should also speed up the website indexing by web search bots [BOT10].

The last, and probably the most desperate method is to remove website from web search engine index. This method is, however, the most inefficient, as for fastest results it requires the webmaster to remove the website from each index manually, as any automatic method (e.g. by using specific robots.txt file) requires the website to be visited by search engine boot, which may take a relatively big amount of time. It should be also noticed, that in many cases, the data removed from the Web may be still (at least partially) accessed by web search engine cache.

4.7. SUMMARY

Unauthorized access to private data put online (also called the web search engine hacking) is nowadays considered as a moderate danger for experienced webmaster. But as it is shown by examples above, there are still many Internet users who are not aware of possible danger. The problem is especially important in cases of commercial data, which usually should be confidential.

The value of data gathered by malicious web search queries varies, but as it was proven above, the attacker may be able (in some cases) to gain access either to private data (like passwords or personal information) or confidential documents at practically no time. Securing private data put online is especially important, as in many cases the users are not aware that e.g. placing online a curriculum vitae with their

Social Security Number can be more dangerous in case of identity theft attempt than having a malicious software installed [IDT10].

Some Google queries can help attacker in exploiting the vulnerabilities of the target system, not only by detecting the software version, but also by finding bugs or backdoors in source code of the webpage or in used software. One of the most popular methods is called CSS (Cross Site Scripting).

It should be noticed, that with the popularization of this subject, the awareness of webmasters and programmers grows constantly, while number of methods of possible attack decreases. Moreover, the web search engines themselves try to reduce the number of malicious queries by blocking some of the query possibilities or removing some of the indexed files or filetypes from their databases. It should be emphasized, that the actions taken by webmasters produce their effects more quickly and efficiently and no professional webmaster should rely only on actions taken by the web search engines.

REFERENCES

- [BOT10] <http://www.google.com/webmasters/bot.html> (accessed 30.08.2010)
- [COM10] http://www.comscore.com/Press_Events/Press_Releases/2009/8/Global_Search_Market_Draws_More_than_100_Billion_Searches_per_Month (accessed 30.08.2010)
- [FIR10] <http://www.firstlevel.pl/udzial-rynku-wyszukiwarek-w-polsce/>, (in Polish, accessed 30.08.2010)
- [IDT10] http://www.idtheftcenter.org/artman2/publish/c_tips/Fact_Sheet_117_IDENTITY_THEFT_AND_THE_DECEASED_-_PREVENTION_AND_VICTIM_TIPS.shtml, (accessed 31.08.2010)
- [JOH10] <http://johnny.ihackstuff.com/index.php?module=prodreviews&func=showcontent&id=27> (accessed 30.06.2010)
- [LIU10] Liu H., Pallickara S., Fox G., *Performance of Web Services Security*, <http://grids.ucs.indiana.edu/ptliupages/publications/WSSPerf.pdf> (accessed 31.08.2010)
- [LON10] Long J.: *Google hacking mini-guide*, <http://www.informit.com/articles/article.asp?p=170880&rl=1> (accessed 30.06.2010)
- [OEC10] Oechslin, P.: *Making a Faster Cryptanalytical Time-Memory Trade-Off*, Lecture Notes in Computer Science. Santa Barbara, California, USA: Springer. ISBN 3-540-40674-3
- [PIO05] Piotrowski M.: *Niebezpieczne Google – wyszukiwanie poufnych informacji*, hakin9 3/2005 (in Polish)
- [SEO10] <http://www.seoconsultants.com/search-engines/> (accessed 29.08.2010)

Speech Technologies on the Internet

Since the Internet was created, the image has been the fundamental information carrier. However, during the last decade sound has started to play a more important role in the global network. For humans, speech is the most natural and fastest form of communication. Therefore, scientists and researchers are constantly working in order to revive communication between man and machine and to make the entire process more human-friendly. People are more likely to be willing to work with systems showing human-like qualities.

The dynamic development of speech technologies is connected with advances in research concerning synthesizers, speech recognition systems, digital telephony as well as Internet technologies. Owing to their advantages, the technologies are used in a variety of ways. Firstly, they are useful in overcoming barriers to disabilities like problems with seeing, hearing or speaking. Secondly, they are helpful in mobile systems in which reading and writing are becoming more complicated due to miniaturization of equipment. Finally, the use of the technologies is more and more popular in other areas including distance education, learning foreign languages or business.

5.1. SPEECH SYNTHESIS

Speech synthesis, also referred to as text-to-speech, is a process in which textual data are sent to the synthesizer where the text is transformed into a speech signal and

is played by means of a sound device. The result of the activities is generation of artificial speech whose quality is evaluated by means of two criteria: intelligibility and naturalness. The ultimate purpose is to create a situation in which the listener will not be able to tell the difference between synthetic and natural speech. In general, transformation of text into speech comprises two stages: the analysis of the text and the synthesis of sound wave.

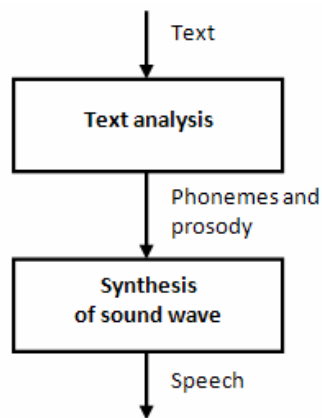


Fig. 5.1. Block diagram of a text-to-speech (TTS) system

The first operation is responsible for processing the natural language and aims at understanding a given text by eliciting the biggest possible amount of linguistic information from the textual version. This stage is accountable for transformation of the text into a phonetic transcription and for generation of suitable intonation and prosody. It consists of three phases: normalization, a phonetic analysis and a prosodic analysis. Normalization is mainly connected with division of a text into sentences, words and recognition of non-standard words (numbers, dates, abbreviations). A phonetic analysis is related to pronunciation dictionaries, proper names and changing graphemes into phonemes. The final stage during text processing is a prosodic analysis made in order to establish prosody and intonation. The functions mentioned above are realized by subsystems belonging to a module of the NLP

synthesizer. They include a pre-processor, a morphological analyzer, a contextual analyzer, a syntactic-prosodic parser as well as a letter to sound module [BEN07a].

DSP (Digital Signal Processing) is responsible for creating speech of the synthesizer and involves processes such as reception of phonemes and prosody from the NLP module, decoding of segments, matching prosody, concatenating segments and speech synthesis. The most important methods of sound wave synthesis include:

- a concatenative synthesis,
- a formant speech synthesis,
- a vocal tract synthesis.

A formant speech synthesis is connected with using appropriate digital filters that, on the basis of developed rules, generate sounds at frequency characteristic of suitable phones. A vocal tract synthesis is a simplified method for simulation of the vocal tract where speech is generated by means of specially prepared rules. At present the most popular and the simplest method is a concatenative synthesis. Synthesizers based on this method concatenate fragments of recorded speech i.e. phrases, words, half-syllables, triphones and diphones. First, a large database of sound segments with recordings of human speech must be prepared, then after marking and processing during the synthesis the segments are selected, modified and concatenated. Thanks to this method, it is possible to obtain intelligible and natural speech of high quality. A more advanced type of this method is a corpus synthesis. In this case, the corpus consists of many acoustic units. The important element here is also the criterion function that enables selection of combinations of acoustic units with the highest quality [BEN07b].

A concatenative synthesis has been used for building many synthesizers, the most popular of which in Poland is IVONA developed by IVO Software. It is an application comprising two modules: a linguistic one and a speech generator. The linguistic module analyzes a given text and converts it into a phonetic transcription whereas the speech generator makes it possible for the synthesizer to speak using individually selected voice in a suitable language. The first element of the system is a universal part of every program for speech synthesis. The second module, however, is connected with a specific voice and may be changed [IVO10a].

The other popular programs for speech synthesis generating utterances in the Polish language are Elan, MBROLA and FESTIVAL. Although speech synthesis nowadays represents a high level there are still problems with obtaining appropriate intonation of voice and stress patterns in specific parts of utterances. Yet, it may be possible only when the synthesizer is made to comprehend what is read. The IVO Software synthesizer uses Rapid Voice Development to minimize the problem. Thus, the synthesizer learns particular concepts itself in order to stress them in an appropriate way [IVO10b].

Speech synthesizers are tools that are more and more widely used in the global network. There are a lot of services that make it possible for their websites to be read to the user. The first example is the Onet.pl portal. Most information in the service may be listened to. It is enough to double-click a suitable icon located to the right, just under the title of the presented article (Fig. 5.2). Onet.pl uses the IVONA synthesizer.



Fig. 5.2. Speech synthesis on websites of www.Onet.pl

Another example of using a synthesizer for reading Internet sites is connected with the IDG portal. There, if you double-click the ear-shape icon, you automatically activate a special window and the Audiobot service that enables listening to the website being read (Fig. 5.3).



Fig. 5.3. Speech synthesis on websites of www.IDG.pl

The following two websites are examples that resemble text-to-speech systems. Festvox is a service equipped with a tool for generating synthetic speech in several languages. In this case, the Festival TTS synthesizer is applied. The service is easy to use as it is enough to type or paste a text into an appropriate dialogue box, select the audio file format and the type of voice (Fig. 5.4).

The last example of a website with a built-in synthesizer is PEDIAPHON that synthesizes articles in Wikipedia into the audio format. What is interesting about this example is the possibility to generate Polish speech. Unfortunately, the quality of the synthesis is on a low level; therefore, the speech is barely comprehensible (Fig. 5.5). It is caused by the fact that the synthesizer uses acoustic units characteristic of English.

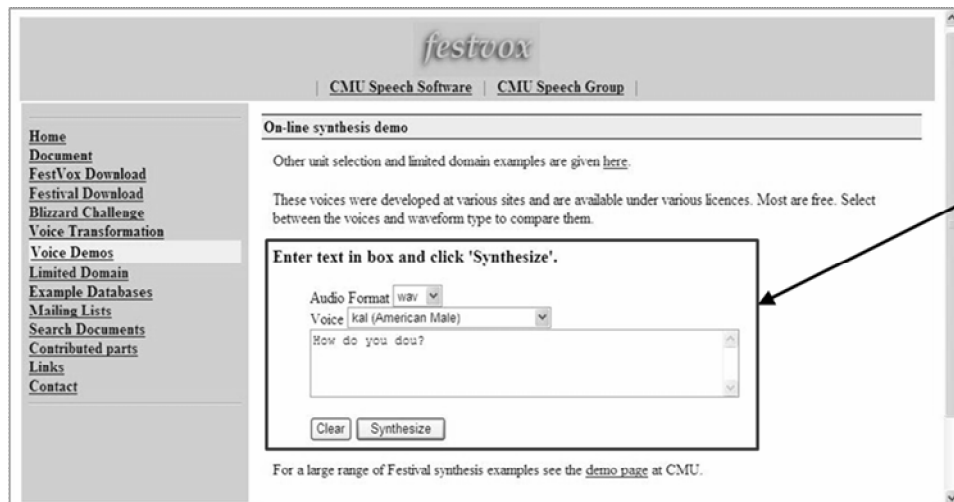


Fig. 5.4. On-line synthesis in Festvox [FES10]

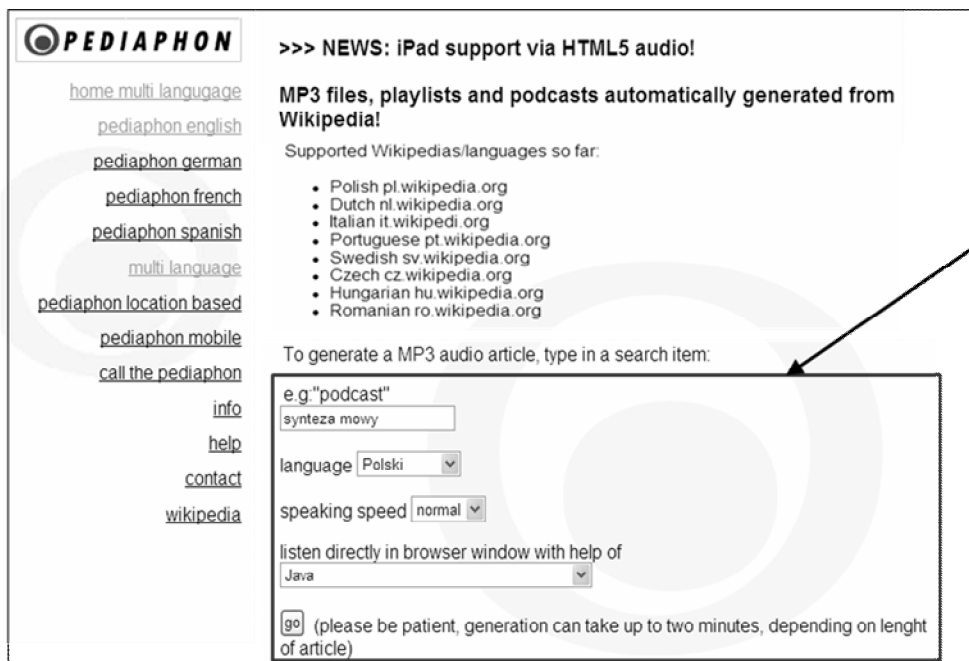


Fig. 5.5. Multi language on-line synthesis in PEDIAPHON [PED10]

5.2. SPEECH RECOGNITION

The second indispensable aspect of communication between man and machine is automatic speech recognition. The speech signal is a very complicated time sequence reflecting a complex character of its articulation. It is characterized by big redundancy and nonstationarity. The speech signal may change depending on the content, the speaking rate, coarticulation, the environment, disturbances etc. In general, there are systems for recognition of continuous speech and of individual words (isolated speech). Some of the systems require a special training session, whereas the others learn themselves. Such systems convert received speech signals into a text or react to them in an appropriate way doing specific actions.

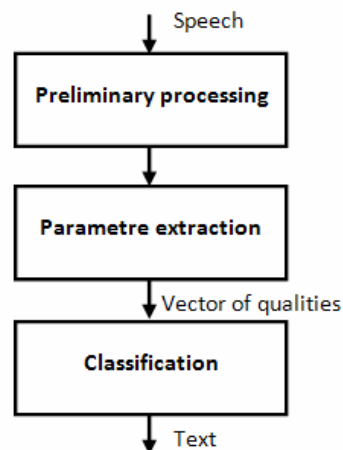


Fig. 5.6. Block diagram of a speech recognition system (ASR)

The parameter representing the quality of the process is the effectiveness of recognition that depends on:

- type of recognition (continuous speech, isolated words),
- type of text (spontaneous text, reading),
- environment (disturbances, noise),
- speaker (dialect, foreign accent).

Automatic speech recognition is a process that is usually composed of three stages. The first step is done in a computer by a sound card. This is where the preliminary processing takes place. During the processing, we observe reinforcement, filtration limiting the bandwidth from the bottom and from the top and processing the speech signal into a digital form. During the second stage, numerous values of parameters are extracted from the signal and the collection of the values creates a vector of qualities. The parameters contain information about the content of the utterance and about individual properties of the speaker's voice. At this stage, the following operations on the signal are observed: preliminary reinforcement, division into segments, the discrete Fourier transform (DFT), filtration with the use of MEL filters, logarithmization etc. During the final stage, the received vector is compared with the vector of reference in order to classify the signal. Incoming segments of images of the utterance are compared with patterns from the database. The patterns are a generalized description of classes of sounds (phonemes, words, sentences) and they are obtained during the process of learning. Speech recognition frequently uses neural networks, hidden Markov models (HMM), algorithms based on similarity functions, the Bayes algorithm, the NN algorithm (the nearest neighbour algorithm) etc [TAD88].

The potential of possible applications for ASR systems is very wide and includes controlling devices, dictating texts, preparing screenplays, simultaneous language translations and security systems using human voice.

Speech recognition is a tricky and complicated process. In addition, the environment may disturb the speech signal. Thus, in order to increase the effectiveness of recognition, speech recognition systems receive data from two channels: audio and video.

What is interesting is the fact that speech recognition is possible via the Internet. On the market there are several solutions using Internet technologies for speech recognition. For example, SkryBot.tv enables speech recognition and retrieving textual information in audio and video files. The recognition that takes place here is independent of the speaker. SkryBot is composed of a decoder and acoustic as well as linguistic models. The creators of the website aimed at automatic generation of subtitles for video recordings available in community portals and files with TV and

radio news. This solution is innovative since it pertains to the Polish language. In the future, its authors plan to extend the functionality of the service with new modules for other European languages. This portal providing audio content in a form of text may be extremely helpful for people with speech disturbances as well as for the deaf who want to learn or improve their pronunciation. Thanks to SkryBot.tv, the user may send files with recordings of utterances so that they are recognized and converted into text. The effectiveness of speech recognition by means of the presented tool exceeds 80%. Voice personalization i.e. teaching the system how to recognize a particular voice is necessary for the software in order to improve speech recognition performance. Therefore, it is important to provide a 30-60 minute recording with the user's voice. Moreover, some other additional conditions have to be met if we want to achieve this result. Firstly, it is the minimum 16kHz / 16bit mono quality of the recordings. Secondly, what is needed is the wav format and -40dB noise intensity as well as using vocabulary similar to the one that appears in TV or radio programmes [SKR10].

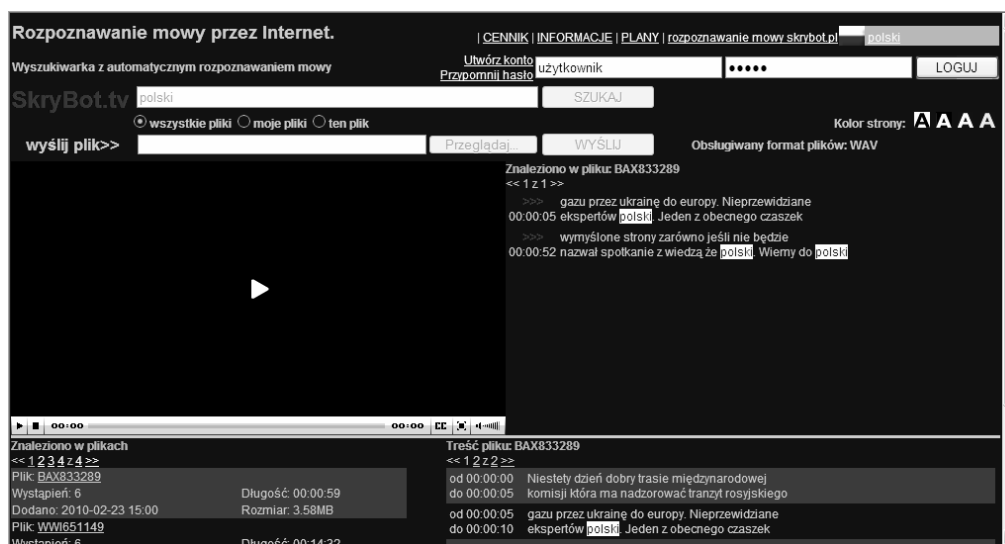


Fig. 5.7. Speech recognition on the Internet on www.SkryBot.tv

During the recent years, also the Internet tycoon- Google has worked on Google Audio Indexing i.e. the GAUDI project that aims at speech recognition and retrieving text in audio and video recordings. This service is to enable searching for text (audio mining) and creating subtitles for films on YouTube. After typing the desired text in an appropriate dialogue box, the search engine will retrieve a list of films including the uttered word or phrase. Furthermore, after selecting a given recording we will also be presented a list of sentences in which the specified word was used. In this way, the user is easily able to jump to the right portion of the video where these words are spoken. So far, this experimental technology was just used as a gadget, called “the Google Elections Video Search gadget”, only available with political channels [SHA10].

5.3. CHATBOTS AND AVATARS

Websites more and more frequently present technological solutions that are supposed to attract attention of Internet users. At present many companies place great emphasis on technologies simulating human appearance and behaviour. Speech is undoubtedly the human quality that distinguishes people from other living organisms. Since it is a fast and convenient way of communication, there are attempts to develop interfaces that exchange information between man and machine using the voice channel.

Nowadays chatterbots also called lingubots are equipped with speech technology systems. They may also be implemented on webchats. They are Internet software that is used for communication between man and machine. The programs are based on intelligent algorithms that enable to conduct a dialogue with man. It turns out, however, that in fact, the systems simulate conversations rather than conduct them. So far, there are two generations of this type of applications and the third one is only in the phase of trials and experiments. The first programs of this type enabled conversations with a computer by means of a textual language. They replaced man and acted as moderators in discussions and forums on the Internet. They analysed and supervised content provided by Internet users. They also automatically gave standard answers to frequently asked questions. The second generation of chatbots

often uses speech synthesis to conduct a conversation. During the process of communication, the service user formulates a question typing it in an appropriate dialogue box. The bot's answers are presented in a textual form in a dialogue box but at the same time, they may be uttered thanks to a specially built synthesizer. The system formulates its answers taking into account the context and the possessed database of knowledge. The conversation with a third –generation bot will be conducted in two directions – solely by means of the voice interface [PLE10a, ADA10a]. Therefore, effective algorithms for automatic speech recognition of continuous speech have to be improved. The main problems are complexity, contextuality and ambiguity of natural languages. In the case of the English language, a relatively high level of recognition was achieved. Yet, non-dominant languages including Polish are in a far worse situation.

Chatbots need to have artificial intelligence so that they could successfully imitate humans or conduct dialogues with man or other bots. It is claimed that proficiency in using a language is one of determinants of intelligence. Measurement of chatbot intelligence is possible thanks to a special Turing test. It verifies competences of machines in terms of using the natural language as well as the ability to think in a way similar to humans. So far, chatbots have not succeeded in passing the test and the simulation of a conversation is still far from being perfect. The area that developed from artificial intelligence (AI) and is responsible for modelling and simulation of human behaviour is called *botic*. This knowledge is used for developing interfaces based on NLP and rules concerning man-to-man interaction that enable more effective human computer interaction (HCI). Second-generation bots are capable of reasoning and therefore they can predict the user's behaviour at the same time adapting their actions in an appropriate way [PLE10b].

Chatbots find it very difficult to interpret the meaning of words and recognize linguistic structures since the tasks belong to the most complex aspects of artificial intelligence. Textual questions posed to programs-bots undergo the meaning analysis. Hence, analytic parsing and association procedures based on neural networks are applied. It is worth pointing out that adequate comprehension of the question guarantees giving the right answer [ADA10b]. At present scientists are developing

dialogue bots in which different methods are combined and used. Beside linguistic methods, they also apply fuzzy logic, context-free grammar, Markov processes, semantic networks, supervised learning and pattern recognition [MAW10].

Nowadays the most popular second-generation chatbots are systems that base their actions mainly on adaptation of the input sentence of the user to the answers in the database. The existing bots have three types of knowledge [CHA10]:

- common knowledge that enables the virtual assistant to conduct a conversation on general topics,
- specialist knowledge i.e. information concerning business operations of a particular company, which enables giving advice,
- outline of personality – information about the particular chatbot (age, sex, etc.).

Bots have been provided with intelligence and the ability to generate speech. Yet, recently a lot of attention has been paid to their visualization. Eye-contact is an important element of communication. Simple gestures such as a smile or nodding the head enable a faster interaction and a more relaxing conversation. Virtual characters imitate human movements, facial expression and gesticulation with the help of analysis of the text. An avatar is a graphical representation of a bot. It resembles a human head or a whole human figure. Second-generation avatars are popular and widely used on many websites. Their visualizations may be static or animated. In most cases, they are short films with virtual or real people who gesticulate and move. Selected scenes are replayed depending on the user's actions, which gives the impression of an interaction with a real person. Third-generation bots will be developed mainly in three-dimensional graphics thanks to real-time animation rendering. It will give them a more realistic, human-like appearance. In addition, the bots' lips movements will be synchronized with the sounds of the speech synthesizer [ADA10c].

Applications of avatars are very widely used in business and marketing. In business, virtual characters may be used on Internet portals, corporate vortals (Intranet), online shops, data warehouses, multimedia kiosks and electronic mail systems [ONL10a]. Internet agents can give answers to frequently asked questions

using colloquial speech. Moreover, they can collect anonymous information concerning needs of potential customers and can learn from the data they have gathered previously. Avatars require no service, work 24 hours a day, are capable of conducting a conversation with a few users at the same time and their aim is to create a friendly atmosphere and develop good contacts with the customer. They are an alternative form of advertising on the Internet. Furthermore, they may act as hosts on websites, assisting users in navigation within a service. Virtual characters frequently act as consultants, advisors or shop-assistants on Internet sites. They help customers in selecting the best offers and inform them about the operations of the company.

In Poland, a few businesses develop virtual characters and offer to locate them on websites of interested customers. For example Onlinetools.pl offers a virtual specialist as a representative of the company on the Internet whose duties include creation of a corporate image in a form of a friendly and innovative organization, answering customers' FAQs and offering them products and services [ONL10b]. Fido Intelligence developed a virtual assistant using a special Text Intelligence innovative technology. Their implemented realization is, for instance, a historical figure – virtual King Stanislaw August Poniatovsky- a bot in the service of Skarbiec Mennicy Polskiej in the Mint of Poland, whose job is to help in choosing numismatics and be prepared to give detailed information about them [CHT10, FID10].



Fig. 5.8. Chatbot representing a king in the service of the Skarbiec Mennicy Polskiej in the Mint of Poland [FID10]

Another company developing visual chatbots is Inguaris. Inguaris created by the organization is a virtual consultant who may be a tradesman, an advisor, a technical assistant, an intelligent search engine and an InfoBot [ING10]. In the Polish market, Stanusch Technologies is the leader in terms of implementations of the technology. The chatbots developed by the company have an incorporated semantics search engine. The semantics search engine requires questions instead of key words and the results listing includes a precise answer instead of a list of links [STA10]. One of examples of webbots developed by the company is Ewa – a virtual advisor who is present on biznes2biznes.com, a platform of cooperation between small and medium-sized enterprises. This virtual character has a big database with information concerning the b2b portal, acts as a guide and provides service support [BIZ10].



Fig. 5.9. Ewa – virtual advisor on b2b platform [BIZ10]

The technologies presented in the chapter may also be used in distance education. An avatar may act as a teacher in e-learning courses, may be a guide during the course, may search for information, motivate and support e-learners. In this case, however, the bot's knowledge may be limited only to the content of the training.

5.4. VOICE PORTALS

Speech technologies are also applied in voice portals that enable access to information available on the Internet and are frequently used in b2c communication. They are interactive IVR (Interactive Voice Response) systems that provide telephone access to voice websites. The basis for voice portals is speech recognition and transformation of textual information taken from a database into an audio form [FUJ10]. Thanks to them users have a fast access to electronic mail and business news (accounts, stock exchange, transactions). After connecting the voice portal, the user first listens to the introduction and then selects suitable options from the menu by dialing or by giving voice commands. Afterwards the system makes the search and retrieves information from the database using scripts located on web servers by means of SQL queries. The result of the operations is retrieved in different ways e.g. in a voice form (speech synthesis), in a fax form, computer data etc. A voice portal

enables automation of many processes, may conduct many conversations at the same time and do it 24 hours a day, which reduces corporate costs. This technology may also be very useful for blind people.

A good example of an institution that implemented a voice portal is the Polish-Japanese Institute of Information Technology. The system enables logging, selecting options from the menu, access to news (RSS services) and announcements from the dean's office, contact with workers from the Institute and the dean's office as well as access to the student account. The main element of the system is the ASR server. It has a built in acoustic adaptation module that enables retraining of the neural networks to recognize new voices. Another key element of the system is the TTS server, a multithread program with an integrated speech synthesizer. In this case, the Primespeech Voice Portal Manager is the system responsible for management of the voice portal and accessible through the Internet browser. The last element of the outlined portal is Primespeech Gateway, software cooperating with the Dialogic telephone card that enables multichannel speech recognition in real time [SZK10, BRO10].

Another example of a voice portal was developed for Zarząd Transportu Miejskiego in Warsaw (Public Transport Authority of Warsaw, a public company created to manage public transport and operators). The portal provides information on how to get from one part of the city to another, makes it possible for passengers to file complaints against drivers or ticket inspectors, gives information about tickets and charges, discount entitlements and about latest changes in timetables [ZTM10]. The advantage of the solution is above all, the instant and automatic access to information with no need for long waiting for connection with the consultant. The portal was highly rated for the quality of the speech recognition system by scientists involved in the LUNA project [PRI10, LUN10].

One more voice portal was prepared for the Warsaw Stock Exchange. It gives telephone access to latest quotations concerning different companies. In this case, it is enough to say the name of the company you are interested in and the voice portal will retrieve information related to the latest quotations. This information is

automatically updated on the Internet [GPW10]. All the three voice portals presented above were created by PrimeSpeech.

For vocal interaction voice portals use the VoiceXML language that was developed by W3C (World Wide Web Consortium). While XML (Extensible Markup Language) became a standard for representing data and structures on the web, VoiceXML became a standard for describing voice applications. The standard defines the way a dialogue is created and conducted between man and computer [MED10]. VoiceXML is a language that gives users voice access to the content of the web and is interpreted by voice browsers. It contains tags informing the voice browser what actions need to be taken. The actions may include speech synthesis, automatic recognition, conducting a dialogue and playing the voice file [UTD10, V3O10].

There is also a possibility to exchange HTML documents (websites) into VoiceXML documents (voice interface). It is feasible thanks to transcoding. However, this operation is complicated as most information on the web was created for graphical browsers. The sequential nature of voice requires a different way of presentation than the parallel nature of graphical interfaces. The graphical browser presents a general image of the website and we can move our eyes from one place to another without losing bearings. Yet, in the case of a voice interface, the users do not know what the website is like before they hear all of it. Hence, this gives graphical browsers a big advantage over a voice interface [CAF10].

5.5. SCREENREADERS AND VOICE BROWSERS

Screenreaders and voice browsers are programs that enable vocal reading of different documents e.g. HTML, PDF, DOC, TXT. The tools are especially useful for people who are blind and those with visual impairment who have difficulties in reading and learning. Furthermore, the tools are helpful for understanding texts if the people have not fully mastered the language of a particular country. They are an invaluable help in e-learning or m-learning, especially for the disabled.

Screenreaders are programs that enable presentation of content on the monitor, including websites, by means of voice. They are of vital importance to mobile

appliances in which reading is difficult due to small size of the screen. Moreover, voice access to the Internet and audio presentation of information may be very helpful during travelling e.g. while driving a car. The programs differ in terms of their reliability and intelligence. The best screenreaders enable users operating most computer programs in different computer systems without looking. In general, screenreaders find it difficult to read websites. Tables, graphics as well as links that are spelt are particularly hard to read. Expressivo is an example of a screenreader that handles reading websites relatively well. In addition, the program makes it possible to listen to audiobooks, news, e-mails, articles and RSS channels. It is even possible to create an audiobook and watch a dubbed film [EXP10].

Voice browsers are programs used for voice reading of websites. They are available free of charge to the end user. However, Website owners pay a charge to enable their site to be read by means of voice. Intelligent Web Reader is a Polish voice reader developed by IVO Software. It is a stand-alone program that makes it possible to read websites using human voice in different languages. The main element of the program is an integrated speech synthesizer. The program is installed on the user's computer and therefore it does not require a fast Internet connection. Thanks to the browser, many websites were given voice access including websites of city councils, local government institutions, financial and insurance institutions, schools, universities, foundations, associations and websites for the blind. While reading the websites, the browser divides them into logical fragments i.e. menu, footnotes, paragraphs, messages. AI algorithms are applied for dividing pages and they interpret the website in the same way as regular readers. Navigation of websites is possible by means of key shortcuts. The disadvantage of the presented browser is the fact it operates just websites with a bought license [SLU10].

BrowseAloud made by Texthelp Systems and ROKTalk developed by ROK Talk Ltd are other examples of voice browsers used for websites.



Fig. 5.10. An image of the BrowseAloud toolbar [BRO10]



Fig. 5.11. An image of the ROKTalk toolbar [ROK10]

Both browsers have many differing features such as increasing text-size, highlighting text as it is read, translation of text to other languages, and changing the background color. These features can be activated from a floating toolbar, the appearance of which can be invoked by the site visitor. While BrowseAloud requires Website visitors to download and to install software, in the case of ROKTalk no download and installation are required. For people who visit a website with ROKTalk, the TTS and other text-accessibility features can be used immediately. BrowseAloud, on the other hand may be used for other purposes such as word processing [CLE10].

5.6. CONCLUSION

Sound technologies are becoming more and more important in the contemporary world as they provide many benefits to users of computers as well as of mobile devices. Their biggest challenge is assistance in overcoming barriers to disability. Thanks to them handicapped people may be active in social life when they participate in discussion groups or learn, for example programming languages. Furthermore, modern sound technologies enhance comfort and make it easier for people to use a computer or the Internet. In many cases, they are indispensable for companies since

they increase quality of business services and reduce corporate costs. It is all due to intelligent automation of customer service processes and increased efficiency of service provided to mass Internet users. Artificial intelligence and speech generated by computers contribute to humanization and personalization of contacts between man and machine. At the same time, however, speech technologies are a challenge to developers of websites and of new software.

REFERENCES

- [ADA10] Adamczyk J.: *Ewolucja internetowych botów*, Marketing w praktyce, 1/2009, <http://www.e-marketing.pl/artyk/artyk110.php>
- [BEN07] Benesty J., Sondhi M. M., Huang Y.: *Springer Handbook of Speech Processing*, Springer-Verlag New York, Inc., Secaucus, NJ, 2007
- [BIZ10] <http://www.biznes2biznes.com/>
- [BRO10] Brocki Ł., Korżinek D., Marasek K.: *Telephony Based Voice Portal for a University*, http://www.ist-luna.eu/pdf/VP_sasr.pdf
- [BRO10] http://www.browsealoud.com/page.asp?pg_id=80002&tile=UK
- [CAF10] VoiceXML Programmer's Guide, <http://cafe.bevocal.com/docs/vxml/vxml.pdf>
- [CHA10] <http://www.chatbot.pl/wiedza/>
- [CHT10] http://www.chatbots.org/virtual_assistant/stanislaw_august_poniatowski/
- [CLE10] <http://clearhelper.wordpress.com/2010/04/29/text-to-speech-services-choosing-roktalk-over-browsealoud/>
- [EXP10] <http://www.expressivo.com/pl/>
- [FES10] <http://festvox.org/voicedemos.html>
- [FID10] <http://fidointelligence.pl/>
- [FUJ10] *Technologies for Voice Portal Platform*, <http://www.fujitsu.com/downloads/MAG/vol40-1/paper07.pdf>
- [GPW10] <http://www.gpw.pl/index.asp>
- [ING10] <http://www.inguaris.pl/>
- [IVO10] <http://www.ivona.com/>
- [LUN10] <http://www.ist-luna.eu/>
- [MAW10] http://maws.ctgt.pl/frame.php?id_article=7
- [MED10] *Voice Portals White Paper*, www.medialab.sonera.fi/workspace/VoicePortals.pdf
- [ONL10] <http://onlinetools.pl/index.html#/pl/index/consultant.html>

- [PED10] http://www.pediaphon.org/~bischoff/radiopedia/index_espeak_mbrola_multilanguage.html
- [PLE10] Pleban B.: *Chatboty jako realizacja testu Turinga, a zastosowania w biznesie*, http://www.ptzp.org.pl/files/konferencje/kzz/artyk_pdf_2010/127_Pleban_B.pdf
- [PRI10] <http://www.primespeech.pl/portal-glosowy-ztm.html>
- [ROK10] <http://www.roktalk.co.uk/index.html>
- [SHA10] Shankland S.: *Google audio search graduates to Labs project* http://news.cnet.com/8301-1023_3-10042536-93.html
- [SKR10] <http://skrybot.tv/>
- [SLU10] Ślusarczyk Cz.: *Dostępność stron internetowych dla osób niepełnosprawnych - problemy i trudności techniczne*, E-mentor nr 2 (9) / 2005, <http://www.e-mentor.eud.kei.pl/artykul/index/numer/9/id/134>
- [STA10] <http://www.stanusch.com/>
- [SZK10] Szklanny K.: *Synteza mowy w e-learningu dla osób niepełnosprawnych*, <http://www.syntezamowy.pjwstk.edu.pl/publikacje/artykulvu.pdf>
- [TAD88] Tadeusiewicz R.: *Sygnał Mowy*, WKiŁ, Warszawa, 1988
- [UTD10] <http://www.utdallas.edu/~gupta/narayanthesis.pdf>
- [V3O10] *Voice Extensible Markup Language (VoiceXML) 2.1*, <http://www.w3.org/TR/voicexml21/>
- [ZTM10] <http://www.ztm.waw.pl/index.php?c=126&l=2>

ITC Applied in Management of Stables - Common Practice and the Future

The title sounds a little bit strange. The chapter should probably appear in an agricultural business or mechanization magazine, but not at a conference on IT in management. This is in fact a problem of management. A yard requires management procedures [AGR10]. A yard is only a kind of description of business which consists of breeding horses, training them and their riders, selling if they are ready and/or purchasing if required. This kind of activity has been present for thousands of years, although over the last decades the purpose of horse breeding has changed. Horses are mostly used in sport and leisure riding (recreation) rather than being agricultural and industrial pulling power or military service. Horses are however still used for butchery in some countries.

One may still encounter opinions (I personally came across such a statement) that horses are the past. Indeed, horses are not meant to replace or to compete with motor vehicles (cars, motor bicycles, quads or tractors). This is indeed the past, even if there are some places in the world where horses are still used as workforce. Horses are to give us pleasure and an active and healthy lifestyle with a good and close contact to nature. So in this case how can we connect IT and communication (electronic) systems with these animals and their handling. There are opinions that automation and IT in stables kill traditional farmers way of living. This is true, the way

of living (not only farmers) is not the same as it was several dozen years ago, and it will be changing in the future. It is very hard to avoid a modern impact on our life. We apply modern solutions to achieve lower economical cost or to increase productivity, but also to have more comfortable living conditions which is significant especially at the very early phase of investment in innovation, when the profitability is not the most important aspect or is rather difficult to obtain.

During research for this work the author has found many computer programs supporting farm management, feeding of animals, expert systems and equipment potentially useful in collecting information about a stable and its inhabitants. Most of them are not dedicated to horse breeders and stable managers. Visiting stables and talking to their owners, managers and riders (mostly in Lublin Voivodship in Poland) the author has found that it is indeed seldom that functions of stable or riding centers are supported by computers or specialized equipment. This work is not to show the real situation in implementation of IT in stabling and its management. It requires a different type research. This is only information regarding potential solutions.

Apart from the above, it is essential to remember that breeding horses creates large potential for industry. It is common to use the notion of "horse industry" which is officially used by European Union [HOR01]. The horse industry consists of various branches of industry wanting to provide for "horse" demand.

Selected industries include:

- 1) feeding industry,
- 2) pharmacy industry,
- 3) textile and leather industry,
- 4) transportation,
- 5) building and construction industry,
- 6) tourism,
- 7) education,
- 8) insurance,
- 9) medical and veterinary services.

This is of great interest to business, creates many vacancies and may be regarded as an extra impact on traditional rural development.

The number and scale of industries involved in horse activity should not surprise anybody if we consider a few top examples from the horse market. It shows that significant amounts of money may be paid for good ideas and solutions in this business.

Example number one: the most expensive stable – Woodlands Stud in Australia was bought by Sheikh Mohammed bin Rashid Al Maktoum in 2008 for \$420M [BOR10]. Example number two: a highly priced stallion – Skip Away, earned for his owners almost \$10M [COZ10]. Example number three: beautiful and expensive mare - 13 years old Kwestura has been sold for €1,125M during Pride of Poland in 2008 [KWE10].

The table 6.1 shows some values concerning size of horse industry in selected countries.

Tab. 6.1. Size of Horse Industry

Num.	Country/State	Number of horses	Numer of riders	Official job places	Market turnover (bln. Euro)
1.	Germany	1 100 000	1 700 000	300 000	2,6*
2.	United Kingdom	1 400 000	2 400 000	250 000	10,00
3.	France	500 000	1 000 000	59 000	6,6**
4.	Sweden	250 000		28 000	2,20
5.	Austria	100 000		15 000	1,26
6.	Australia	220 000			6,30
7.	Poland	330 000	130 000	20 000	
8.	California	700 000		311 000	4,10***
9.	Floryda	500 000		440 000	3,00***
10.	Kentucky	320 000		194 000	2,30***
11.	USA	9 200 000		4 600 000	39,00***

* betting turnover not included

** only betting turnover

*** amounts in USD

Source: Own table based on [ŁOJ05, AHC10, WOR07].

The above values may cause deep concern on the profitability of investment in horse business or industry. The list above does not include industries such as IT and automation, but these industries also start to cooperate with horse and stable owners.

A wish to provide best possible conditions and ensure safety for the expensive

(and mostly just loved) horse both during the breeding process and training make the owners search for modern solutions, methods and possibilities which are available through new technology. IT and communication systems are still a new, rapidly developing and modern technology. An increasing applications in other human activities make it cheaper and available even for small size farms and stables.

Most farmers – horse breeders as well as riders do not probably realize how helpful the new technology can be and to what extent it can support their operations. It may decrease both the cost and risk and allows us more time for riding. It is important to remember that maintaining continuity of data collecting is a management obligation. AgriFood Skills Australia, one of eleven Industry Skills Councils in the country, which is obliged to provide accurate industry intelligence on current and future skills requirements and training needs for the agrifood industry, describes competency standards for stud stable management (certificate IV – RTE4101A) as follows “Gather and manage information for managing the stable – 2. Documents within the organization that detail the requirements of the production program are identified and obtained. 3. Record keeping systems in place within the stud are maintained and managed to ensure sound records may be accessed at any time.” [AGR10a].

It is necessary to define what kind of stable in Poland this article relates to. A short list of activities which may be provided by such stables, which may also be called: horse riding centre, horse riding school, horse riding club, horse training centre, sport stable, horse stud, etc. Most of these describe most important aspect of a stable activity, but do not eliminate others.

The functions include:

- to breed horses,
- to run a high standard livery ,
- to feed horses,
- to keep them in good health,
- to train horses,
- to train riders,
- to support riders having horses at the stable.

It is necessary to realize that a stable/yard should not be defined as only a building to keep horses in. This is a complex of buildings, constructions and equipment which provides horses with safe living environment, enables to proliferate, be trained and supports people taking care of them and riding them.

In this case management of a stable/yard must be taken into consideration equally to all activities in planning, organizing, motivating, and controlling people, horses and equipment to keep horses in the best possible conditions and ensure their safe use by people.

The above definition of a stable with their functions and targets shows a large number of reasons demanding the implementation of modern solutions.

The reasons are reasons:

- 1) parameters of stable microclimate,
- 2) feeding horses,
- 3) identification of horses,
- 4) safety of horses in the stable (buildings, arenas, paddocks),
- 5) proliferation,
- 6) risk decreasing for a horse and a rider in riding outside of property,
- 7) marketing, accountancy, planning, communication with customers,
- 8) horse training in various riding disciplines,
- 9) safety of horses and riders during training and sport competition/shows.

6.1. PARAMETERS OF MICROCLIMATE IN A STABLE BUILDING

It is obvious that living conditions have an influence on animal's. For this reason horse breeders are obliged to ensure good living conditions for their horses. It is usually achieved by keeping stalls clean and dry. Most farmers do not measure any parameters of stable microclimate except temperature. If the stalls are clean, adequately sized and ventilated problems of chemical air composition are easily avoided. The problem appears if the outside temperature decreases much below freezing and it is necessary to close all windows and doors of the building. A well

designed stable and ventilation system should certainly be prepared for such circumstances, but in fact most are not.

A horse with a body weight about 500 kg emits to the air 120 dm³ carbon dioxide, 83 cm³ water vapor, 698 W heating energy. According to Polish standards concentration of carbon dioxide in a stable should not exceed 3.000 ppm, hydrogen sulfide gas 0,5 ppm, ammonia 20 ppm, air flow should be below 0,3 m/s, relative humidity below 80%, and temperature between 5 and 28 C degrees [ROM04, GAJ10]. It is necessary to measure the stable's parameters to be able to understand its microclimate – at least some of them. The equipment for measurement of the above parameters is accessible on the Polish market. LB710 with an extra supporting equipment was tested in stable conditions [FIE08].



Fig. 6.1. VZ7722 – indicator of temperature, humidity and concentration of carbon dioxide

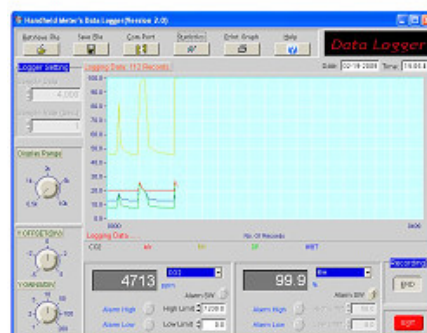


Fig. 6.2. VZ7722 software screen (seen on PC) for data analysis

A professional thermo-hygrometer might be used for continuous work and controlling external equipment (ventilators, electromagnetic valves) as well as for routine measures in the same time of carbon dioxide, temperature and humidity. User may fix alert values, which if exceed, will trigger an optic and sound alarm. The port RS 232 creates potential cooperation with PC. It allows to record current indications. Collecting data requires installation of software (on PC) which (as it happens) must be separately purchased [AIS10] It is important to have the possibility to transfer data to the PC, which may be located elsewhere, eg at a farmer's home. It is important to receive a warning if maximum parameters are exceed. This also relates to other gases which are unwelcome in a stable: ammonia and hydrogen sulfide. Usually it is possible to use one detector or measurement system for both.



Fig. 6.3. MX 2100 – Multi gas measurer

It is particularly important in stables where stalls are not cleaned daily and manure stays there for several days. The droppings and most of urine are of course covered by dry bedding but the concentration of ammonia is growing. All types of ventilation systems (gravity or mechanical) may to be controlled automatically but the problem of cost and scale of investment may be a barrier. Depending on the stable size it might be easier to control ventilation manually, but in any case it is necessary to have information about parameters of microclimate in a stable. Data to a PC or

a mobile phone of person being in charge of a stable gives them a chance to react swiftly.

The simplest is example concerning radical drop in outdoor temperature (for instance during windy winter night) causing decrease of temperature in the stable below zero. A farmer would probably find ice in waterers and cracked pipes in the morning. It is of course also not suitable environment for horses, especially if a farmer is caught by surprise and horses have not been rugged. Overheat is by far more disturbing for horses. This may occur on a winter night when a stable door is closed, and humidity, as well as concentration of carbon dioxide and ammonia rises.

Continuous access to temperature readings assisted by an alert signal (particularly important at night) would enable a farmer to avoid potential problems with horses' health and cost of water system repair. Collecting historical data for measured parameters allows to analyze them and rise horse comfort and health. Problems may appear not only during winter time but also in spring and autumn when stable windows or even door remain open. Cold wind is more dangerous to horses' health than low temperature. It is therefore important to know and react (reaction may be automatically triggered by the system) if air draft exceeds 0,3 m/s and is too strong. A portable anemometer AZ9871 [KLI10] which measures airflow, relative humidity and temperature of air is a good example. The equipment registers measured data, cooperates with PC and may be mounted on a tripod anywhere. Measuring range 0,3 – 35 m/s.



Fig. 6.4. AZ9871 – Anemometer

To choose any type of measuring equipment for a stable it is essential to remember that a stable environment is dusty and devices must be dustproof and able to operate in wider range of temperatures as well as humidity.

6.2. HORSE NUTRITION

Horse eats and drinks, it is obvious but it is not easy to calculate how much they should consume to be in good condition, but not cost more than it is necessary to spend. It is necessary to consider horses' workload (hard or light work), type of work (jumper, racer, puller, recreation horse) and conditions (pregnant mare, reproductor) while determining food dosage. Undoubtedly breed, type and weight of a horse are important as well as environmental conditions. There are various recommendations of working horse nutrition based on researches of horse's demand for carbohydrates, fat, proteins: Canadian norms, N. Hansson, Bünger, K. Nehring, G. Karlsen, M. Tomme [PRU06]. The way of calculating food doses for sport horses consisting of eight phases is described for instance by S. Pilliner and Z. Davies [PIL03]. Tables describing of horse workload trained in western style disciplines as well as racing and polo are available [DAS10].

Good advice on horse feeding may be available but most stables probably do not use them and do not calculate individual doses of nutrition for each horse. Firstly, it is difficult to qualify the horse work and their real workload. Secondly, it requires knowledge about real nutritive value of collected and used food, which also changes with time. Thirdly it takes extra time to calculate and then to prepare doses. This will change if the calculations and power expenditure evaluation are technically supported and easier.

If the kind of a horse is known, it is necessary to evaluate its energy expenditure. It is possible to use tables showing what kind of work is performed by a horse (also sports one) and what is the estimated energy expenditure. It is not possible to determine nutrition exclusively on the basis of energy expenditure calculation and it requires exact observation of an animal [PAN94]. Much more exact methods are

available e.g. *calorimetry*. This however is very difficult to use with animals and in stable conditions simply impossible. Another method is tachycardia, which is measurement of a heart rate. Measuring this parameter it is necessary to remember that it may be done only during the dynamic work (in a case of sport horse it is not a problem), but it is also necessary to remember that thermal and psychological stress have strong influence for the result. The method is not very accurate, [BUG10] but gives important information during training as well as data for analysis after certain number of trials, which is important for evaluation of horse workload and fodder doses.

At least two heart rate measurers for horses are available on the market. Polar Equine in Zone, which allows a rider to gain knowledge of how hard the horse is working during an exercise or a ride, and the Forerunner 305 with a high sensitivity GPS receiver included, new courses feature and robust wireless heart rate monitor for optimal performance.



Fig. 6.5. Horse Heart Monitor - Polar Equine In Zone



Fig. 6.6. V-Max Heart Belt & Garmin 305 Forerunner [DIS10]

It is obvious that the main application of the above equipment is in horse training and increase of the horse performance. It may also increase safety (in a case of Forerunner) during horse riding in forests by decreasing a risk of getting lost.

Supply of drinking water for horses is yet another issue. Lack of water creates metabolic disorders and is more nagging for horses than lack of fodder [FED03]. This means the breeder must be sure that horses in the stable have water to drink, especially if automatic waterers are used. To be sure a horse drinks water and how much of it, a water consumption indicator should be installed. An example of one is model ACC.WCI-40, which records the cumulative time an animal drinks from its waterer, let know if an animal is not drinking water and it can be located within 100 m of the waterer being monitored [NEL10].

Evaluation of horse power expenditure is the second step (the first one is to determine the class of a horse) to determine nutrition. Transferring records of horse workload in longer period to PC will allow ongoing analysis and future comparative analysis. It may then be used to calculate a balanced rations, which may be done “by some rather basic mathematical calculations or by a series of complex procedures to formulate least-cost rations by computer”.¹⁷ This type of software is available on the Polish market: DLG, INRAtion [INR10], Dawka, WinPasze [MRO10], Winmix (French system INRA 88) [WMX10], Hybrymin, Spartan. Excell higher advanced formulas may be used for this type of calculations [ARK10]. Most of the above mentioned software relates to cattle nutrition. Only Hybrymin software [HYB10] manual in its list of the animals contents: horses, sport horses, reproducers and mares. There were some initiatives undertaken to decrease the cost of software which is important for small farms. The result is for instance Trilpasz program, described by authors as cheaper than most of software available on the market [CZE08]. The application allows to calculate feed doses according to DLG system.

The nutritional value of fodder especially very often used in Poland for horse feeding oat and hay depends strongly on the conditions of storage and moisture of forage. Too much water in fodder causes excessive growth of microorganisms responsible for rotting components contained in the forage [LER08]. Polish standards [PNR96] allow moisture of oat up to 15%. In fact a farmer should control the condition

of oat grain while in storage as well as hay, for which maximum moisture is 17%.

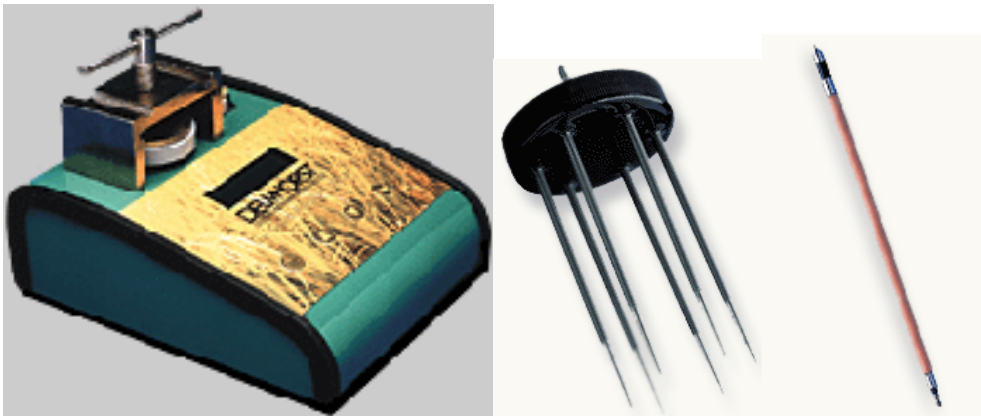


Fig. 6.7. G7, Delmhorst – Grain moisture meter and probes [DEL10]

The meter (Fig. 6.7) with accessories allows to test moisture level of various corn grains, hay and straw. Features of the equipment are: built-in temperature correction over the range of 0 - 70° C, measures moisture level over the range of 6% - 40% depending on grain, built-in calibration check, and adjustable set point with buzzer to alert when a pre-selected %MC has been reached. It is difficult to find moisture meter compatible with PC.

Finally, controlling of physical condition of a horse and its nutrition must be evaluated by owner or rider on daily basis and may be supported by weighting every two – three weeks.



Fig. 6.8. The example of equine scale (weighbridge), and hand held indicator,
<http://www.equiscales.com/>.

Monitoring horse weight gain or loss supports establishing correct feed regime.

It is easy to estimate a horse's weight using the following formula – horse girth x horse girth x horse length / 11.000 (weight in kilograms) [MCK07]. Length should be measured from the chest to the rump. The weight calculated by the formula is not accurate. There are at least several formulas useful for horse weight calculation. The usual difference is in the denominator, which it may be 8.71716 or 12.000 [PEA07]. Another formula to estimate a horse's weight – C, shows relation between horse "heart" girth – O, and horizontal horse girth – P corrected by type and/or race factor (experimentally fixed) – W [SAS87]:

$$C = O \times P \times W$$

It is advisable to use an equine specific weighbridge in order to determine accurate weight. Results may be transferred from digital scales to a PC.

6.3. THE HORSE IDENTIFICATION

There are various purposes of horse identification. Since foals have been weaned they must be distinguished each from another. It may be a problem in case of large number of horses in big studs. The transportation of horses requires documents

confirming their and their owner's identity. Participation of horses in sport competitions is also not possible without ID documents. But probably the most common reason for horse identification is horse trading. In this case the problem is more complex. There are various types of horses being bred for different purposes. A high value horse needs not only evaluation of his conformation, movement and health conditions. Since the genotype influence on the features and performance is well known the transaction of purchasing a horse requires a confirmation of its origin. General identification of a horse covers many visual features such as colour, markings (head and legs) or easy to measure such as height, girth, weight [HOR10]. State regulations usually require visual description [STA02, OKO05]. Additionally the date of birth is stated in documents, but evaluation of horse age is not so easy especially if it is aged and it happens that passports include inaccurate horse age [PAS10]. All the information used to identify a horse, including physical description, could match many other horses. To avoid this some breeds associations require an individual mark for identification purposes. Horses racing in North America is such an example with obligatory lip tattoo. This tattoo is inside the upper lip and is linked to the registration papers to identify the horse and owner [DUL10] An extra individual marking of horses seems to be necessary.

Horses are also a target for theft. The problem is that only colour, markings, age and sex which are used to identify a horse in these documents, and the physical description could match many other horses. So firstly, it is important to provide the police with a reliable method of identification. Secondly, even if the owner finds and recognizes the horse it must be proved beyond doubt that the horse has been stolen and in fact belongs to the person. ID certificates (passports), branding and even lip tattoo may not be sufficient. DNA testing which is a very good and applied method is time consuming and cannot be used by the police on the road. An electronic microchip may be a solution. "Electronic ID (EID) is considered high-tech branding and is computer compatible. A tiny microchip, laser etched with a unique alpha numeric code, is implanted in the nuchal ligament (along the crest of the neck) and read with a radio frequency scanner. The injection procedure is performed by a licensed veterinarian. The chip cannot be easily seen and does not cause any significant tissue

irritation. The information contained in the chip (a number code) is specific for an individual horse and can be obtained by using a special radio-frequency scanner. EID is difficult to remove surgically without extreme trauma to the nuchal ligament and possible death of the horse” [JOH10]. The microchip injection and reading with a scanner is only part of a solution. The data base of this specific information must be active and available for instance in case of buying a horse and checking if the horse has not been stolen. There were some initiatives undertaken in Poland a few years ago. KONID is an example of such system [MAT10].



Fig. 6.9. The microchip, the scanner, and the way of implanting it

Unfortunately the proposal was rejected. Presently the solution may be again interesting because of obligation imposed by European Union and in Poland implemented by the Act of Changing Rule about Livestock Identification and Registration System [PZJ10]. The system of horse identification might be also useful in large studs thanks to transponders, especially if staff are often changing and should limit number of frauds in horse trading. There may be some other potential applications of horse microchipping if they send signal which could be responded and located by satellites or by mobile phones network antennas. That subject will be discussed later.

6.4. THE HORSE SAFETY IN THE STABLE

It is difficult to keep an eye on horses 24 hours a day, but there are many reasons which force an owner or stable personnel to do that: disease with rapid progression, or when a horse becomes stuck in a stall after rolling. It is also important to notice quickly any damages in a stall, e.g. broken waterers, feeders or other equipment of a stall, especially if the damage may be dangerous for a horse. It also allows the stable personnel to watch mares and observe the preparation for foal birth without disturbing them. Bigger stables may employ sufficient number of personnel including night guards, but even in this case the guard must make rounds every hour or two and during intervals between rounds horses are not actually watched. Smaller stables usually rely on fate and nobody supervises horses at night and sometimes even during daytime. Another reason is to supervise stable personnel during their work. It increases level of labour quality and safety of working personnel too. The monitoring may (should) cover not only a stall, but also paddocks which would allow to avoid unpleasant consequences of broken fences and horses escapes or fighting. At least they will be promptly separated and the reason of any potential injury would be obvious. If there is no dedicated guard it is advisable to implement not only visual but also audio control which allows a farmer to hear if something is going on in the stable or on the paddocks. To organize this method of monitoring we need CCTV system –

Close-circuit Television System which may be equipped with audio subsystem, eventually with the option to record the voice. The picture, depending on the requirement may be recorded or only for viewing. The system is able to transfer the picture during daytime as well as night. Night vision CCTV simply put night vision ability to see in the dark. Night vision CCTV cameras provide high quality colour video during the day. When the light level drops below a certain point the security camera automatically switches to IR, or infrared, night vision video. When the light level increases to a particular point the camera automatically switches back to colour video [UKC10]. Cameras may be wireless or wired. They must be suitable for indoor or outdoor operation. The indoor kit should consist of a camera and dustproof cover. The outdoor housing should have an internal heater. Of course the system needs brackets, connectors, switches, patches, cables and multiplexer. The system also requires a DVR camera (for 8 or more units) with USB Backup, H.264, with the possibility of remote viewing through browser and mobile phone. UPS (Uninterruptable Power Supply) is necessary. A monitor and keyboard must be available in the guard or the watchman room and must work LCD. It is also important to project remote light switch in the stable, even if the system is working in IR. It will allow a more exact observation in case of an emergency.

A professional digital recorder of vision, working in Triplex system allows watching, recording and reading video material at the same time. A www server gives a chance to read camera pictures via internal computer network or internet.



Fig. 6.10. The keyboard and LCD monitor – pictures of stalls, outdoor camera mounted on the stable wall (watching paddocks), and the view of cameras showing paddocks and riding arena.¹

¹ The system assembled in a small stable in Lublin Voivodship, Poland (made by Atest Ltd Co.), allows permanent short and long distance supervising of horses and personnel activity.



Fig. 6.11. Pictures from nine cameras seen on the computer screen

The monitoring system cooperates with local home server and intranet network as well as with a wide range internet outlet – DSL Neostrada (2.048 Mb/s).

Audio system is supported by an electronic babysitter, which allows to place a microphone in paddocks while horses are there. A two way channel of audio communication allows hands free reporting from and ordering to the stable personnel while the owner is outside of the stable building. Of course walkie-talkies are useful if workers are out of the stable building, for instance in paddocks.

Every camera should be installed in a corner of a stall in a way that at least part of a stall where gate, feeder, waterer, salt or mineral licks and place for hay are situated are visible. It requires wide angle lens and precisely chosen place of installation, which should be also considered due to sun reflex in various daylight as well as artificial light reflex after sunset. It is obvious that all electronic equipment and cables must be high enough – beyond the reach of a horse.

Horses are not always together in paddocks. Mares and stallions are always separated as well as stallions each from another. There are some horses which do not tolerate their colleagues and also must be separated in their “private” paddocks. In the bigger stud it requires organization – logistic system. Usually stable-boy open and close right stalls and paddocks. It takes time and requires more staff. An automatic

open/close gate system might be useful in this situation. The disadvantage is that it is usually necessary to provide power supply for gate operators, but if possible the solar powered unit may be used.

Swing gate operators are the only ones to be used in paddocks. Fences are low and gates do not have frames high enough to allow horses to enter. Additionally paths for horses and fences in paddocks are often designed in such a way that a wing of a gate may open a path to one paddock and in the same time close another one.

Operators for stall gates depend on the gate construction – if they are sliding the slide operator should be used.



Fig. 6.12. Swing gate operator with solar panel and slide one [LIF10]

A well designed project will allow stable managers to program opening and closing selected gates and creating paths in paddocks for selected horses only. Of course the system works if there is power supply and must be well secured for fast manual operation in case of lack of electricity or being out of order. Control signal for gates is usually transmitted wireless. If a vinyl fence is used for paddocks it needs only light duty gate operator and solar supply should be efficient. The horses should be trained to go out of stalls if gates are opened and particularly to leave paddocks while gates are opened, if not it may need slight intervention of stable-boy.

6.5. THE HORSE AND THE RIDER SAFETY IN AND OUT OF A RIDING CENTRE

A yard for recreational horse riding is obliged to ensure safety for the riders and horses in two situations – rides outdoors or in an indoor arena as well as outside of

the riding centre. To watch the rider on the arena is easier. Even if the trainer is not there, he may use the CCTV. Well cooperating rider may inform the trainer about the horse heart beating or blood pressure parameters. It is also possible that in the riding centre there may be automatic monitoring and recording data which is more useful and important in sport riding. This would give a manager greater certainty that in case of an emergency during a ride (although accompanied by appropriately trained instructor) a fast call for help will be made by the instructor if necessary. Knowing the surrounding terrain (especially forests where the visibility is highly limited) the instructor may ask for the help. The situation may be more difficult for the group if the instructor is injured or unconscious. Usually mobile phones are helpful, but a person calling for help needs to know precisely where the group is located. Stress is not helpful and may cause mistakes. Of course there are stables providing two guardians even with a small group but it costs more. Some riders choose to ride on their own. It is rather seldom that a yard would rent a horse to an unknown rider with unknown ability to ride, but it often happens that riders use horses which do not belong to them. Some instructors take horses to train them in terrain. It is also very often that people having their own horses in horse hotels take them and go separately, even not reporting to anybody in the yard about their predicted route. In a case of an emergency it is rather difficult. First the yard manager needs to know that a problem has occurred. Sometimes the horse comes back without a rider, or time of absence is too long. Second he/she has to know where the help is needed. It means that the real time locating system would highly raise the safety level and rescue efficiency. Such systems are used for car tracking, but some of equipment is presented as very useful for many other purposes: auto enthusiasts, campers, anti-theft protection, parents of teenage drivers, couriers, municipal managers, ATV'ers, auto rental managers, snow plow drivers, firefighters, adult children of aging parents, food delivery service, law enforcement, park rangers etc.



Fig. 6.13. The USFT PT-X5: Live Personal & Vehicle Tracking Device [USF10]²

There is no reason why it cannot be used for horse riders tracking if they are outside of a riding centre. Their situation is very similar for ATV'ers. The suggestion in case of riders would be to have two emitters (trackers), one for a rider and the other for a horse in a case of a fall of a rider and an escape of the horse. The first important information for a manager tracking the couple would be separation of signals coming from two devices – the rider and the horse. The manager may contact the rider by mobile phone. It also allows the monitoring persons to watch the speed and route of the horse in real time or to play back the recorded data after return. It provides an opportunity to avoid or at least to minimize the risk of horse injury or improper use of the animal, particularly if it is rented.

6.6. HORSE TRAINING

Most of the above potential solutions relate to sport riding as well. The trainer and trainee need faster and more detailed information about a ridden horse and its organism parameters. It is important to know blood pressure and HR of a horse. The equipment for heart beat monitoring has been described in Chapter 3 in case of evaluation of horse's workload which is necessary to estimate the size of feeding

² Transmits its location every 20 feet if walking, 10 seconds if driving. Small - just 2.28"W x 3.9"L x 0.90"D. No external antenna. Internal rechargeable Li-Ion battery for up to 10 Days of 10s updates. Web-based tracking - no software to install - accessible from anywhere. Virtual Fence and Speed Alert notifications by email or text message.

doses. Watching the HR of a ridden horse allows the trainer to evaluate the horse's condition, its level of stress and effort which the horse may attempt without health risk. The best method would be to send wireless signal of monitoring devices to the trainer's computer. The limit alerts may be fixed and the trainer would be warned if exceed. Information about the HR should be recorded and completed data must be analysed after training and compared with similar data gathered over time. If the methods of touchless monitoring breathing functions [MUR04] and heart rate are developed for measurement from larger distances in free movement it would be possible to monitor these parameters both during training sessions and competitions where stress of a horse and rider is much higher. The measure may be done also during horse work in exerciser. Old type of exercisers required to tie horses to the arms of exercisers what was stressful for horses.



Fig. 6.14. The horse exerciser (four compartments) and its control panel (made by Consult S)³

Today exercisers are automatic with computer controlled time, speed and direction. Horses are not “chained” to a machine, but walk freely in a kind of stalls – compartments. If horses are not used to the equipment or are aggressive towards each other an electric fence may be additionally used on the walls separating the horse compartments

The horse condition is not only the problem, the other one is technique of riding.

³ Fully automatic. Eleven programs from relax to intensive walk for speed 4, 6, 8 km/h, change of direction in fixed time and its combination. Location – the small stable in Lublin Voivodship, Poland.

Training requires fast feedback about mistakes which gives a chance for correction. It strongly depends on the kind of competition undertaken. It is a completely different situation during for example jumping training or endurance rides. A trainer can usually observe the pair move and concludes the source of problems during jumping. But the communication between trainer and rider is difficult (short range radio with headset for both might be useful). The rider cannot see himself/herself and the horse. CCTV system in the arena (good for jumping and dressage) allows to record the training session and with more advanced systems it is possible to play it back immediately and watch the situation on a large screen. Of course it must be a multi-camera wireless system which requires placing cameras in carefully selected positions so that the recorded picture shows for instance distances of jump out points to fences, the position of horse and riders in jump etc.

6.7. THE STABLE AS A BUSINESS

The administration of a stable needs not only field activity, but also running typical managerial administration, which requires communication with clients – riders, potential clients, horse owners having their horses in the stable as well as making promotion to collect new horses to the yard and riders for training. Most yards in Lublin Voivodship providing commercial services have dedicated websites. Their role is mostly to inform potential clients about stableing opportunities: arenas, trainers, horses (also for sale) and specialization of training (recreation, dressage, jumping etc.). It is possible to find information about extra activities such as competitions and results, riding holidays and other special events. Usually it is possible to find contact data such as phone number or e-mail address and the location with access to road map. Most yards include photo galleries of their surroundings, horses, trainers and commonly organized events.

It would be useful to have an interactive booking system of rides and hotel rooms because depending on the size of horse stalls and number of riders in training as well as size and number of arenas the manager may limit time of arenas usage.

Unfortunately no stable in Lublin Voivodship offers such opportunity. The same concerns hotel rooms for riders.

Simple data bases concerning horses at yard's disposal as well as those in the hotel and their owners would be useful. Some yards collect information about horse medical history. Suppliers of fodders, manure collecting farmers, veterinarians and farriers data should also be in the system. An accountancy software is also required.

Even small yards require at least common administrative software adapted to its needs. The manager of the stable should not forget about guests coming to the stable such as parents with their riding children who sometimes are not interested in contact with real horses or are afraid of them, but also about beginners scared a little bit of riding real horse. In that case an "automatic" horse would be a useful solution encouraging potential customers.



Fig. 6.15. The horse simulator created by Racewood and the software developed by Equitech Software

The above robot may ensure lessons on various levels in dressage which gives

also professional riders a chance to develop their riding skills without tiring a horse.

“Full riding lessons can be conducted on the horse and it is good for developing riding fitness. ... The machine also informs riders about their level of accuracy including circle sizes and proximity to markers” [TYZ06].

6.8. CONCLUSION

If a rider trains only for jumping they may have some simple problems like observing their way of jumping and to correct bars/poles in fences if knocked down or changing the height. Any change of training parcours as well as during competition needs time and assisting people to carry bars. The solution would be virtual parcours controlled by computer, with a possibility to program the type, height and width of fences and their location in the arena. Laser and holography technique may be very helpful. There would be a problem of horse not feeling the impact of bars, but probably this problem could be solved with special hoof covers giving the sensation of hit.

There are ‘intelligent’ houses in the construction market supporting their inhabitants in lighting, heating, audio controlling and remote access to all functions, the construction of an intelligent stable would be a challenge for engineers and programmers.

REFERENCES

- [AGR10] AgriFood Skills Australia, AHCA50308 *Diploma in Livestock (specialising in horse breeding)*, <http://www.agrifoodskills.net.au/sub-pages/ahca50308-diploma-in-livestock/>, 28.09.2010
- [AGR10a] AgriFood Skills Australia, AHCA40308 *Certificate IV in Livestock*, <http://www.agrifoodskills.net.au/sub-pages/ahca40308-certificate-iv-in-livestock/>, 28.09.2010
- [AHC10] American Horse Council website. <http://www.horsecouncil.org/nationaleconomics.php>, 26.09.2010

- [AIS10] *Urządzenia pomiarowe, termo-higrometry.* <http://www.aisko.pl/vz7722-miernik-co2-temperatury-wilgotnosci-p-265.html?osCsid=6e7aebaf492e85f4e306110a801cccb5>, 28.09.2010
- [ARK10] *Programy komputerowe w żywieniu zwierząt, przedmiot nauczania – tematyka ćwiczeń.* AR Kraków, <https://usosweb.ar.krakow.pl/kontroler.php?action=actionx:katalog2/przedmioty/pokazPrzedmiot%28kod:H.ZZW.PKZZ9.SI.HZOHX%29>, 04.10.2010
- [BOR10] Borthwick M.: *Dubai horse racing industry defiant.* BBC Middle East Business Report, Dubai, http://news.bbc.co.uk/2/hi/in_depth/business/2009/business_of_sport/8584923.stm, 06.09.2010
- [BUG10] Bugajska J.: *Ocena obciążenia pracą fizyczną dynamiczną na stanowisku pracy.* CIOP, http://ergonomia.ioz.pwr.wroc.pl/download/obciazenie_praca_dynamiczna-teoria.pdf, 04.10.2010
- [COZ10] *Cóż to był za ogier! (What a stallion it was!).* <http://sport.wp.pl/kat,1025735,title,Coz-to-by-l-za-ogier-Zarobil-10-mln-dolarow,wid,12270110,wiadomosc.html?ticaid=1af49>, 06.09.2010
- [CZE08] Czerniawski Sz., Krzyś A.: *Wspomaganie obliczania dawek paszowych poprzez wykorzystanie programu komputerowego Trilpasz.* Inżynieria Rolnicza 5(103)/2008, Uniwersytet Przyrodniczy we Wrocławiu
- [DAS10] *Formulating Rations for Horses*, Equine Section, Department of Animal Sciences, University of Kentucky, <http://www.uky.edu/Ag/AnimalSciences/pubs/asc115.pdf>, 04.10.2010
- [DEL10] G7, *Delmhorst – miernik wilgotności ziarna.* http://www.moisturemeterstore.com/moisture_meter_m6500.shtml, and <http://centrumpomiarow.pl/mierniki-wilgotnosci/g7/>, 05.10.2010
- [DIS10] *The Distance Depot*, http://www.thedistancedepot.com/VMax_Horse_Equine_Heart_Rate_Monitors_s/67.htm, 04.10.2010
- [DUL10] Dulay C. P.: *Horse Identification Lip Tattoos.* <http://horseracing.about.com/od/statisticsdata/a/aalipattoo.htm>, 05.10.2010
- [FED03] Fedorski J.: *Poradnik dla hodowców i miłośników koni.* PWRiL, Poznań 2003

- [FIE08] Fiedorowicz G., Łochowski B.: *Mikroklimat stajni w okresie zimowym*. Problemy Inżynierii Rolniczej nr 4/2008, IBMiER Warszawa
- [GAJ10] Gajda O.: *Paragrafy w stajniach. Warunki przetrzymywania koni, , Konie i rumaki*. <http://www.konieirumaki.pl/component/content/article/73-konie-i-rumaki-sierpien-2009/139-paragrafy-w-stajniach-warunki-przetrzymywania-koni.html>, 28.09.2010
- [HOR01] *The Horse Industry in European Union*, Final Report 2001, Swedish University of Agricultural Sciences, Uppsala, 2001
- [HOR10] *4 H Horse Reference Manual, Horse Identification*, [http://www1.agric.gov.ab.ca/\\$department/deptdocs.nsf/all/4h7933](http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/4h7933), 05.10.2010
- [HYB10] *Hybrimin Futer 2003 Detale*, <http://www.hybrimin.de/content/index.php?id=28&lang=pl>, 04.10.2010
- [INR10] INRation, PrevAlim, <http://www.inration.educagri.fr/pl/presentation.php>, 04.10.2010
- [JOH10] Johnson P.: *The Importance of Horse Identification*, <http://www.petplace.com/horses/the-importance-of-horse-identification/page1.aspx>, 05.10.2010
- [KLI10] *Klimatyzacja PL, branżowy portal internetowy*, http://www.klimatyzacja.pl/index.php?akt_cms=2295&cms=174, 28.09.2010
- [KWE10] *Klacz Kwestura sprzedana za ponad milion Euro. (Mare called Kwestura sold for over one million Euro.)*. Janów Podlaski, <http://www.dziennikwschodni.pl/apps/pbcs.dll/article?AID=/20080810/AKTUALNOSCI/360650440>, 06.09.2010
- [LIF10] *Liftmaster gate operators*. <http://www.liftmaster.com/consumerweb/pages/productmodeldetail.aspx?modelId=1463>, and <http://www.liftmaster.com/consumerweb/products/pfISlide.htm>, 02.11.2010
- [LER08] Leroch R.: *Pasze w magazynie*, Farmer PL, 15.02.2008, http://www.farmer.pl/srodki-produkcji/pasze-zywienie/pasze_w_magazynie_c5e8405c21de9161396a.html, 04.10.2010

- [ŁOJ05] Łojek J.: *Koń, zwierzę z przyszłością, (Horse, the Animal with Future)*, HIPPICA PRO PATRIA, 2005
- [MAT10] Matuszczyk J.: *Bezpieczeństwo twojego konia*, <http://arch.voltahorse.pl/articles.php?id=1326>, 05.10.2010
- [MCK07] McKiernan B.: *Estimating a horse's weight*, PrimaFact 494, NSW Department of Primary Industries, Australia 2007, http://www.dpi.nsw.gov.au/_data/assets/pdf_file/0008/109988/estimating-a-horses-weight.pdf, 05.10.2010
- [MRO10] Mroczo L.: *Programy komputerowe dla rolnictwa. WinPasze 3.0.*
- [MUR04] Murthy R., Pavlidis J., Tsiamyrtzis P., *Touchless Monitoring of Breathing Functions*. University of Huston, Texas, University of Athens, Greece, Proceedings of 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, September 1-5 2004
- [NEL10] *Nelson Waterers Accessories (Automatic Horse Waterers Accessories)*, <http://www.nelsonmfg.com/horse-waterers/700/accessories/>, 05.10.2010
- [OKO05] Okoński J.: *Identyfikacja koni sporządzanie opisu i diagramu do sportowych paszportów koni zgodnie z instrukcją PZJ*. prezentacja, RACOT 2005
- [PAN94] *Normy żywienia koni. Wartość pokarmowa pasz*. Praca zbiorowa. IFiZZ PAN, Omnitech Press, Warszawa 1994
- [PAS10] *Stable Close Equine Practice*, <http://www.horsevet.co.uk/passports.php>, 05.10.2010
- [PEA07] Pearson N.: *Condition Scoring and Weight Estimation of Horses*. Agriculture Notes, Attwood, 2007
- [PIL03] Pilliner S., Davies Z.: *Jak osiągnąć mistrzowską formę koni*. SIMA WLW, Warszawa 2003
- [PNR96] Polska Norma PN-R-74106: 1996
- [PRU06] Pruski W., Grafowski J., Schuch S.: *Hodowla koni*. PWRiL, Warszawa 2006
- [PZJ10] PZJ, <http://pzj.pl/modules/news/article.php?storyid=1592>, 11.10.2010

- [ROM04] Romaniuk W., Overby T.: *Systemy utrzymania koni – poradnik*, Instytut Budownictwa. Mechanizacji i Elektryfikacji Rolnictwa (publikacja sfinansowana przez EU), Warszawa 2004
- [SAS87] Sasimowski E., Budzyński M.: *Żywnienie koni*. PWRiL, Warszawa, 1987
- [STA02] Stachurska A.: *Identyfikacja koni*. Wyd. AR Lublin, 2002
- [TYZ06] Tyzack A.: *Riding schools embrace horse robot*, 10 May, 2006, <http://www.horseandhound.co.uk/news/397/84556.html>, 08.09.2010
- [UKC10] UKCCTV, <http://www.ukcctv.co.uk/night-vision-ir-cctv-security-cameras.xhtm>, 12.10.2010
- [USF10] *The USFT PT-X5: Live Personal & Vehicle Tracker*, <http://www.vehicle-tracking-gps.com/pt-x5-personal-tracker.htm>, 02.11.2010
- [WMX10] *Winmix wersja 2.0*, <http://www.djg.pl/cdata/File/opisy/WMX2.pdf>, 04.10.2010
- [WOR07] *World horse population estimated at 58 million*, 12.09.2007, <http://www.horsetalk.co.nz/archives/2007/09/105.shtml>, 07.09.2010

Electronic Contacts between the Applicant and Office after Amendment of the Act on Informatization of Activities Carried out by the Public Task Performing Entities

Amendment of the act on informatization of the activities carried out by the public task performing entities was enacted by the Polish Sejm (the lower chamber of the Polish parliament) on 12 February 2010 and changed also the provisions of nine other legal acts. One of these is the act of 14 June 1960 – Administrative Procedures Code (Official Journal of Laws No 98, sec. 1071 with amendments of 2000, hereinafter referred to as the KPA). The KPA regulated, among others, the way of submitting, issuing and delivering the letters between applicants and offices. This legal act was an obstacle on the way to electronic administration, because it mostly applied to documents drawn up on paper. Also the tax ordinance was modified (Official Journal of laws No 137, sec. 926). None of the articles of the amended act says that official web pages should function as social portals. It is, however, the worldwide trend which is also followed by our administrative entities.

7.1. ELECTRONIC CONTACTS IN ADMINISTRATIVE PROCEEDINGS IN VIEW OF AMENDMENTS IN THE KPA AND UINF

According to article 14 § 1 KPA from before the amendment, the cases and affairs should be dealt with in writing. When amended, it admits an alternative. Specifically, this alternative says that affairs can be dealt with in writing, as has been the case so

far, or in the form of an electronic document pursuant to the provisions of the act on informatization of the entities which perform public tasks, of 17 February 2010 (hereinafter referred to as UINF), delivered by electronic communication means. Thus, affairs can be dealt with in the form of an electronic document delivered by electronic communication means. Other provisions of the KPA say that the electronic communication means may only be the electronic in-box in the ePUAP system or another one. It seems that creating an entry about delivery through an electronic in-box, instead of using electronic communication means, would contribute to a higher coherence of regulations.

Paragraph 4 added to article 46 KPA says that in order to deliver a document in the form of an electronic document, the public administration organ (subject to § 6) sends to the addressee's electronic address the information containing two indications and one instruction. According to the first indication, the addressee may collect the document in the form of an electronic document. The second indication is an indication of the electronic address from which the addressee may collect the document and under which she/he should confirm the service. The instruction refers to the way the document is received, and first of all the way of identifying under the indicated electronic address in the teleinformatic system of the public administration organ and information about the requirement to sign the official certification of receipt as specified in art. 20a UINF.

The public administration organ, in order to deliver the electronic document, should first send to the addressee's electronic address the information that she/he may collect the document and how to do this. In turn the addressee may subsequently collect the document under the electronic address indicated in the information and she/he should under that address confirm its delivery. Associated with the receipt of the electronic document is the requirement to sign the official certification of receipt through a safe electronic signature, verified by a qualified certificate, or by means of a signature confirmed by the trusted profile ePUAP, as well as by means of other technologies enabling identification in the teleinformatic system.

Article 46 KPA § 3 deals with the effectiveness of the delivery of a letter by electronic communication means. If within 7 days from sending the letter the public

administration organ receives a confirmation of the letter delivery, the delivery is effective. Otherwise, if no confirmation of the delivery is received within 7 days from the day the letter was dispatched, the office delivers the letter in a traditional way determined for the letter in a form different from the electronic document form.

There is a discrepancy in the KPA between articles 46 § 3 and 44 § 4. As article 44 § 4 deals with delivering a „document in the form of an electronic document“, article 46 § 3 deals with the effectiveness of delivering the “letter” through electronic communication means. The sense is saved by article 43 which says that the „document in the form of an electronic document” from article 44 § 4 is identical with the “letter” from art. 44 § 3 KPA.

Another serious problem results from the fact that while delivering the letter through electronic communication means NOBODY sends the letter and a term appears which does not start the course of the matter. If the letter is not dispatched specifically, one has no concrete date from which to count down 7 days. Therefore it may result from articles 46 § 4 and 46 § 3 that in the situation when the delivery of a letter through electronical communication fails, the letter will never be delivered electronically. But such a conclusion should not be applied in practice. Lawyers advise to count down the mentioned 7 days, when regulations do not specify the moment of sending the letter through electronic communication means, by taking a decision in the office. It can be e.g. counting down 7 days from the moment of sending to the addressee’s electronic address the information (sending the information is not tantamount to sending a letter, but at least it is the first activity about which the addressee can learn) about the awaiting letter and how it will be collected or from the moment of placing the letter at a place from where the addressee can collect it. What can happen in the case the counting down starts from the moment the letter is placed at the place from which it can be collected? If the organ does not send to the addressee an email with the information about the awaiting letter or when the teleinformatic system does not automatically generate such an email, the addressee does not learn that a letter is awaiting him and will not collect it. If the addressee does not collect the letter, the office will not receive a confirmation of receipt and may deliver it in a traditional way. What kind of situations may arise when the office counts

down 7 days from sending the information about the waiting letter? If the email with the information about the waiting letter, whether sent by the organ or automatically generated by the system, reaches the addressee but the addressee does not collect it, then the office does not obtain the service confirmation either and will be entitled to deliver the letter on paper.

Another problem is the lack of the delivery object. In § 7.3 the decree of the Minister of the Interior and Administration of 27 November 2006 on drawing up and delivering letters in the form of electronic documents (Official Journal of Laws No 227, sec. 1664) says that a copy of electronic document may be made in the form of a certified computer print. Legal regulations forbid the delivery of electronic documents on an informatic data carrier. We have to get used to the possibility of delivering the paper copies of electronic documents.

The serious problem with establishing the dates of letters delivery results from copying largely art. 46 § 4 KPA from § 4.1 of the mentioned decree of the Minister of the Interior and Administration. Amending UINF, after five years, led to wasting the chance to adapt this article to art. 46 § 3 KPA or vice versa.

The amended art. 391 KPA says that: „The delivery is carried out by electronic communication means pursuant to the provisions of the act of 18 July 2002 about electronic providing of services (Official Journal of Laws No 144, sec. 1204, with subsequent amendments), if the party or another participant of the proceedings: 1) applied to a public administration organ for the delivery or 2) expressed her/his consent for delivering her/him letters by these means.” So it indicates that not the organ but the party or another participant decide about the delivery way and thereby the form, whether written or electronical, of the delivered document.

The modified art. 63 § 1 KPA contains the catalogue of the methods of submitting applications. According to that article, they may be submitted: „in writing, telegraphically, by telefax or verbally to the protocol, and also using other electronic communication means through the electronic in-box of the public administration organ... “.Of the allowable methods of submitting applications, removed was submitting them by teleprinter and ELECTRONIC MAIL. One could hardly be surprised by disappearance of the teleprinter from the regulation because of

a sporadic use of this oldest form of submitting applications. Electronic mail may still be used in offices e.g. to collect applications for access to public information.

The amended version of art. 107 KPA indicates that the decision may also be issued in the form of an electronic document affixed with a safe electronic signature verified by a valid qualified certificate. Consistent with the contents of this article is the amended art. 109 § 1 KPA, according to which the decision is delivered to the parties either in writing or through electronic communication means, of course through the electronic in-box. Problems with delivering the decision may also occur in this case, similarly to those described for delivering letters. If, despite efforts made, the public administration organ fails to deliver the decision issued in the form of an electronic document, by electronic communication means, it can do this in the traditional way.

A decision may be issued, by virtue of the amended art. 124 KPA, in the form of an electronic document affixed with a safe electronic signature verified by a valid qualified certificate and the certificate issued in this way may be delivered in an electronic form. On the other hand, the amended art. 125 KPA says that the „Decisions which can be complained against at the administrative court are delivered in writing or through electronic communication means.”

The certificates may be issued in the form of an electronic document affixed with a safe electronic signature, if so demanded by a person applying for the certificate (art. 217 § 4 KPA). Such certificates may be delivered both by the electronic in-box or on an informatic data carrier.

The public administration organs were imposed some limitations of demanding certificates to confirm the facts or legal status known to them *ex officio* (amended art. 220 KPA). This normalization is not a novelty, because it precisely repeats the normalization from before the amendment of UINF. Transferred from the existing legal status were the normalizations forbidding the organ to demand certificates to confirm the facts or legal status which can be established by the organ according to its records, registers or other data, and also by virtue of official documents presented by the applicant (identity card, registration certificate or other documents). Particularly important is § 1.2 art. 220, which says, among others, that the „public administration organ must not demand a certificate to confirm the facts or legal status if they can be

established by the organ according to ... public registers held by other public entities which can be accessed by the organ electronically ...". We could not possibly predict today how the provisions of this article will apply in practice or when the organs acknowledge that they have access to public registers held by other offices.

One of the most serious barriers in electronic contacts between applicants and offices was the requirement to affix the application with a safe electronic signature verified by a valid qualified certificate. Starting from the day the UINF amendment came into force, i.e. from 17 June 2010, in the case of the applications and other letters addressed to the offices in administrative proceedings, the clients should have the possibility to affix them with either a qualified signature or a new type of electronic signature i.e. the signature confirmed by the ePUAP trusted profile. Three decrees of the Minister of the Interior were to appear with the act, without which the trusted profile cannot work. Because of at least three months' vacatio legis for the decrees, the earliest time of implementation of the trusted profile may be September this year. The range of the gratuitous use of the signature confirmed by the ePUAP trusted profile was confined to signing the letters addressed to public administration organs in general administrative proceedings and in tax proceedings. So in this amendment of UINF it was not the legislator's intention to give the Internet clients, free of charge, the electronic signature to sign all declarations of will, including agreements.

7.2. CHANGES IN ELECTRONIC CONTACTS WITH TAX OFFICES

Changes introduced to the act of 29 August 1997 – tax ordinance (hereinafter referred to as: ORDП) are strictly connected with those in the KPA act. On the other hand, supplements and modifications in the KPA are a consequence of introducing into UINF some concepts of electronic in-box and official certification of receipt and a new method of identification of teleinformatic systems user – ePUAP trusted profile. Changes in ORDП result mostly from adaptation to new solutions provided in the UINF.

Section 13 of ORDП, added to art. 3, refers within definition of the electronic document to art. 3 sec. 2 UINF. Pursuant to this UINF article, assumed as the

electronic document is the semantically separate set of data arranged in a specific internal structure and recorded on the informatic data carrier”.

Submission of applications, also in the electronic form, was comprehensively regulated in the provisions of art. 168 ORDП. Before the amendment of UINF, the following declarations could be submitted through the electronic communication means:

- tax returns for the value added tax,
- about the amount of profit (loss),
- for the tax on games,
- annual.

However, all the mentioned documents had to be affixed with a safe electronic signature verified by a valid qualified certificate. After the amendment, this requirement was liberalized through introduced gratuitous signature confirmed by the ePUAP trusted profile.

Art. 168 was added to art. 14h ORDП, containing a catalogue of regulations applied respectively in cases of individual interpretations. This enabled the people interested in submitting applications for individual interpretations to use the electronic communication means when submitting them, without the need to affix them with a safe electronic signature verified by means of a valid qualified certificate.

Due to the UINF amendment, art. 143 § 1a ORDП was specified more precisely, which enables a distribution of the tax organ’s certification rights onto other persons than those who fulfill the organ’s functions. The article indicates that the authorization given to the officer who acts on behalf of the tax office may comprise signing of letters in the form of electronic documents by an electronic signature verified by a valid qualified certificate. Decentralization of certification rights is consequently aimed at an improvement of the tax office’s work. But such authorization should not be confused with authorization to carry out the tax inspection (art. 283 § 1 ORDП) which cannot be issued in an electronic form because of the lack of a relevant regulation. On the other hand, the authorization itself, dealt with in art. 143 § 1a, is not delivered to the party in connection with launched proceedings and it does not have to be included into the case file.

The existing possibility of delivering the letters, except for the electronic communication means (if the tax organ was requested to deliver the letters electronically), was extended by those cases where the party gives its consent for such delivery (art. 144a ORDП). In its present form, ORDП does not have an option to deliver the letters electronically. This rests with the tax organ.

Specified more precisely were the principles of establishing the proceedings starting date at the party's request, launched by the electronic communication means (in art. 165 § 3b ORDП). Specifically, this is the day when the demand is introduced into the tax organ's teleinformation system. When the information reaches a relevant server of the tax organ, it should be understood as receiving the demand submitted by the electronic communication means.

Eliminated was the possibility of submitting applications telegraphically, by teleprinter and telefax (art. 168 § 1 ORDП), which dictated by the development of technology and the fact that the obsolete devices are no longer used. For applications submitted by the electronic communication means the contacts were provided between the tax organ and tax-payer through the tax organ's electronic in-box (created according to the UINF). Instead, in the KPA the possibility to submit applications telegraphically or by telefax remained, and there are no precisely specified requirements for the signature on the applications submitted by these devices. By the "application" we mean a demand, explanation, appeal and complaint as well.

Specified (in art. 168 § 3a ORDП) were the requirements to be met by the application submitted as an electronic document with reference to the solutions of teleinformation system user identification adopted in the amended UINF. Indicated were (art. 168 § 4 ORDП) the terms of confirming the submission of application as an electronic document, including adaptation changes connected with introducing the official certification of receipt in the UINF.

According to the regulations of art. 178 ORDП from before the amendment, the party has the right to survey the case file, make notes, copies and transcripts at the tax organ's seat during its working hours, in the presence of the tax organ's employee. Paragraph 4 (added to art. 178 ORDП) enables the party, but after identification in the

way determined in the UINF, an on line access to the letters submitted in the form of an electronic document and delivered in this form by the tax organ. Both before and after the amendment (pursuant to the provisions of art. 267 § 1 sec.3 ORDP) the party is encumbered with the costs of transcripts or copies, as specified in art. 178. Furthermore, issuing a certificate for the application is subject to a stamp duty. The certificate means certification of conformity of a duplicate, transcript, excerpt or copy carried out by governmental or self-governmental administration organs. The stamp duty amounts to 5 PLN, both for a full and only just started page.

Maintained was the obligation to attach to the application a receipt of payment of the due stamp duty or computer print of such a receipt to the public administration organ which considers the case.

Amendment of the UINF was not assumed to change executive provisions relating to the stamp duty. The inconvenience of having to attach to application a receipt of paid stamp duty results from §3 sec. 1 of the decree issued by the Minister of Finances (on 28 September 2007) on payment of the stamp duty. In the era of information and developed electronic banking, maintaining in the act of the obligation to present the receipt of payment of the stamp duty is an immense deficiency, forcing the Internet client to go to the office or send paper receipts of payment, and the clerks – to enforce accomplishment of the obligation to provide the document traditionally, on paper. The real possibility to settle affairs through electronic communication means will apply only to those people who are entitled to subjective exemption from the stamp duty, whereas all other people (including the physically handicapped) can deal only with those affairs where the stamp duty payment is not obligatory.

In 2010 the possibility to submit annual tax returns through the Internet was extended by the possibility to submit the annual tax returns by the Internet, hoping that over one million tax-payers would use this form. The following forms with enclosures could be sent electronically:

- PIT 36 – tax return showing the amount of generated income (incurred loss) by entrepreneurs;

- PIT 36L – tax return showing the amount of generated profit (incurred loss) by entrepreneurs paying the flat tax;
- PIT 37 – Annual settlement of incomes;
- PIT 38 – Tax return showing the amount of generated income (incurred loss) from the purchase or sale of shares;
- PIT 39 – Tax return on the generated income (incurred loss) from sale of properties.

Only the married couples which jointly settle their incomes had to apply earlier for the tax organ's certificate about choosing such a method of settling the tax. From the beginning of January till the end of April almost 2 million of various types of tax returns were submitted (almost twice more than throughout the entire 2009). From 355 thousand of 24 million tax-payers, only about 420 thousand forms were obtained.

According to the Ministry of Finances, such a low popularity of this service results from that the fact that the previous year income was not known and from the fear of the new situation and being accustomed to the paper confirmation with a seal. On the other hand, the tax-payers reported some problems with software and fear of intercepting the data, because the forms are sent without the transmission being coded.

In the e-taxes project the fact should be discerned that this time the tax office realized that tax-payers do not always use the Windows operative programme: the application to submit the forms was also available for Linux and MacOS systems.

Hoping that popularization of e-Taxes may yield as much as 1 billion PLN savings annually to the budget, the Ministry of Finances undertakes further initiatives:

- launching the tax-payers' central register, owing to which when completing the e-returns we do not need to give all identification data;
- submitting the amendment to e-PIT without the e-signature and sending by married couples joint tax returns according to simpler rules (at the beginning of 2011);
- opening individual tax accounts available on-line; it will be possible to preliminarily fill in the returns, e.g. on the tax-payers' behalf, through tax administration (during 2011–2013).

7.3. ELECTRONIC CONTACTS THROUGH SOCIAL UTILITY SERVICES OF WIKI TYPE

After the amendment the citizens are expected to show more interest in the public administration offices platform; perhaps it could even be converted into a social portal form.

Almost 1/3 of all adult Poles, i.e. nearly 2/3 of Internet surfers (acc. to CBOS report of July 2009) are registered in a social utility portal, while only 11% of internet surfers do not have accounts on them (acc. to PBI study). The office's page may be also treated as one of the instruments for constructing a local community, which – similar to social utility portals - perhaps may be based on electronic exchange of information between institutions and inhabitants of the region.

Enterprises and higher schools all over the world use the open source solutions, very often including services based on Wiki. The best known example of the use of such a system is Wikipedia. Services using the Wiki environment are oriented to mutual work on the contents interesting to users. They are a variant of the Content Management System, i.e. a system of managing the contents of the server allowing every entitled person to publish her/his own materials. They enable cooperation of many users at creating the internet contents, often without the need for authorization and with public access to editing of the contents of a given internet service.

Because of a considerable flexibility of Wiki solutions, this platform has many potential applications, useful also in the public administration sector. On what basis? If only due to:

- fast rate, simplicity of creating, co-creating and updating of documents, studies, translations, or various publications;
- versioning (possibility to store each consecutive version of the article);
- easy creation of references both to internal and external resources;
- simpler way of formatting and tagging, as compared to HTML;
- possibility to create groups having specific access rights;
- possibility to fulfill the role of files repository, e.g. in the form of legal acts sets.

Wiki will also perform well in the official Intranet as a platform for publishing and exchange of information between employees. It is very useful in management of informatic projects, informing on an on-going basis about the project's state, cooperation in project documentation, at group creation of project pages, lists of tasks to be performed, or calendars. It may be also used to create development plans or thematic dictionaries.

The potential inherent in Wiki was noticed in the USA, where Barack Obama appointed in March 2009, as the Chief Information Officer, Vivek Kundra, who opted for using Web 2.0 instruments in administration. Kundra wants to use the cheap Wiki software in office practice to facilitate cooperation of a larger group of people. He also realized that administration has a lot of information recorded on disks of governmental computers and plans to render them available to the population to enable all those interested to create their own services based on these resources [BEN09].

The Polish public administration noticed the potential hidden in social utility services. The public institution profile on Facebook, Twister or Blip may constitute a permanent instrument of promotion, conscious communication with clients, and may become a standard for effective marketing through social utility services.

Owing to the fan page profiles of the Masovian Regional Office, Lower Silesia, Łódź Regional Office, Prime Minister's Office, Cultural Warsaw, city of Wrocław or Pomeranian police, the number of people interested in the information presented by the mentioned institutions is over a dozen or so thousand. What's most important, the services enable direct contacts with the region's inhabitants and effective communication both ways.

Owing to the possibility of commenting on entries, inviting to various events, publishing of photos and films, and especially the less formal character of social utility services, these instruments are more friendly to users, as compared to the official web site.

The clerks and internet clients have a chance to conduct a communication dialogue instead of monologue. The clerks and officers should find important and useful information they can derive from inhabitants' entries and comments. Aiming at a dialogue with inhabitants, the social utility portal should present the contents most

close and important to people. It should contain information about the upcoming cultural events, changes in public communication, citizens' actions or tourism.

Microblogs on which information may be published from the level of the web page, any mobile device with access to the network, mobile phone through text messages or internet communicator, constitute an exceptionally good system of early warning about accidents, emergencies and natural disasters.

As to the requirement of putting up short and concise information, the conciseness of published contents is the primary feature, characteristic of microblogs. A single entry in the case of Blip must not be longer than 160 characters, whereas on Twitter – 140 characters. Thus, the need to edit a short message is forced by the mechanism of the microblog's functioning, but in respect of administration they do not exhibit any serious information value.

An advantage of the presence on social servers is also generating of entries on official web sites. For example:

- launching the profile of the Masovian Regional Office on Facebook in January 2010 increased the number of entries on the official web site of Masovia by approx. 20%;
- in just one month, i.e. February 2010 owing to the profiles set up on Facebook, Twister and Blip the web site of the Lower Silesia region noted additionally about 220 visits;
- during two months, owing to the Cultural Warsaw profile on Facebook, 8000 visits to the city's official web site were generated directly from the information published there;
- through the mechanism rendered available by the creators of the tested new Warsaw municipal web site allowing to share by just one click the articles found on the web site – on own profiles of over 200 social utility services.

People turn to the internet to find the current, fast transmitted, short and pithy information. The specificity of social utility portals can meet these requirements and actually present messages as soon as they are created. The employees of the

Masovian Regional Office plan to use, apart from Facebook, also Twitter to inform e.g. about floodings in Masovia.

The primary purpose for which public institutions started using this instrument was first of all a free warming up of their own image. On Blip, apart from the Polish Postal Service, Social Insurance Company, Polish State Railways, also present are now among others: the Prime Minister's Office, Ministry of Foreign Affairs, Ministry of the State Treasury. But in the internet surfers' opinion the information presented by these institutions is dry, official and rather unattractive. For example the Ministry of Foreign Affairs „blipped” with the following words: “In view of the state of emergency announced by the authorities of Thailand, the Ministry of Foreign Affairs advises caution to those who travel to that country”.

Using the internet as an easy and effective way of communication to increase the citizens' interaction and participation in public life is thereby reinforced by new technologies and applications, such as blogs, Rich Site Summary (RSS) and „mobile administration” (providing not only the local access but almost complete coverage by wideband services).

Offices should not have any problems with conciseness of information presented on microblogs (because of the limited number of characters in a single entry), but they should work a bit on the form and contents of messages. It is worthwhile, because the social utility services of Wiki type can enhance the offices' brand and image – free of charge.

Interesting and useful proposals for constructing the office's positive and friendly image may be found on the profile of the Masovian Regional Office on Facebook, e.g. about closing the regional road for TIR lorries; the Lower Silesia Marshal Office created for internet surfers a possibility to participate in the competition for a hit of Wrocław agglomeration. A modern railbus on the railway line from Wrocław to Trzebnica was chosen. The popularity of the promotional action carried out on Facebook in February, called: „THE LOWER SILESIA HAS ITS COAT OF ARMS AND FLAG and we have them for our fans” surprised even its organizers. The region inhabitants who declared their willingness could gratuitously obtain the flag and a lot of people, proud of living in the Lower Silesia and ready to manifest such membership,

sent emails asking for the flag. „Also with me it is flapping beautifully – reported the inhabitants of other towns and villages in the region.” [BIE10]. Another idea evidencing high initiatives of Lower Silesia officers is a new action owing to which the Lower Silesia enthusiasts are given stickers for cars with the region’s logo.

The employees of the press team of the Regional Police Headquarters in Gdańsk constructed a positive image using the Web 2.0 instruments, creating the „Pomeranian police” channel in YouTube service and a profile of the same name on Facebook. In this way they wanted to show the society that their work consists in prevention and help for the people rather than in spectacular action and chase. As the first film in the YouTube service the Pomeranian policemen presented the material about six-year old Krystian’s dream coming true: they allowed him for one day to impersonate the role of the law enforcement officer. Other good actions of that police entity were e.g.: searching on Facebook for witnesses to an assault or carrying out a campaign for voluntary blood donation. In turn, the traffic police in Słupsk boasted of their new purchase saying: „the legend of American motorization – beautiful 250-horse power Cadillac with videorecorder, the purchase of which was financed by our compatriots living in the USA.”

When promoting the office on Facebook or Blip, one should resign from formal language, as this could particularly irritate the internet surfers and discourage them from following the profile. The press team of the Pomeranian police in its profile on Facebook wittily presented the useful information about localizations of radar systems for speed measurements: „You can find below a link with a list of sites where at the beginning of April you can take a photo of yourself while driving”).

Using the Web. 2.0 type instruments, one can not only stop the office monologue, start a dialogue with the internet surfer and acquire feedback information from him, but also give him a chance to evolve from the recipient into a sender and creator of the contents, among other through the function of commenting on the entries. The aim is that the internet surfers informed each other about important and interesting events in their towns and communes. For example on the “Cultural Warsaw” profile the organizers of the capital city’s cultural events invite to various events, and the fans review the performances and concerts themselves.

Those offices which will treat the Internet as a bit of news which one should have because it's right to have it will not construct their positive image. The studies (six of seven companies do not have any official policy of using social utility portals) carried out by IESE Business School (Spain), E. Philip Saunders, College of Business at the Rochester Institute of Technology (USA) and Henley Business School (Great Britain) indicate that the lack of official strategy of using the social utility portals may cost the company a lot in the case of uncontrolled leak of data. The use - by public administration- of social utility instruments is a novel phenomenon in that using the results of studies carried out for business sphere one should try and get prepared to it by developing a policy of using such technology and deciding which information will be published. It is worthwhile to choose a person or even team responsible for it.

The intentionally created policy of using the Web. 2.0 instruments must result from the analysis of expectations and needs of the profile founders and users. The profile on social utility services assures a higher attractiveness, as compared to microblogs, but for unknown reason the profiles are used only as microblogs. Another interesting observation: Facebook's functionality is higher than that of Nasza Klasa portal founded by Wrocław inhabitants, but it is on Klasa that over 20 towns run their portals, including: Swarzędz, Legnica, Bydgoszcz, Gorzów Wielkopolski, Częstochowa. This may result from the fact that Nasza Klasa declares supporting the town profiles, e.g. allowing for special graphic distinguishing of their accounts.

The mentioned instruments allow the citizens to easily propose their governments their own ideas and place comments on environmental protection, regional policy and even political issues, and similarly to obtain feedback information so that both parties participating in such an interactive contact derive advantages from such arrangement.

Tab. 7.1. Desirable characteristics of information placed by offices on social utility portals

DESIRABLE CHARACTERISTICS OF INFORMATION PLACED ON SOCIAL UTILITY PORTALS	DESCRIPTION OF CHARACTERISTICS
Information placed on portal should be selected. Placing any information should be preceded by specifying the policy of using the social utility portals	Including the knowledge about who the users are and what sort of information they expect
Entries should not be written in a typical official language; the vocabulary should at best be short, witty and adjusted to the medium's nature	Attractive information titles and short sentences should be used, which intrigue and encourage the reader to click on the link to the message on the official page .
The presented information should be update and useful, closest and most important to people	It should be the information which is expected by a given portal community, but the purposes important for the office for which the profile was set up must not be omitted.
To assure a positive image of the office	Not only the official web site, but the profile on a popular social utility portal may be an extension of the entity's information activities.
To enable presentation of certain situations in the form of video material without any drastic depictions.	E.g. a film showing how to secure apartments against burglaries, floods; video material from breaking up a criminal group dealing with sale of cars; information transmitted to people in such way will be much easier to remember.

7.4. CONCLUSIONS

When amending the acts we should not focus mainly on linguistic formulation of articles but depending on their purposes we should enable their safe use.

The problem with establishing the terms of delivering letters may result in that if a matter cannot be settled in the form of an electronic document delivered by electronic communication means, it has to be settled in the traditional, written form.

The modified art. 391 KPA (dealing with the way and form of the document to be delivered) will surely contribute to increased turnovers in companies creating the systems used for electronic management of documentation. However, many public entities are not financially prepared to informatization of their activities, i.e. the e-government.

Eliminating the electronic mail from the catalogue of application submission methods will make the administration clients wishing to submit an application through the Internet, forced to do this only and exclusively through the electronic in-box. This will surely increase the popularity of the in-box as a communication means with the office. However, it will not be the means of choice, but of necessity.

Regulations on telegraphic applications submitted by teleprint and telefax were eliminated by amendment of UINF from ORDP, but not KPA. So in this respect the project initiator's inconsistency relating to ORDP and UINF appears.

We should expect that limiting the number of issued certificates to confirm the facts or legal status will bear a disadvantageous financial effect for organs issuing such certificates. After all, the receipts from issuing certificates constitute the communes' incomes.

Moreover, if the legislators cannot amend legal acts without inconsistency and omissions which should be removed, they should do this often, and the frequency of every five years seems to be insufficient.

Towns and communes more and more often are market oriented when creating their images and perceive the profits derived from promotion in the Internet. The internet pages of communes or offices will never generate as many visits as the profile on the popular social utility portal, where internet surfers meet as they have their acquaintances there or just look for some entertainment. Administration entities,

aware that the main course of life runs on the social utility services and determined to participate in it, create their profiles exactly there. However, using the Web 2.0 instruments should be preceded by defining which information will be published and depicted in the office's policy of communication with the strategic surroundings, i.e. inhabitants and investors. Before sensitive personal data are disclosed, detailed principles of functioning on social utility portals should be presented in statutory rules, or guidelines for the entire public sector should be worked out.

REFERENCES

- [BEN09] Bendyk Edwin, *Sieć ze starych sznurków*, <http://www.polityka.pl/archive/do/registry/secure/showArticle?id=3362576>, 21.07.2009
- [BIE10] Bielińska-Kuniszewska K., Kuniszewski S.: *Użyteczne socjal media*. w: IT w administracji nr 05, PRESSCOM, maj 2010
- [CHM10] Chmielarz W.: *Studium rozwoju systemów e-administracji w Polsce*, <http://web.ae.katowice.pl/stanley/konferencja/pdf/Chmielarz.pdf>, 6.10.2010
- [GOV10] Ustawa o informatyzacji działalności podmiotów realizujących zadania publiczne z dnia 17 lutego 2005 Dz.U. nr 64, poz. 565 z 2005 r., z późn. zm.; stan prawny na dzień 29 lipca 2010 r., <http://www.polskieustawy.com/print.php?actid=3098&lang=&adate=20100729>
- [KAM10] Kamiński M.: *Zaufaj profilowi*. w: IT w administracji nr 04, PRESSCOM, kwiecień 2010
- [RZY10] Rzymowski J., Promiński Z.: *Kodeks postępowania elektronicznego*. w: IT w administracji nr 04, PRESSCOM, kwiecień 2010

Features of Software Development for Natural Language Processing

Natural Language processing (NLP) is a field of computer science concerned with the interactions between computers and human (natural) languages. NLP also has many common areas with linguistics.

The following is an incomplete list of tasks in NLP:

- Machine translation: automatically translate text from one human language to another.
- Part-of-speech tagging: determine the part of speech for each word in a sentence.
- Question answering: determine an answer to a question asked in natural language.
- Relationship extraction: identify the relationships among named entities in a text.
- Topic segmentation and recognition: separate a text into segments each of which is devoted to a topic, and identify what topic a segment is devoted to.
- Word sense disambiguation: determine a meaning of word with multiple meanings which suits given context.

Some of these tasks have a very close relation to semantics (meaning).

8.1. APPROACHES TO AND METHODS FOR SOLVING OF NATURAL LANGUAGE PROCESSING PROBLEMS

Machine translation methods

Machine translation (MT) is a subfield of computational linguistics, which studies the use of computer software to translate text or speech from one natural language into another. At a basic level, the MT performs a simple replacement of words from one natural language into words of another. Use of more complex methods makes it possible to attempt a more complicated translation, allowing better handling of differences in linguistic typology, phrase recognition and translation of idioms.

The translation process can be described as:

- Decoding of the source text's meaning.
- Recoding of that meaning in the target language.

Behind this seemingly simple procedure lies a complex cognitive activity. To decode the meaning of the original text as a whole, the translator must interpret and analyze all the features of the text - a process that requires a deep knowledge of grammar, semantics, syntax, idioms, etc. the source language and the culture of its speakers. Translator needed such as in-depth knowledge in the target language for the conversion of meaning.

Therein lies the difficulty of machine translation: how to program a computer so he could "understand" the text as it makes people and "create" a new text in the target language, which would be written like a man.

There are several approaches to this problem.

Rule-based MT

The rule-based machine translation consist of several approaches: transfer-based machine translation, interlingual machine translation and dictionary-based machine translation.

Interlingual MT

Interlingual machine translation is one of the classic approaches to machine translation. In this approach, the source text is transformed into an interlingua, an abstract language-independent representation. The target language is then generated from the interlingua.

Advantages of interlingual approach are:

- it requires fewer components in order to relate each source language to each target language;
- it takes fewer components to add a new language;
- it supports paraphrases of the input in the original language;
- it allows both the analyzers and generators to be written by monolingual system developers;
- it handles languages that are very different from each other.

Disadvantage of interlingual approach is that the definition of an interlingua is difficult and maybe even impossible for a wider domain.

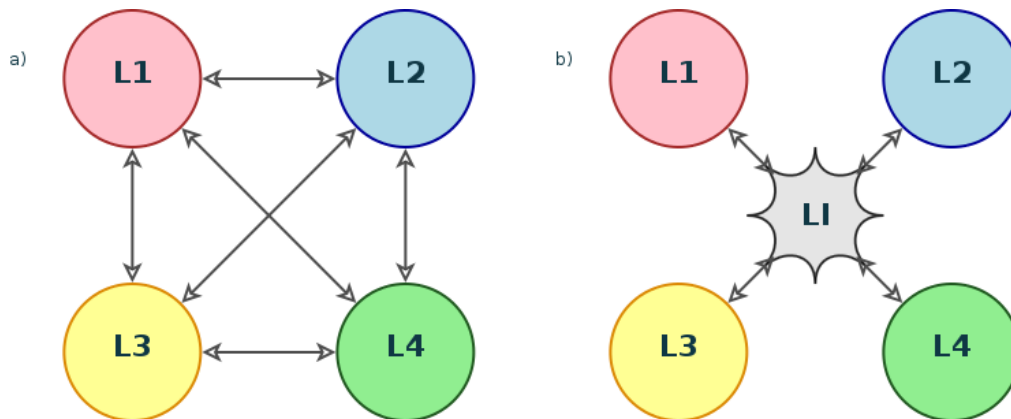


Fig. 8.1 a) Translation graph for direct transfer-based machine translation (4 languages, 12 modules required); b) Translation graph for using a bridge language (4 languages, 8 translation modules required)

Sometimes two interlinguas can be used in translation. It is possible that one of the two covers more of the source language's characteristics, and the other covers

more of the target language's characteristics. The translation process is shown in the next picture.

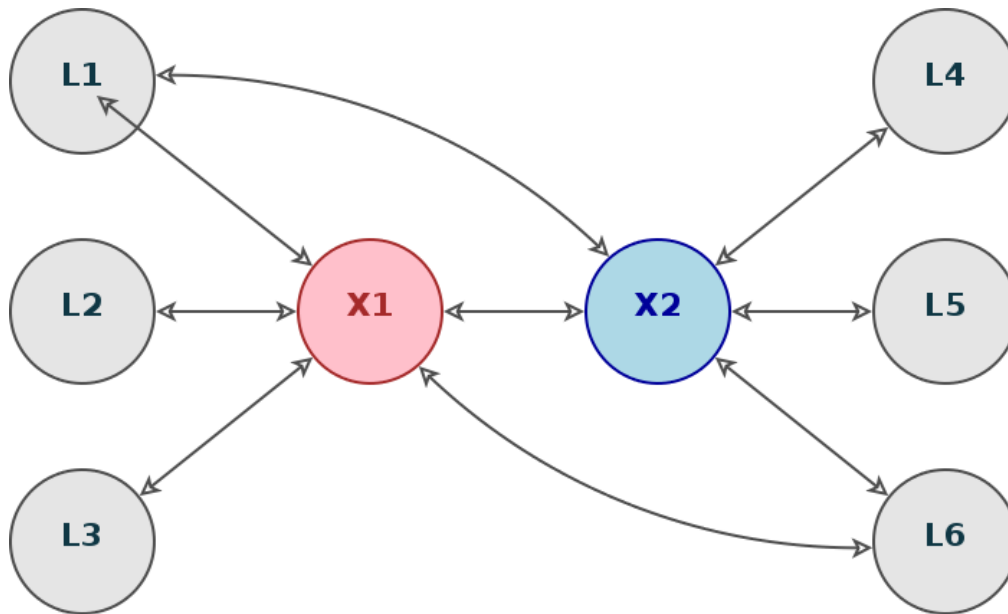


Fig. 8.2: Translation using two interlinguas

Transfer-based machine translation

Transfer-based machine translation is based on the idea of interlingua and is currently one of the most widely used methods of machine translation.

Main idea of transfer-based machine translation is next: it is necessary to have an intermediate representation of the original sentence in order to generate the correct translation. In transfer-based MT intermediate representation has some dependences on the language pair involved. The way in which transfer-based machine translation systems work is simple: it applies sets of linguistic rules which are defined as correspondences between the structure of the source language and that of the target language. The first step includes analyzing the input text for morphology and syntax to create an internal representation. The translation is generated from this representation using bilingual dictionaries and grammatical rules.

Quality of translation using this approach depends on the language pair it is applied to.

Statistical approach

Statistical machine translation (SMT) is an approach to machine translation where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora.

The idea behind statistical machine translation comes from information theory. A document is translated according to the probability distribution $p(e | f)$ that a string e in the target language is the correct translation of a string f in the source language. The problem of modeling the probability distribution $p(e | f)$ has been approached in several ways. One approach is to apply Bayes Theorem, $p(e|f) \propto p(f|e)p(e)$, where the translation model $p(f | e)$ is the probability that the source string is the translation of the target string, and the language model $p(e)$ is the probability of seeing that target language string. This decomposition splits the problem into two subproblems. Finding the best translation \tilde{e} is done by picking up the one that gives the highest probability:

$$\tilde{e} = \arg \max_{e \in e^*} p(e | f) = \arg \max_{e \in e^*} p(f | e) p(e) \quad (8.1)$$

Texts are typically translated sentence by sentence. Language models are usually approximated by smoothed n-gram models, and similar approaches have been applied to translation models, but there is some complexity due to different sentence lengths and word orders in the languages. The statistical translation models were initially word based, but significant advances were made with the introduction of phrase based models.

Word-based translation

In word-based translation, the fundamental unit of translation is a word in some natural language. Typically, the number of words in translated sentences are different, because of compound words, morphology and idioms. The ratio of the lengths of sequences of translated words is called fertility, which tells how many foreign words each native word produces. Necessarily it is assumed by information theory that each

covers the same concept.

Simple word-based translation can't translate between languages with different fertility. Word-based translation systems can relatively simply be made to cope with high fertility, but they could map a single word to multiple words, but not the other way about.

The word-based translation is not widely used today, phrase-based systems are more common. The alignments are used to extract phrases or deduce syntax rules. Matching words in bi-text is still a problem actively discussed in the community.

Phrase-based translation

The aim of phrase-based translation is to reduce the restrictions of word-based translation by translating whole sequences of words, where the lengths may differ. The sequences of words are called blocks or phrases. Usually they are not linguistic phrases but phrases found using statistical methods from corpora. Restricting the phrases to linguistic phrases decreases the quality of translation.

Example-based MT

Example-based machine translation (EBMT) approach to machine translation uses of a bilingual corpus with parallel texts as its main knowledge base at run-time.

At the foundation of example-based machine translation is the idea of translation by analogy. Example-based machine translation systems are trained from bilingual parallel corpora, which contain sentence pairs like the example shown in the table. Sentence pairs contain sentences in one language with their translations into another. So called "minimal pair" consists of the sentences that vary by just one element. These sentences make it simple to learn translations of subsentential units. Composing these units can be used to produce translations in the future.

EBMT is best suited for sub-language phenomena like phrasal verbs. Phrasal verbs have highly context-dependent meanings. Phrasal verbs produce specialized context-specific meanings that may not be derived from the meaning of the constituents.

EBMT also can be used to determine the context of the sentence.

Hybrid MT

Hybrid machine translation (HMT) combines the strengths of different translation methodologies. The approaches can be used in a number of ways:

- Rules post-processed by statistics: Translation is performed using a rules based engine. Statistics are then used in an attempt to improve the output from the rules engine.
- Statistics guided by rules: Rules are used to pre-process data in an attempt to better guide the statistical engine. Rules are also used to post-process the statistical output to perform functions such as normalization.

8.2. APPROACHES TO EVALUATING THE QUALITY OF MACHINE TRANSLATION

A measurement of the quality of the machine translation output is usually called a metric. The task for any such metric is to assign scores of quality in such a way that they correlate with human judgment of translation quality.

The measure of evaluation for metrics is correlation with human judgment. This is generally done at two levels, at the sentence level, where scores are calculated by the metric for a set of translated sentences, and then correlated against human judgment for the same sentences. And at the corpus level, where scores over the sentences are aggregated for both human judgments and metric judgments, and these aggregate scores are then correlated.

Good performance of a metric, across text types or domains, is important for the reusability of the metric. A metric that only works for text in a specific domain is useful, but less useful than one that works across many domains. Another important factor in the usefulness of an evaluation metric is to have good correlation, even when working with small amounts of data.

Attributes that a good automatic metric should have are:

- correlation;
- sensitivity;
- consistency;

- reliability;
- generality.

Nowadays there are several approaches used to evaluating the quality of machine translation.

BLEU

BLEU (Bilingual Evaluation Understudy) is an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another. Quality is considered to be the correspondence between a machine's output and a human-translator's output. BLEU was one of the first metrics to achieve a high correlation with human judgments of quality and remains one of the most popular.

Scores are calculated for individual translated segments—sentences—by comparing them with a set of good quality reference translations. The scores are then averaged over the whole corpus to reach an estimate of the translation's overall quality. Intelligibility or grammatical correctness is not taken into account.

BLEU is designed to approximate human judgment at a corpus level, and does not show good results when used to evaluate the quality of individual sentences.

BLEU's output is always a number between 0 and 1. This value indicates how similar the candidate and reference texts are, with values closer to 1 representing more similar texts.

BLEU uses a modified form of precision to compare a candidate translation against multiple reference translations. The metric modifies simple precision:

$$P=m/w_t \quad (8.2)$$

where m is number of words from the candidate that are found in the reference, and w_t is the total number of words in the candidate.

In order to produce a score for the whole corpus the modified precision scores for the segments are combined, using the geometric mean multiplied by a brevity penalty to prevent very short candidates from receiving too high a score.

BLEU has frequently been reported as correlating well with human judgment, and

remains a benchmark for the assessment of any new evaluation metric. There are however a number of criticisms that have been voiced. It has been noted that although in principle capable of evaluating translations of any language, BLEU cannot in its present form deal with languages lacking word boundaries.

It has been argued that although BLEU has significant advantages, there is no guarantee that an increase in BLEU score is an indicator of improved translation quality.

NIST

NIST is a method for evaluating the quality of text which has been translated using machine translation. Its name comes from the US National Institute of Standards and Technology.

It is based on the BLEU metric, but with some alterations. Where BLEU simply calculates n-gram precision adding equal weight to each one, NIST also calculates how informative a particular n-gram is. That is to say when a correct n-gram is found, the rarer that n-gram is, the more weight it will be given.

NIST also differs from BLEU in its calculation of the brevity penalty insofar as small variations in translation length do not impact the overall score as much.

Word error rate

Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system.

The general difficulty of measuring performance lies in the fact that the recognized word sequence can have a different length from the reference word sequence. The WER is derived from the Levenshtein distance, working at the word level instead of the phoneme level.

This problem is solved by first aligning the recognized word sequence with the reference word sequence using dynamic string alignment.

Word error rate can then be computed as:

$$WER = \frac{S + D + I}{N} \quad (8.3)$$

where:

S - number of substitutions,

D - number of the deletions,

I - number of the insertions,

N - number of words in the reference.

When reporting the performance of a speech recognition system, sometimes word accuracy (WAcc) is used instead:

$$WAcc = 1 - WER = \frac{N - S - D - I}{N} = \frac{H - I}{N} \quad (8.4)$$

where:

H is N-(S+D), the number of correctly recognized words.

METEOR

METEOR (Metric for Evaluation of Translation with Explicit ORdering) is a metric for the evaluation of machine translation output. The metric is based on the harmonic mean of unigram precision and recall, with recall weighted higher than precision. It also has several features that are not found in other metrics, such as stemming and synonymy matching, along with the standard exact word matching. The metric was designed to fix some of the problems found in the more popular BLEU metric, and also produce good correlation with human judgment at the sentence or segment level. This differs from the BLEU metric in that BLEU seeks correlation at the corpus level.

As with BLEU, the basic unit of evaluation is the sentence, the algorithm first creates an alignment between two sentences, the candidate translation string, and the reference translation string. The alignment is a set of mappings between unigrams. A mapping can be thought of as a line between a unigram in one string, and a unigram in another string. The constraints are as follows; every unigram in the candidate

translation must map to zero or one unigram in the reference translation and vice versa. In any alignment, a unigram in one string cannot map to more than one unigram in another string.

Each stage is split up into two phases. In the first phase, all possible unigram mappings are collected for the module being used in this stage. In the second phase, the largest subset of these mappings is selected to produce an alignment as defined above. If there are two alignments with the same number of mappings, the alignment is chosen with the fewest crosses, that is, with fewer intersections of two mappings. From the two alignments shown, alignment (a) would be selected at this point. Stages are run consecutively and each stage only adds to the alignment those unigrams which have not been matched in previous stages. Once the final alignment is computed, the score is computed as follows: Unigram precision P is calculated as:

$$P = \frac{m}{w_t} \quad (8.5)$$

Where m is the number of unigrams in the candidate translation that are also found in the reference translation, and w_t is the number of unigrams in the candidate translation. Unigram recall R is computed as:

$$R = \frac{m}{w_r} \quad (8.6)$$

Where m is as above, and w_r is the number of unigrams in the reference translation. Precision and recall are combined using the harmonic mean in the following fashion, with recall weighted 9 times more than precision:

$$F_{mean} = \frac{10PR}{R + 9P} \quad (8.7)$$

The measures that have been introduced so far only account for congruity with

respect to single words but not with respect to larger segments that appear in both the reference and the candidate sentence. In order to take these into account, longer n-gram matches are used to compute a penalty p for the alignment. The more mappings there are that are not adjacent in the reference and the candidate sentence, the higher the penalty will be.

In order to compute this penalty, unigrams are grouped into the fewest possible chunks, where a chunk is defined as a set of unigrams that are adjacent in the hypothesis and in the reference. The longer the adjacent mappings between the candidate and the reference, the fewer chunks there are. A translation that is identical to the reference will give just one chunk. The penalty p is computed as follows,

$$p = 0.5 \left(\frac{c}{u_m} \right)^3 \quad (8.8)$$

Where c is the number of chunks, and u_m is the number of unigrams that have been mapped. The final score for a segment is calculated as M below. The penalty has the effect of reducing the F_{mean} by up to 50% if there are no bigram or longer matches.

$$M = F_{\text{mean}}(1 - p) \quad (8.9)$$

To calculate a score over a whole corpus, or collection of segments, the aggregate values for P , R and p are taken and then combined using the same formula. The algorithm also works for comparing a candidate translation against more than one reference translations. In this case the algorithm compares the candidate against each of the references and selects the highest score.

8.3. TECHNIQUES OF SEMANTIC MODELING

In the machine translation system based on rules, the source code the first thing

analyzed morphologically and syntactically, to get a syntactic representation. Various methods of analysis and transformation can be used to obtain the final result. To choose methods and emphases are heavily dependent on system design, however, most systems include at least the following steps: morphological analysis, lexical categorization, lexical conversion, structural transformation, morphological synthesis, syntactic transformation (surface), semantic transformation (deep). Dwell in particular on semantic change and semantic modeling techniques. First language semantics - is a branch of linguistics which studies the semantic aspect of language that is meaning, the meaning of linguistic units (morphemes, words, phrases, etc.)

Semantic transformation (deep). This creates a level of semantic representation, which depends on the source language. This presentation may consist of a series of structures that represent meaning. In these systems, the translation is usually done predicates. Translation is also usually requires a structural transformation. This level is used to translate between more distant languages, or languages that have had no genetic relationship (Spanish - English or Spanish - the language of the Basques, etc.).

Methods of formal semantics

There are many methods, but the most famous and popular are the following:

- Method of component analysis,
- The method of semantic cases,
- Semantic networks.

Method of component analysis

In English, a separate line of compositional semantics. It was assumed that by a finite set of semantic components can be described as an unlimited set of lexical items. Technique selection semantic factors is to consider the allocation of certain words and signs, dividing the words into different classes and semantic groups, for example, on grounds such as animate / inanimate, male / female gender, etc. can be identified and more differentiated features for word classes, such as animals, fish, birds, people, etc. Meaning of each word thus appears as the set of semantic factors.

Consider a concrete example of this method. Take, for analysis of the word "journal". First, we must find a word or phrase indicating the kind of things, which is a kind of journal. This word - a periodical.

1. The value of this generic terms (hyperons) is the first semantic component within the definition of the word "journal". It displays the general features of the magazine with other publications of this kind (2 trait = frequency of publication). These common symptoms are called integral semantic features.

2. Search for all words denoting other kinds of periodicals and identify those attributes for which logs are different from other kinds of periodicals. Such signs are called differential semantic features.

In addition to magazines, periodicals are newspapers, newsletters, catalogs. From newspapers magazines differ in that they are stitched. If printed publications are not stitched, it is not a magazine. From the newsletter and directory journal is different on other grounds not related to registration of the publication and its contents. For example, create directories for publishing data about the product. Thus, the interpretation of the word "journal" includes, besides the integral sign, 2 differential. For the magazine are the components that characterize it from the look and feel of the content.

Method of component analysis is actively developing overseas. There are different theories of compositional semantics (Katz and Fodor scientists - pioneers, Barbara Patty, Anna Vezhbitskaya, etc.). For example, Vetbitskaya comes from the fact that the values of all words in all languages can be described using the same limited set of elements as irreducible atoms in physics, ie semantic primitives: many pronouns, numerals, verbs (doing, being able to think, speak, to know, have), size (large and small), adverbs (where, when), etc.

Basically, for the method of component analysis distinguished the thesis that "the meaning of sentences is the sum of meanings of its constituent words."

The method of semantic cases

Great contribution to the development of language for writing semantic structures and forms, introduced by Charles Fillmore. He accepted the hypothesis component

structure values and the idea of sequential expansion of word meaning into simpler components up to the semantic primitives or atoms of meaning. Sharing the common views on the predicate argument structure, he concludes that it is necessary to specify not only the number of arguments of the predicate, but their role is semantic content. He identifies the following roles:

- 1) agent - animate the initiator of the action,
- 2) object - a thing which is the subject of,
- 3) counterparty - the force against which the action,
- 4) recipient - the person for which an action,
- 5) tool - the physical cause of action / motivation,
- 6) source - the original state of the object to the action.

He also offered a detailed concept of lexical meaning. He is a classic of lexical semantics abroad.

The common conception of lexical meaning based on the concepts of layering, ie, includes shades of meaning, stylistically and emotionally expressive elements of values.

He goes further and adds value in two parts: the actual value and presubpozitsiya. For example, in saying "Vasya - not a bachelor" does not assert that Vasya was not a man. Ie if we assume that the word "bachelor" - an adult male never-married, then the negative is only the second part after the comma, which is the actual value.

The main result of this research is to review Fillmore usual scheme of entries in dictionaries. He believes the primary means of vocabulary task of semantic role structures and rules of their transfer to surface structures that are common to the concepts of such structures in the Russian research management models YD Apresian.

This theory of semantic cases and semantic dictionaries developed in the Moscow school of semantics, which created a model "text - the meaning - the text" and in particular the explanatory combinatorial dictionary of modern Russian language. For such a dictionary was developed apparatus of lexical features that are similar in the sense of the unit semantic cases Filmore. In 1984 was published version

of the dictionary "New Explanatory Dictionary of Synonyms of the Russian language."

Semantic networks

Technology of semantic networks can lead civilization to a new level. It is for the reason that the government of U.S.A., France, Germany and other countries are investing in these developments a huge fiscal resources. These technologies are expected to address the context and as a result - the establishment of information systems, artificial intelligence. Programming will be possible in natural languages, the creation of smart weapons of the battlefield, advanced search and expert systems, and much more.

Unambiguous definition of the semantic web is not currently available. In knowledge engineering it's means a graph showing the meaning of the complete image. Graph nodes correspond to concepts and objects, and arcs - relationship between objects. Formally, the network can be defined as follows:

$$H = \langle I, C, G \rangle \quad (8.10)$$

where:

I - a set of information units;

C - many types of relationships between information units;

G - the map that defines the specific relationship of existing types between elements.

Semantic Web as a model most often used to represent declarative knowledge.

One of the first well-known models based on the semantic web is a TLC-model (Teachaple Language Comprehender - affordable mechanism for understanding the language), developed by Kuillian in 1968. Model was used to represent the semantic relationships between concepts (words) to describe the structure of long-term memory in human psychology.

Semantic networks, designed as a common unit of knowledge representation, from the very beginning actively used for the construction of systems of natural language processing. We consider various ways of representing cognitive content of statements in the NL with semantic networks. The simplest - the logical predicate:

typed semantic relation connects two nodes. Since this representation does not allow to denote events in the node-set is turned on and the predicate itself and the circumstances of the action associated with it a set of relations. Semantic network of this type, shown in Fig. 8.3, describes the importance of English phrases On Wednesday morning John hit Mary in the park by the fist as a set of nodes corresponding to objects or conceptual notions and related some of directed bonds.

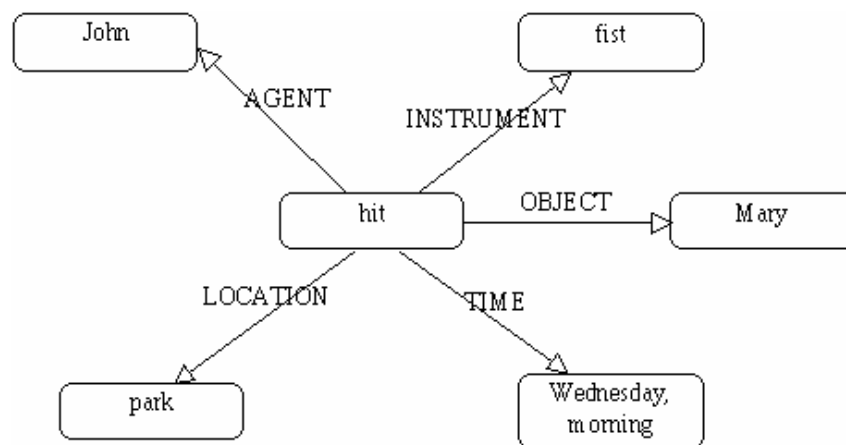


Fig. 8.3.

G.S. Tseitlin offered his version of the presentation value of ER-expressions with the help of semantic networks. His associative networks contain nodes that express some entity (the relevant objects in the text or in the outside world), and directed arcs connecting these nodes. Site content may be a number, a string of symbols, the procedure or a finite set of other nodes. Arcs are named, and the names of all arcs leaving the node must be distinct, often they are intended to semantic and syntactic roles. In particular, the words "open" and "close" can be used to refer action with a wide range of objects: open / close door, a bottle, a computer file. In terms of associative networks, these words can also be represented as arcs that connect some types of objects and method for appropriate action.

System SNePS (Semantic Network Processing System) is widely used as a means of developing pilot applications using natural language, since it includes

basic reporting mechanism and the automatic construction of semantic networks with minimal structure and mechanism of the withdrawal on those networks. The mechanism of ATN-grammars can also be described in terms of SNePS (ATN-grammar of English is part of the basic package SNePS). The close connection of this system with processing tasks ER is also due to the fact that architecture is not focused on the manual creation of semantic networks, and on its construction as a result of extracting knowledge from various sources, often ER-texts.

Theory of rhematic graphs usually use the mechanism of network representations to reflect the phenomena characteristic of language as a communication tool rather than as a mechanism for modeling the knowledge about the world (ontology). This theory is based on the mathematical theory of lattices (partially ordered sets) and provides a view of both the semantic, syntactic and phonetic information. Two rhematic graph, the relevant proposals "Larry was reading some trash" and "Larry was reading a comic bought at the station" and depicted in Fig. 8.4, constructed from the parse tree of the proposals by converting it into an acyclic graph. This process involves the fusion of the leaf nodes belonging to the same object, and making order in accordance with the focus of attention: in Fig. 8.4, dashed lines link the reference point (point of reference) and a point of interest.

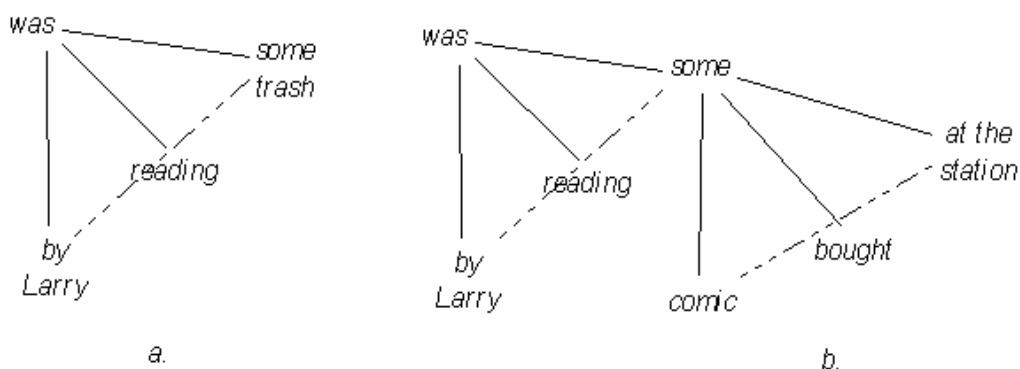


Fig. 8.4.

Based rhematic graphs combined into a single view when analyzing a text (see

the analysis of the proposal "Larry was reading a comic he had bought at the station" in Fig. 8.5), is a logical conclusion by modus (passive-active relation between action and situation), the type of operation (primary, secondary, tertiary) and arrays containing the classical class-subclass relations and phase (direction of action, including source, purpose and position).

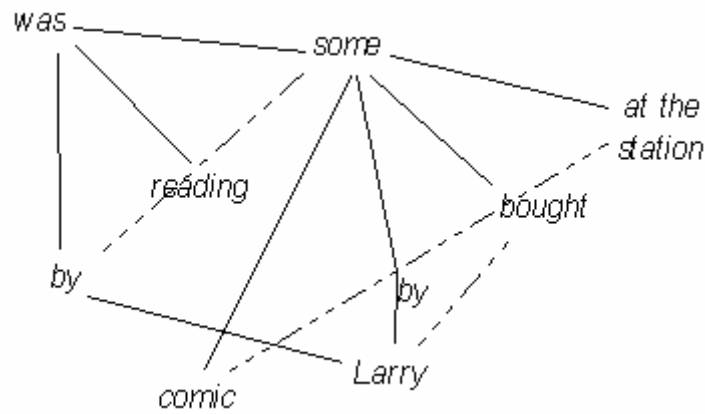


Fig. 8.5.

Joint lattice of these relationships is presented in Fig. 8.6.

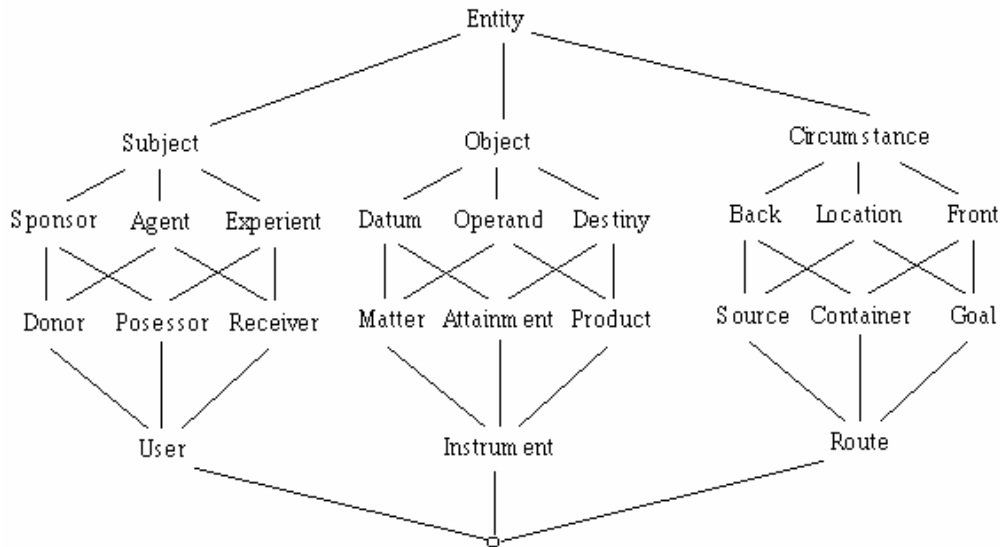


Fig. 8.6.

System SNOOP (System with Networks and Object-Oriented Productions), offers another way to represent semantic networks for language problems by integrating network representation with the methods of object-oriented programming and production rules. The network nodes and relations between them belong to the classes, described by a separate network inheritance in which the class is defined by a set of possible fields (which allows nodes to have an internal structure) and properties associated with objects relevant to this class. The properties of objects of classes described by the groups of production rules, each of which consists of the sample, allowing to navigate through network conditions at the sites found by the sample, and actions to change the internal structure of these sites and change the network (creation and deletion of nodes and relations).

Classification of semantic networks

For all semantic networks is true separation of the arity and the number of types of relationships.

The number of types of relationships, the network can be homogeneous and

heterogeneous:

Homogeneous networks have only one type of relationship (arrows), for example, such is the above-mentioned classification of species (with the only attitude). In heterogeneous networks the number of types of relationships more than two. The classic illustration of this model of knowledge representation represent just such a network. Heterogeneous networks are of great interest for practical purposes, but also a great challenge for research. Heterogeneous networks can be represented as a tree-like interweaving of multilayer structures. An example of such a network may be Semantic Web Wikipedia.

Number of types of relationships in the semantic network is defined by its creator, based on specific goals. In the real world, their number tends to infinity. Each relationship is, in fact, a predicate, simple or compound. Speed of work with the knowledge base depends on how effectively implemented treatment programs necessary relations.

- The arity:
 - Are typical network with **binary** relationships (connecting exactly two concepts). Binary relations are very simple and convenient to represent the graph as an arrow between two concepts. In addition, they play a crucial role in mathematics.
 - In practice, however, may need relationships that connect more than two objects - **N-ary**. This gives rise to complexity - how to portray such a relationship on a graph, not to be confused. Conceptual graphs remove this difficulty by presenting each relation in the form of a single node.

In addition to conceptual graphs, there are other versions of semantic networks, this is another basis for classification (**implementation**).

8.4. FEATURES OF SOFTWARE DEVELOPMENT FOR NATURAL LANGUAGE PROCESSING

Based on what was described above it is possible to make some conclusions about features that natural language processing software definitely needs to have. At first

glance a theory of the approaches seems to be enough to produce good software tools for the tasks of NLP. But as the state of the art of such software tools shows, it is not as easy. Development of such software requires some complex features that are described below.

Development of machine translation software

- Software for MT should be developed on a modular basis. As translation process usually consists of several stages.
- Texts to be translated should be analyzed not only on sentence level, but on phrase (collocations or n-grams) and word level as well.
- When using rule-based approaches some detailed information about the grammar rules of source and target languages is very important. This is why MT software will differ for different language pairs. And translation of every pair should be viewed separately, as different programming task.
- Statistical and example based approaches require linguistic corpora available for every language involved and bilingual corpora for language pairs. The larger corpora are, the better the quality of translation is. If there is no corpora or very few of them for language(s) involved in translation, then probably the task of creating some of such structures should be considered. Which is not an easy task itself.
- Hybrid approach, though it may seem the optimal one, is quite dangerous because without proper linguistic modeling it may combine disadvantages of approaches used.

As an example of machine translation software developing we can provide a Kazakh-English machine translation system. The system is developing in Kazakh National University since 2008. It has following characteristics:

- Machine translation from Kazakh into English is done with studying linguistic features of both languages.
- Rule-based approach and Statistical approach to machine translation cannot produce good translation separately, that is why hybrid approach is being used.

- In order to produce translation as close to publication quality as possible MT system has pre-editing of source text.
- As there is no corpora for Kazakh language MT system will have some tools to help to produce monolingual and bilingual corpora.
- MT system has following parts (modules):
 - sentence boundary identification;
 - length check;
 - morphological analyzer;
 - syntactical parser;
 - bilingual dictionary;
 - morphological generator;
 - syntactical generator;
- Morphological analyzer:
 - splits sentences into words;
 - carries out vocabulary control;
 - analyzes words for morphological features;
 - uses regular grammar for parsing;
 - saves unknown words for future adding to a dictionary;
- Syntactical parser
 - performs part-of-speech tagging;
 - outputs scheme of the sentence in source language;
 - uses Link grammar theory;
 - parsing success control;
- Bilingual dictionary:
 - consists of several dictionaries (general, term-base, named entities);
 - contains word pairs and collocation pairs in Kazakh and English;
 - contains morphological and syntactical data that cannot be described in grammars;
- Morphological generator:
 - generates words in corresponding morphological form in target

- language;
- uses regular grammar to generate word forms;
- uses a statistical module for word sense disambiguation;
- Syntactical generator:
 - produces scheme of the sentence in target language;
 - puts translated words and collocations in target language into the scheme;
- Source text and output target text are saved for human revision and future adding to monolingual and bilingual corpora.

Morphological analyzer uses regular grammar for parsing separate words. Which is quite usual usage of that type of grammars. Syntactical parser instead of context-sensitive grammars uses Link Grammar theory for parsing sentences.

Link grammar is a theory of syntax which builds relations between pairs of words, rather than constructing constituents in a tree-like hierarchy. There are two basic parameters in the link grammar: directionality and distance. Here is an example of how this grammar is used in parsing.

Sentence: **Марат жаңа кино көрді.**

Simple link grammar rule describing this type of sentences looks like this:

<subject>: S+;
<adjective>: A+;
<object>: A- & O+;
<verb>: S- & O-;

It means that in this type of sentences <subject> can have 1 link of S type from the right, <object> can have 1 link of A type from the left and 1 link of O type from the right and so on.

Parse tree of the sentence:

```

+-----S-----+
|       +---A---O---+
|       |       |       |
Марат жаңа кино көрді.

```

Corresponding link grammar rule for English translation, that will be produced by syntactical generator:

<subject>: S+;
<verb>: S- & O+;
<determiner>: D+;
<adjective>: A+;
<object>: D- & A+;

Translation: **Marat watched a new movie.**

Parse tree for sentence in English:

```

      +-----O-----+
      |      +---D---+
+---S---+ | +---A---+
|       | | |       |
Marat watched a new movie.

```

This simple example confirms that machine translation is not as trivial task as it seems from the first look.

Development of quality evaluation software for machine translation

Quality evaluation software is not as difficult as MT software. But still it has its own characteristics. Bilingual corpora specific for domain used in translation are needed for program learning and performing the evaluation. Also, because of the algorithms' features described above, supervision of a professional translator when conducting an evaluation is highly recommended.

REFERENCES

- [JUR08] Jurafsky D., Martin J. H.: *SPEECH and LANGUAGE PROCESSING*, Prentice Hall, 2008
- [BOR09] Boretz A.: *AppTek Launches Hybrid Machine Translation Software* SpeechTechMag.com, 2009
- [BAB07] Babych B., Hartley A., Sharoff S.: *Translating from under-resourced languages: comparing direct transfer against pivot translation*. Proceedings of MT Summit XI, 10–14 September 2007
- [PAP02] Papineni K., Roukos S., Ward T. and Zhu W. J.: *BLEU a method for automatic evaluation of machine translation*, ACL-2002: 40th Annual meeting of the Association for Computational Linguistics
- [BAN05] Banerjee S. and Lavie A.: *METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments*. Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics, 2005
- [SLE93] Sleator D., Temperly D.: *Parsing English with a Link Grammar*. Third International Workshop on Parsing Technologies, 1993

Application of MALDI-TOF Mass Spectra Processing and Classification

Mass spectrometry is widely used proteomic technique. It is used for tissue samples processing analysis. Mass spectrometry has widely application in tumour tissues analysis and it gives good opportunities for searching of earlier diagnostic, new medicals and better monitoring of diseases progress and effects of treatment [COM07]. Moreover, it is non-invasive method – usually serum samples are analyzed.

The goal of mass spectrometry is to measure dependencies between m/z (mass to charge) value and ions intensities. Samples are ionized and obtained ions are accelerated by an electric field, they drift through a drift tube and finally they hit in a detector. Ions reach the detector in different time because they have different velocities. The smaller weight single ion has, the faster it is. In this way measures are performed and this type of detector is called TOF (time of flight). Spectrometers counts number of ions which hit the detector in a specified time interval. MALDI-TOF (*Matrix-Assisted Laser Desorption/Ionization*) [POL07] is popular and widely applicable technique in proteomic research. It is typical soft ionization technique using laser energy. Samples are mixed with a highly absorbing matrix. After this process analyte molecules are sputtered and spared [MOR05]. Results are presented as two dimensional matrixes which form mass spectra, where M/Z values are on horizontal axis and ions intensities on vertical one.

Popularity of MALDI-TOF technique in the proteomic research relies in high-resolution measurements in reasonable time. It is also not very expensive method. Proteomic data are of a great importance for the problem of an early-stage cancer detection. However, a problem exists is high-dimensional data sets which need to be processed in a reasonable time.

Important issue is also classification of mass spectrometry data. Classification gives opportunities to differ ill patients from healthy donors or predict positive or negative reaction on the medical treatment. However, classification of mass spectrometry data is hard due to high-dimensionality of data. Such input data-set for classification usually contain several hundreds or even thousands of features. Number of observations are significantly smaller. From the statistical point of view using such number of features is useless. Protein data are usually highly correlated. It is obvious, because such data reflects multiple correlated processes of human organism. However, the great majority of classification methods have assumption that data should be uncorrelated. Researchers need to use reduction and selection techniques. Such methods attempt to find the smallest data sub-set chosen with defined criteria among the whole data set. Too large number of features has an adverse impact on the classification results and it makes the analysis very difficult [MOR05].

9.1. RELATED WORKS

Technology of mass spectrometry is widely used in tasks of proteins/peptides measurement. Correctness and accuracy of those measures is essential because it is used in further biological analysis. Proper peaks identification process should be prefaced with initial preprocessing of raw mass spectrometry data. There are tools, which enable less or more complex mass spectra analysis. Basic steps of preprocessing are: smoothing, baseline correction, normalization, trimming. There are no universal standards assigning proper order of processing those steps. Each proposed method of peaks detection [NOR07] proposes different set and order of preprocessing methods.

Among widely known packages of spectra analysis is commercial application ProteinChip Software [KAT06] using moving average filter for smoothing, convex hull for baseline correction and total ion current normalization. Peaks detection is implemented in two steps: searching with low sensitivity of easy to find peaks and using of high sensitivity searching for smaller peaks. PROcess [XLU05] is other bioinformatics tool available in BioConductor package. Normalization is done by unifying areas under the curve for all spectra from a dataset. Smoothing of local minima with interpolation or regression is used for baseline correction. Peaks detection is performed with searching of local maxima. Local maxima needs also to meet conditions of minimum value of signal to noise ratio and minimum value of area of single peak.

Peaks detection based on local maxima are also implemented in such tools as SpecAlign [13], Cromwell package [MOR05, NOR07] and PrepMS [TIB04], whereas SpecAlign uses also such parameters as minimal height of peak and size of moving window for peaks. MassSpecWavelet [LIN06] is another Bioconductor package, where searching for peaks is performed with undecimated wavelet transformation.

Among popular preprocessing methods there are: moving average [CAR05] and peaks subtraction for baseline correction [PET07] and moving average, wavelet transformation and different types of filter (Savitzky and Golay, Gaussian, Kaiser) [GRO08, REI09] for smoothing.

Among different peaks detection method one can find: comparison of apex and surrounding noise level [SMO07], minimal value of true peak area threshold [VAP98, TIB04, MOR05]. A mean spectrum [MAN07] and peaks clusters [MAN08] is also applied. Peaks were also decomposed with a sum of constituent functions [MCL00]. Dijkstra [ZHA07], Kempka [RAN05] and Noy & Fasulo [DIJ06] proposed also a decomposition with mixture models. Those papers however do not use mixture models in the context of a biomarkers search.

Another issue is a problem with a normalization [POL07] and a redundancy of biological data. There are several big proteomic and biological databases, which offer access to a more or less structured data. However, proteomics is quite a young field of science and data gained by different research centers were devoid of a general

enforced structure. Such structures are created and introduced nowadays, but it is a hard task which needs a lot of time to be accomplished. That is why different databases have the same data labeled with different standards. This makes usage of those data especially difficult in the context of a software development.

9.2. METHODS OF DATA ANALYSIS

Gaussian Mixture Model [KEM04] (GMM) decomposition is used to model spectra and find their peaks. In the single GMM model one peak is represented by a particular Gaussian distribution. A proper determination of GMM components gives a precise spectrum decomposition. Parameters describing peaks may be used in further analysis and biological context determination.

A mixture model is defined as a combination of a finite number of single probability distributions and it is defined with eq. (9.1).

$$f^{mix}(x, \alpha_1, \dots, \alpha_K, p_1, \dots, p_K) = \sum_{k=1}^K \alpha_k f_k(x, p_k) \quad (9.1)$$

where K is the number of mixture components, $p_k, k = 1, 2, \dots, K$ describes mixture parameters including weights $(\alpha_k, k = 1, 2, \dots, K, \sum_{k=1}^K \alpha_k = 1)$.

In a Gaussian distribution a set of parameters in the mixture is composed of means μ_k and standard deviations σ_k . Each mixture component is given also with weight, which represents its share in the whole mixture.

The more complex is a spectrum, the larger is number of mixture components and parameters. Number of components in a typical spectra vary between forty and four hundreds. An Expectation - Maximization (EM) algorithm is one of well known methods of a hidden and unknown parameters estimation. This method is used for solving model of preprocessed spectra. Gaussian distribution was used for spectra modeling because it is suitable for a spectra noise modeling. This makes analysis faster and it enables omitting of earlier smoothing of spectra.

The EM algorithm is an iterative, nonlinear method. It is composed of two steps: an Expectation (E) and a Maximization (M) performed in a loop. The E step (eq. 9.2)

calculates a conditional probability of belonging of sample x_n to k th component. The M step is the most time consuming part of EM and it consists in calculation of new parameters values. Eq. 3 presents a form of M step adjusted to the Gaussian distribution.

$$p(k | x_n, p^{old}) = \frac{\alpha_k^{old} f_k(x_n, p^{old})}{\sum_{k=1}^K \alpha_k^{old} f_k(x_n, p^{old})} \quad (9.2)$$

$$\begin{aligned} \mu_k^{new} &= \frac{\sum_{n=1}^N x_n p(k | x_n, p_{old})}{\sum_{n=1}^N p(k | x_n, p_{old})}, k = 1, 2, \dots, K \\ (\sigma_k^{new})^2 &= \frac{\sum_{n=1}^N (x_n - \mu_k^{new})^2 p(k | x_n, p_{old})}{\sum_{n=1}^N p(k | x_n, p_{old})}, k = 1, 2, \dots, K \\ \alpha_k^{new} &= \frac{\sum_{n=1}^N p(k | x_n, p^{old})}{N} \end{aligned} \quad (9.3)$$

Application of EM algorithm needs to be adjust to two dimensional mass spectrometry data. A weighted version of EM [PET07, GRO08], was developed. The second dimension of mass spectrometry data represents intensities, which may be described as repeats of corresponding m/z values. Therefore, appropriate multiplications needs to be performed. A weighted version of the EM remains the E step unchanged (eq. 9.2), because the value of probability that a sample x_n belongs to k th component does not change. So, independently of number of the x_n repetition in a vector, calculated probability is always the same. The calculations of the M step need to be performed in a slightly different way. Proper formulas are presented in eq. 9.4.

$$\begin{aligned} \mu_k^{new} &= \frac{\sum_{n=1}^N x_n y_n p(k | x_n, p_{old})}{\sum_{n=1}^N p(k | x_n, p_{old}) y_n}, k = 1, 2, \dots, K \\ (\sigma_k^{new})^2 &= \frac{\sum_{n=1}^N (x_n - \mu_k^{new})^2 p(k | x_n, p_{old})}{\sum_{n=1}^N p(k | x_n, p_{old}) y_n}, k = 1, 2, \dots, K \\ \alpha_k^{new} &= \frac{\sum_{n=1}^N p(k | x_n, p^{old}) y_n}{N} \end{aligned} \quad (9.4)$$

The EM algorithm needs for initial values and stop criterion. It is essential to choose proper initial values, because it reduces time needed for calculations and probability of local maximum problem occurrence. Important issue is also to estimate number of mixture components, because it needs to be known before starting the algorithm. The choice of convergence criterion and accuracy is equally important. Criteria may be based on consecutive values of parameters or likelihood function and it may be calculated with a chosen distance.

9.3. CLASSIFICATION ISSUES

On the basis of the single learning set multiple different classifiers could be built. The best classifier would be chosen on the basis of the number of misclassifications. However, truly classification probabilities are unknown and they might be estimated from a validation probe. The validation probe is a random sample, independent of the learning probe. Membership to classes of its objects are unknown. Misclassification probability of specific classifier is estimated on the basis of the validation probe. Classifier evaluation should be done using observations independent of those from the learning probe. In other cases the classifier is biased. In practice, the usually chosen proportion is the division: 50 % on the learning probe and 25 % each for the validation and test probes [COM07]. However, the division depends on the specificity of the data set.

One of widely used classifiers in proteomic research is The Support Vectors Machines (SVM) [N.Vapnik [NOR07, POL07, PLE10]]. SVM classification uses appropriately designated discriminant hyperplane. Searching of such hyperplane needs Mercer theorem and optimization of quadratic objective function with linear restrictions. In case of fully separable learning subsets, the SVM idea is to find two parallel hyperplanes, which delimit the wider area do not containing any probe elements. So, hyperplanes need to be based on some of the probe elements -they are called support vectors. The discriminant hyperplane is put in the middle of the resultant area.

In case of not linearly separated learning sub-sets the penalty is introduced. The best separation can be obtained for higher dimension space.

The form of the SVM rule is presented in (eq. 9.5).

$$f(x) = \text{sgn}\left(\sum_{\text{sup. vect.}} y_i \alpha_i^0 (x_i x) + b^0\right) \quad (9.5)$$

where: α are Lagrange's coefficients and b is a constant value. For inseparable classes the additional restrictions take the form of (eq. 9.6).

$$\begin{aligned} x_i w + b &\geq 1 - \xi_i, y_i = 1 \\ x_i w + b &\geq -1 + \xi_i, y_i = -1 \end{aligned} \quad (9.6)$$

where: ξ_i is a constant value $\xi_i \geq 0$.

The more complicated classification problems with larger number of features may be solved with kernel functions. Such construction enables to obtain non-linear shapes of discriminant hyperplanes. The SVM rule with kernel $K(x_i x)$ takes the form of (eq. 9.7).

$$f(x) = \text{sgn}\left(\sum_{\text{sup. vect.}} y_i \alpha_i^0 K(x_i x) + b^0\right) \quad (9.7)$$

One of the most popular kernel function is radial kernel (eq. 9.8).

$$K(x_i x') = \exp(-\|x - x'\|^2 / c) \quad (9.8)$$

A classifier cannot be build on thousands of features. The existence of the curse of dimensionality [MAO00] proves, that the complexity of the classifier has an effect on the classification quality. The more complex classifier is, the higher should be the proportion between number of observation and number of features [WOL96]. Dimension reduction techniques needs to be used. There are two types of those methods:

1. features extraction – data are undergone transformation – new data set is obtained
2. features selection – sub-set of the most optimal data is chosen

One of commonly known classification and features extraction methods is Partial Least Squares (PLS) [WOL96], which is a special case of Principal Component

Analysis (PCA) [VAP98]. Features selection in PLS method is performed with use of two dimension (X and Y) data. So it enables using structure of the hole learning data set.

The idea of PLS method is to find latent vectors. Latent vectors are calculated from the covariance between X and Y. The decomposition of X and Y is done to low-dimensional space of hidden variables. Independent variables X are decomposed according (eq. 9.9).

$$X = TP^T + E_x \quad (9.9)$$

where $T^T T = I$, I - identity matrix, T - score matrix and P - loading matrix. A product of T and P gives good estimation of X matrix.

Dependent variables Y are decomposed as (eq. 9.10).

$$Y = UQ^T + E_y \quad (9.10)$$

The final model of PLS describing $Y \Leftrightarrow X$ regression is (eq. 9.11).

$$Y = X(PB_1Q^T) + E = XB + E \quad (9.11)$$

One of well known features selection method SVN-RFE (Support Vector Machine Recursive Feature Elimination) [BAR02], which uses propagation backward. The procedure starts with full range of input features and features are ranged and successively removed. Only one feature is removed in a time. SVM weights coefficients are used as a rang criterion, so SVM-RFE method is closely related to SVM classification. In SVM-RFE procedure SVM classification might be formulated as in (eq. 9.12).

$$\begin{aligned} \min_{w,b,\xi_i} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i^2 \\ & y_i(wz_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \end{aligned} \quad (9.12)$$

Eq. 9.10 is solved with (eq. 9.13).

$$\min_{\alpha_i} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \tilde{k}(x_i, x_j) - \sum_{i=1}^n \alpha_i \quad (9.13)$$

where $\tilde{k}(x_i, x_j)$ is a kernel function.

The SVM-RFE objective function is (eq. 9.14).

$$J = \frac{1}{2} \|w\|^2 \quad (9.14)$$

Very common technique of feature selection is T test. The most significant features according the T test are chosen. For each feature a T test range is calculated with (eq. 9.15).

$$c_i = \frac{|\mu_i^+ - \mu_i^-|}{\sqrt{\frac{(\sigma_i^+)^2}{n^+} + \frac{(\sigma_i^-)^2}{n^-}}} \quad (9.15)$$

where: μ_i^+, μ_i^- denote the mean values for i^{th} feature calculated for respectively positive and negative samples. Similarly σ_i^+, σ_i^- denote standard deviations and n^+, n^- denote numbers of positives and negatives learning samples.

The T statistics treats all feature as independent. This assumption is usually not met. However, T test is successfully used for protein data classification.

9.4. SOFTWARE PROJECT

The mass spectra analysis software has been divided into three parts. The first one is dedicated to preprocessing and decomposition of single or a multiple spectra. The second part enables performing a biological analysis of achieved results. The third part is classification module. Modules are intended to work independently. They might exchange data or be used as two separated programs (fig. 9.1).

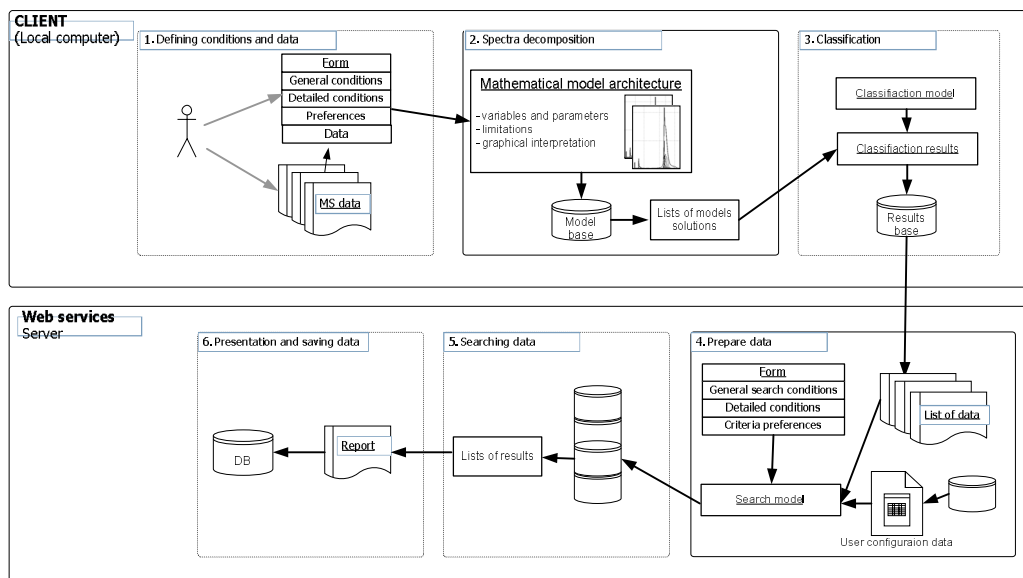


Fig. 9.1. A flowchart presenting data processing in the system

Module I - spectra analysis software

The spectra analysis module is standalone application created in Matlab environment. It is composed of a main decomposition module and a few associated sub-modules [PLE10]. Fig. 9.2 presents the first level of a data flow diagram of the spectra analysis module.

The core functionality of this module is the spectra decomposition with the GMM mixture model. Two analysis scenarios are possible: processing all spectra sequentially, one by one or using the mean spectrum. Processing many spectra this way lengthens their calculations time, especially when the defined accuracy is high. In such cases calculations using the mean spectrum might be preformed.

Using the mean spectrum is faster than processing all spectra sequentially. However, this method needs two more preprocessing steps: interpolation, which standardizes m/z values (X axis) and normalization to TIC (Total Ion Current) used for spectra adjustment. A mean spectrum is calculated with eq. 9.16:

$$y_{-sr_i} = \frac{\sum_{j=1}^K y_{ij}}{K}, j = 1, 2, \dots, N \quad (9.16)$$

where: y_{sr_i} indicates i^{th} element in the output vector of mean values. y_{sr_i} corresponds to i^{th} element x_i . A X axis, after interpolation process, is uniformed for all analyzed spectra. y_{ij} represents the i^{th} element of the intensities vector of j^{th} spectrum. K is the number of spectra in the data-set and N is the uniformed size of all spectra.

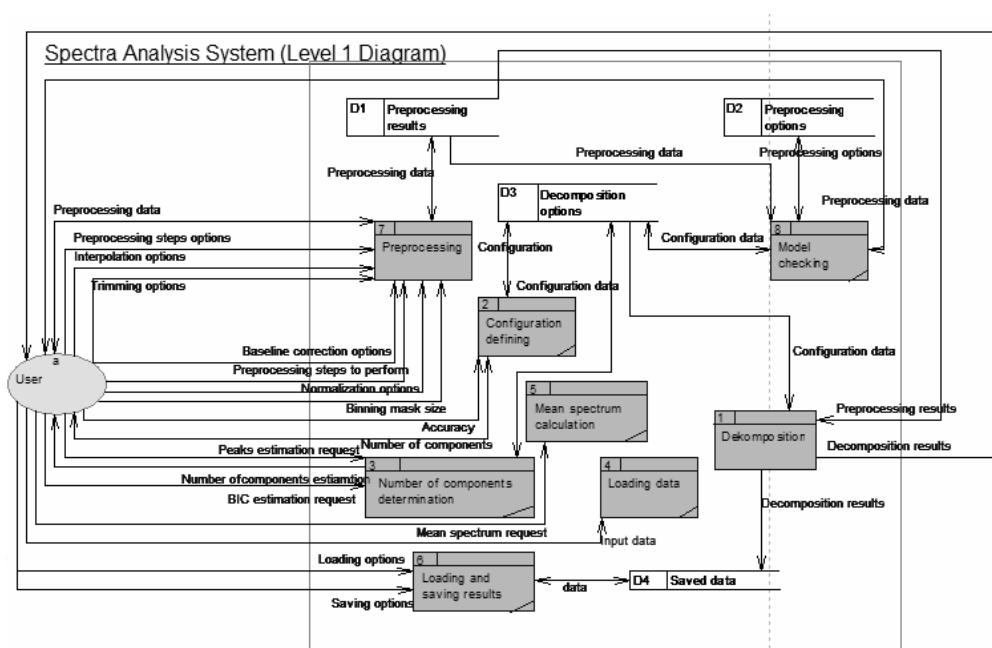


Fig. 9.2. DFD diagram of the spectra analysis module [PLE10]

A spectrum modeled with the GMM are decomposed with the EM algorithm. In case of the mean spectrum obtained values of means (μ_k) and standard deviations (σ_k) remain constant for every single spectrum in the set. Only weights needs to be obtained separately for each spectrum and they are estimated with Least Squares methods.

Decomposition can be done after few steps of a data preparation which are processed by additional modules. Loading data module is responsible for loading textual data and its preparation to analysis. Important submodule is configuration. Settings involve a specification of:

- a special type of decomposition (with or without mean spectrum);
- a source and range of the EM initial data by randomization or with support of standard peaks determination algorithm based on local maxima and minima. Using of this method helps to find initial values in the more probable areas than values found by randomization;
- the accuracy of calculations used in stop criterion;
- preprocessing options (binning mask, baseline correction window, normalization and interpolation options, trimming and propriety checking);
- the number of components, which may be inputted directly or estimated. Estimation may be done with standard peaks determination algorithm based on local minima and maxima. Longer, but giving clear and precise results is using BIC simulation.

Bayesian Information Criterion (BIC) [23] is criterion based on a value of a likelihood function. According BIC, the optimal number of criterion should maximizes eq. (9.17)

$$-2 \cdot \ln p(x | k) \approx BIC = -2 \cdot \ln L + k \ln(n) \quad (9.17)$$

where: x is a sample of size n and k is the number of estimation parameters.

$p(x | k)$ is the likelihood of the observed data. L is the maximized value of the likelihood function for the estimated model. BIC calculations last longer, because decomposition needs to be repeated many time for each of considered number of components.

The model checking module is designed to validate and control system settings before or at the beginning of decomposition. It enables viewing results of preprocessing performed by the Preprocessing module and checks possibility of use of defined settings. In particular, it enables checking baseline correction results and gives information about initial distribution of mixture model components.

The mean spectrum calculation module enables checking of the mean spectrum shape after defined preprocessing steps.

Module II - classification module

The classification analysis might be performed using three dimension reduction techniques – SVM-RFE, PLs and ttest. The module was implemented in Matlab. The main classification is done with SVM classifier linear or with radial kernel. Classification and reduction performance is done for defined values of SVM parameters and number of selected features. The SVM parameters are: value of box constraints (C) for the soft margin and the scaling factor (sigma). The default values are the most accurate values obtained in earlier simulation study. The classifier was learned at the breast cancer patients data and healthy donors. Data of classification module might be transferred from analysis module or they can be loaded manually. Obtained results might be also transferred to the biological context module.

Module III - biological context module

Results of the decomposition may be saved and transferred to the biological context module. Data might be also entered manually. The analysis consists of a four-steps path, whereas each step enables displaying data with a different level of biological context. Levels need to be achieved sequentially. The user may stop at any level and may go back to all previously generated levels. Details of simulation are presented in Fig. 9.3.

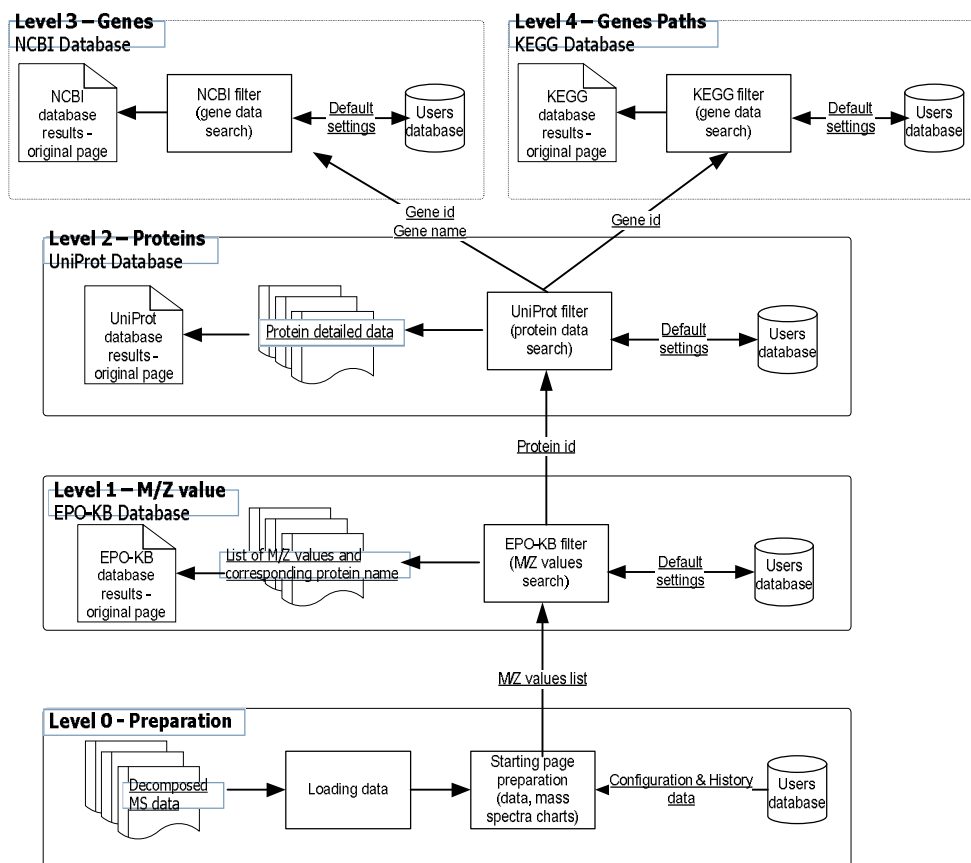


Fig. 9.3. Flowchart of the biological context module

Loading data might be processed in the Level0 of the module. Data can be transferred from the spectra analysis software or choosing particular MS from a disk data. Level0 is also place for specifying detailed search criteria, like: species, the MS platform, the possibility of double and triple charges. These criteria may be inserted manually or loaded from the user profile. Registered users can save results, view the history, add comments and send data to the other users.

The Level1 is the level of m/z values. All data chosen in Level1 are matched for further analysis. EPO-KB (Empirical Proteomic Ontology Knowledge Base) database [LUS08] is the main service of this level. The m/z values are sent to EPO-KB. After

applying defined criteria m/z names of the found proteins are returned with a defined percentage tolerance. The results are presented for one spectrum at the time with references for further analysis and links to original results in the EPO-KB service. The original results include particular detailed information about a specific protein.

The Level2 is a protein level, which is available after choosing proteins in the Level1. Detailed information about specific protein is obtained from an UniProt database [UNI10]. UniProt database enables filtering by such searching criteria as a name of the protein and a species type. This level contains detailed information about proteins, such as entry name, status of reviewing process, organism, gene names and identifiers, features, GO annotations. Links to original UniProt pages of the specific proteins are also available.

The Level3 of the module is a genes level. The Level2 enables displaying detailed information about genes coding a particular protein chosen at an earlier level. Those information are presented through a NCBI (The National Center for a Biotechnology Information) [WHE09] service. Searching is based on a gene name or a gene identifier. The gene identifier enables to return more accurate and precise information about a particular gene (role, status, lineage and related data). The Level4 of the module gives possibility to get a gene pathways data from the KEGG (Kyoto Encyclopedia of Genes and Genomes) [KAN08] database. User can find here such information as details of a pathway, a structure, sequences, references to other databases.

9.5. SUMMARY

The presented project is a comprehensive bioinformatics tool enabling spectra preprocessing and analyzing. The use of the Gaussian Mixture Model decomposition enables a particular work with different types of spectra and options settings support minimizes risk of improper parameters selection. Modules of classification and biological context makes the application useful for different types of users. Integration with biological databases gives access to significant information in one place and

fasters work. Classification module is also available. It has implemented three dimension reduction techniques with adjusted parameters values.

REFERENCES

- [BAR02] Barnhill S., Vapnik V., Guyon I., Weston J.: *Gene selection for cancer classification using support vector machines*. Machine Learning, 46: pp. 389–422, 2002
- [CAR05] Cartwright H.M., Wong J.W., Cagney G.: *Specalign - processing and alignment of mass spectra datasets*. Bioinformatics, 21:2088–2090, 2005
- [COM07] Coombes K., Baggerly K., Morris J.: *Pre-processing mass spectrometry data*. In: Fundamentals of Data Mining in Genomics and Proteomics, W Dubitzky, M Granzow, and D Berrar, eds. Kluwer, 79-99. 2007, Boston.
- [DIJ06] Dijkstra M., Roelofsen H., Vonk R. J., Jansen R. C.: *Peak quantification in surface-enhanced laser desorption/ionization by using mixture models*. Proteomics, vol. 6., pp. 5106–5116 2006
- [GRO08] Gropl C., Hildebrandt-R.Hussong A., Lange E., Pfeifer N., Schulz-Trieglaff O., Zerck A., Reinert K., Kohlbacher O., Sturm M., Bertsch A.: *Openms – an open-source software framework for mass spectrometry*. Bioinformatics, 9:163, 2008
- [KAN08] Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y.: *KEGG for linking genomes to life and the environment*. Nucleic Acids Res. 36, D480-D484 2008
- [KAT06] Katajamaa M., Miettinen J., Oresic M.: *MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data*. Bioinformatics, 22: 634-636, 2006
- [KEM04] Kempka M., Sjodahl J., Bjork A., Roeraade J.: *Improved method for peak picking in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Rapid Commun. Mass Spectrom., vol. 18, pp. 1208–1212 2004

- [LIN06] Lin S., Du P., Kibbe W.: *Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching*. Genome analysis, 22:2059-2065, 2006
- [LUS08] Lustgarten, J.L., et al.: *EPO-KB: a searchable knowledge base of biomarker to protein links*. Bioinformatics 24(11):1418-1419 2008
- [MAN07] Mantini D., et al.: *LIMPIC: a computational method for the separation of protein signals from noise*. BMC Bioinformatics 8:101 2007
- [MAN08] Mantini D., et al.: *Independent component analysis for the extraction of reliable protein signal profiles from MALDI-TOF mass spectra*. Bioinformatics, 24: 63 70 2008
- [MAO00] Mao J., Jain A.K., Duin R.P.W.. *Statistical pattern recognition: a review*. IEEE Trans. PAMI, 22(1): pp. 4–37, 2000
- [MCL00] McLachlan G. J., Peel W.: *Finite Mixture Distributions*, Wiley, 2000
- [MOR05] Morris J., Coombes K., Kooman J., Baggerly K., Kobayashi R.: *Feature extraction and quantification for mass spectrometry data in biomedical applications using the mean spectrum*. Bioinformatics, 21(9): 1764-1775. 2005
- [NOR07] Norris J., Cornett D., Mobley J., Anderson M., Seeley E., Chaurand P, Caprioli R.: *Processing MALDI mass spectra to improve mass spectral direct tissue analysis*. National institutes of health. 2007, USA
- [NOY07] Noy K., Fasulo D.: *Improved model-based, platform-independent feature extraction for mass spectrometry*. Bioinformatics, vol. 23, no. 19, pp. 2528–2535 2007
- [PET07] Petrucci F., Pieragostino D., Del Boccio P., Di Nicola M., Di Ilio C., Federici P. G.: Sacchetta S. Comani Mantini, D. and A. Urbani. *Limpic: a computational method for the separation of protein signals from noise*. BMC Bioinformatics, 8:101, 2007
- [PLE10] Plechawska M.: *Application of gaussian mixture model and proteomic databases in the mass spectra analysis – architecture of software of comprehensive mass spectrometry data processing*. Studia Informatica. Proceedings of BDAS 2010

- [POL07] Polanski A., Polanska J., Pietrowska M., Rzeszowska J., Stobiecki M., Tarnawski R., Skladowski K., Widlak P.: *Application of the Gaussian mixture model to proteomic MALDI-ToF mass spectra*. Journal of Computational Biology. Gliwice 2007
- [RAN05] Randolph T., et al.: *Quantifying peptide signal in MALDI-TOF mass spectrometry data*. Molecular & cellular proteomics : MCP, 4(12):1990-9 2005
- [REI09] Reinert K., Kohlbacher O.: *Openms and topp: Open source software for lc-ms data analysis*. Proteome Bioinformatics, 604:201–211, 2009
- [SMO07] Smolka A.J., Morris J.S, Coombes K.R., Baggerly K.A., Almeida J.S., Karpievitch Y.V., Hill E.G.. PrepMS: *Tof ms data graphical preprocessing tool*. Bioinformatics, 23:264–2265, 2007
- [TIB04] Tibshirani R., Hastie T., Narasimhan B., Soltys S., Shi G., Koong A., Le QT.: *Sample classification from protein mass spectrometry, by 'peak probability contrasts*. Bioinformatics 2004, 20:3034-3044
- [UNI10] UniProt Consortium: *The Universal Protein Resource (UniProt) in 2010*. Nucleic Acids Res. D142-D148 2010
- [VAP98] Vapnik V.N.: *Statistical Learning Theory*. Wiley, 1998
- [WHE09] Wheeler DL et al.: *Database resources of the National Center for Biotechnology Information*. Nuc. Acids Res. 37 D5-D15 2009
- [WOL96] Wold H.: *Estimation of principal components and related models by iterative least squares*. Multivariate Analysis, New York: Academic Press pp. 391-420, 1996
- [XLU05] X.Lu Q.Shi J.D.Iglehart-L.Harris A.Miron X.Li, R.Gentleman. *Seldi-tof mass spectrometry protein data*. Statistics for Biology and Health, Part I, Bioinformatics and computational biology solutions using R and Bioconductor, pages 91–109, 2005
- [ZHA07] Zhang S.Q., et al.: *Peak detection with chemical noise removal using Short-Time FFT for a kind of MALDI Data*. Proceedings of OSB 2007, Lecture Notes in Operations Research, 7, 222-231 2007.

Application of Principal Component Analysis and Projection to Latent Structures for Modeling Processes of Aging and Destruction of Polymeric Materials

Nowadays chemometrics is a universally recognized tool within analytical chemistry and biology. However, beyond this area the Multivariate Approach (MVA) is not as widely applicable as it actually deserves. Multivariate Statistical Process Control (MSPC) is the only exception that proves the rule. Even so, the author believe that MVA could be of a large merit being applied in engineering, in general, and in the optimization of the material formulations, in particular. Recent studies of the Latent Variables (LV) based optimization performed by J. MacGregor [YAC04, MUT06] and A. Hoskuldsson [POM06, HOS07] have demonstrated a great potential of such an approach.

10.1. MULTIVARIATE DATA ANALYSIS

Multivariate Data Analysis is used for a number of distinct, different purposes. All purposes can be divided into three main groups [ESB00]:

- Data description.
- Discrimination and classification.
- Regression and prediction.

The complexity of objects and problems in civil engineering also requires modern MVA for data analysis. An engineer or material scientist is increasingly faced with a need to use mathematical and statistical methods in everyday practice. MVA in general and chemometric methods in particular may essentially enhance the capability of newer, mostly multicomponent, methods.

Nowadays, chemometric techniques are applied in various areas of civil engineering, such as:

- monitoring of historical buildings with non-intrusive measurements, image analysis,
- texture analysis, for example in estimation of heterogeneity in concrete slabs,
- ecological analysis of pollution,
- analysis of aging and degradation of various materials,
- quantitative structure – property relationship.

The texture analysis is the investigation of aging and degradation of PVC films with the help of soft models.

These are various types of constructions. In all these buildings a soft composite material is applied.

This material is a composite with a textile reinforcing basis of high-strength synthetic fiber and a film covering (matrix) of elastomer or thermoplastic. This second layer serves for fixing and protection of a reinforcing basis against influence of atmospheric factors. This layer also provides tightness and air impermeability to a material.

10.2. THE MATERIAL DURABILITY RESEARCH

The material durability is firstly determined by the durability of the polymer matrix. Therefore we experimented with the film samples made of the matrix material.

Our goal of research work was to understand whether the pigment structure influences on the polymer durability.

There are the following problems in prediction of polymer materials durability:

1. Degradation processes take place at various structural levels;
2. There is no general theory of polymer aging and destruction;
3. The traditional kinetic approach fails in description of aging phenomena including both chemical and physical ways of degradation.

Our problem is multi-dimensional and multi-criterial. So we have to apply the soft models for prediction of polymer materials durability.

Studying the aging of various materials it is necessary to trace the kinetics of material properties. These properties are determined by the technical demands in normative documents. On the other hand,

- developing the practical methods for durability prediction it is necessary to choose such characteristics that clearly represent the relationships between the structural changes and material quality,
- certainly, it would be nice to get these characteristics without complex methods of structural analysis,
- basically, it is impossible to single out a parameter that unequivocally represents both physical and chemical processes in polymeric materials at their aging.

Our goal is to construct a model for description of material aging. For this purposes the dependence of material properties upon matrix structure is studied. The following physico-mechanical properties are measured and used for investigation:

- durability σ and elongation at break ϵ ,
- the initial module of elasticity E ,
- creep at ageing Δ ,
- color ΔE ,
- shine,
- transparency α ,
- limiting wetting angle θ .

Color was measured by means of colorimetric system CIE Lab.

This is three-dimensional color space with axes L, a and b. L characterizes brightness of color and it changes from 0 (black) up to 100 % (white). Other two

coordinates, a and b, represent the color tones. Index «a» defines the color position on the color wheel from green up to red, and index «b» defines another color component within a range from dark blue up to yellow.

The samples of various thickness and color were tested in the climatic chamber, made by the author. The following aging conditions were simulated: UV sun radiation, rain, temperature and tensile stress.

The orange pigment changing the color protects polymer from photodestruction. Blue the pigment under influence of atmospheric factors keeps the color. However acting as the sensitizer destroys polymer.

10.3. RESEARCH RESULTS

These data (Fig. 10.1. and Fig. 10.2) were explored by principal component analysis and the regression models were constructed.

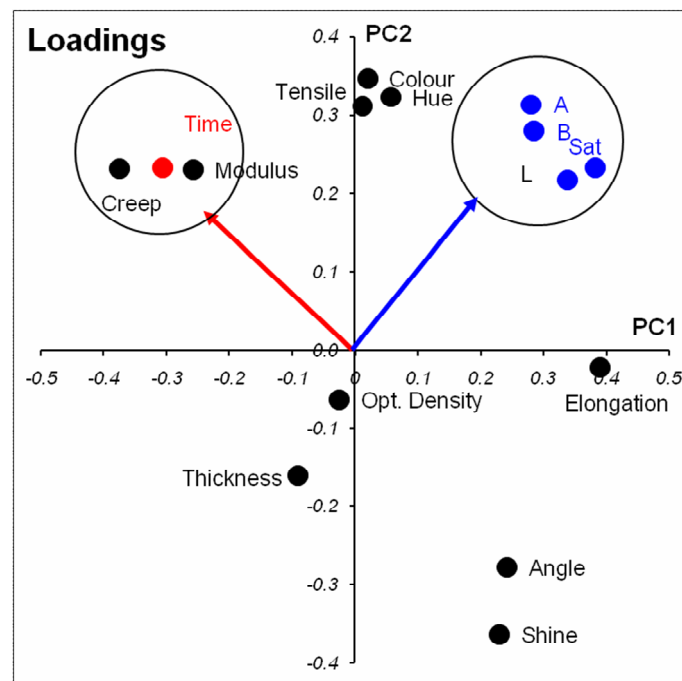


Fig 10.1. Loading plot (PC1 vs. PC2)

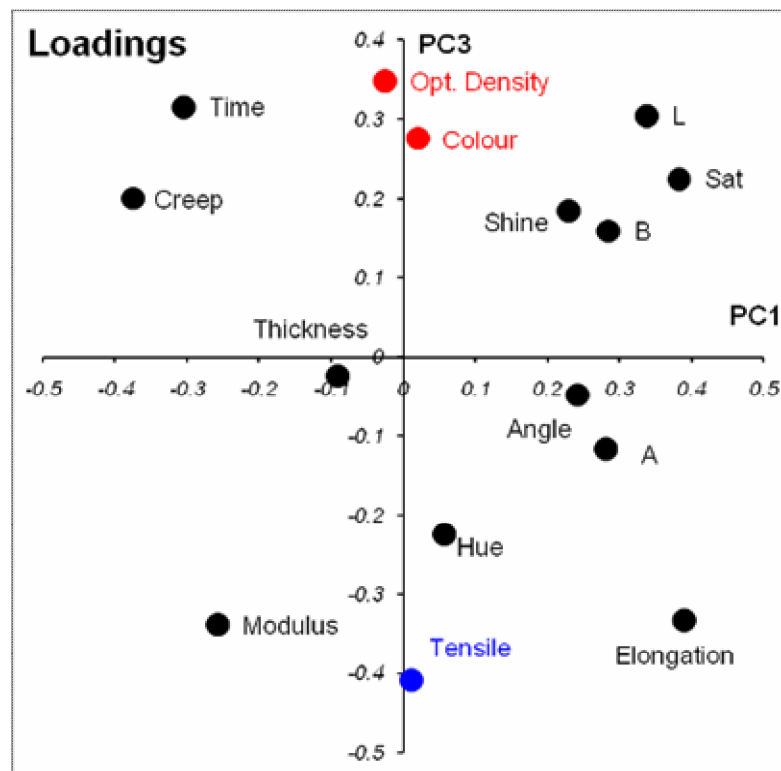


Fig. 10.2. Loading plot (PC1 vs. PC3)

Loading plot (PC1 vs. PC2) shows that two groups of variables equally contribute to the span of principal component space (Fig. 10.1).

The first group of variables comprises parameters of color in the colorimetric to system CIE Lab, and the second group of variables is directly connected with aging process. Time of aging and the sample color are two orthogonal independent factors. They are the most essential variables in the model.

From the loading plot it is also seen, that shine negatively correlates with durability. Reduction of shine signals about the surface defects which are connected with the material durability. The shine and limiting wetting angle are positively correlated. This can be seen in the loading plot where these variables are located closely. In the loading plot (PC1 vs. PC2) the optical density has insignificant influence

on the material durability. At the same time in PC1-PC3 plot (Fig.10.2) an essential negative correlation between these two properties is clearly seen.

Moreover, the score plot (Fig.10.3) shows the separation of samples into two groups. First group includes dark-blue-green-blue, and the second group consists of yellow-orange. Such pattern signalize that the material color strongly influences on the properties.

In other words, changes in material durability are influenced both by the surface phenomena, and the structural changes.

Therefore to construct a model that represents the dependence between the mechanical and physical properties, it is necessary to consider not only the pair variable correlations but also take into account the interference of the most varied factors. This can be done applying the PLS method.

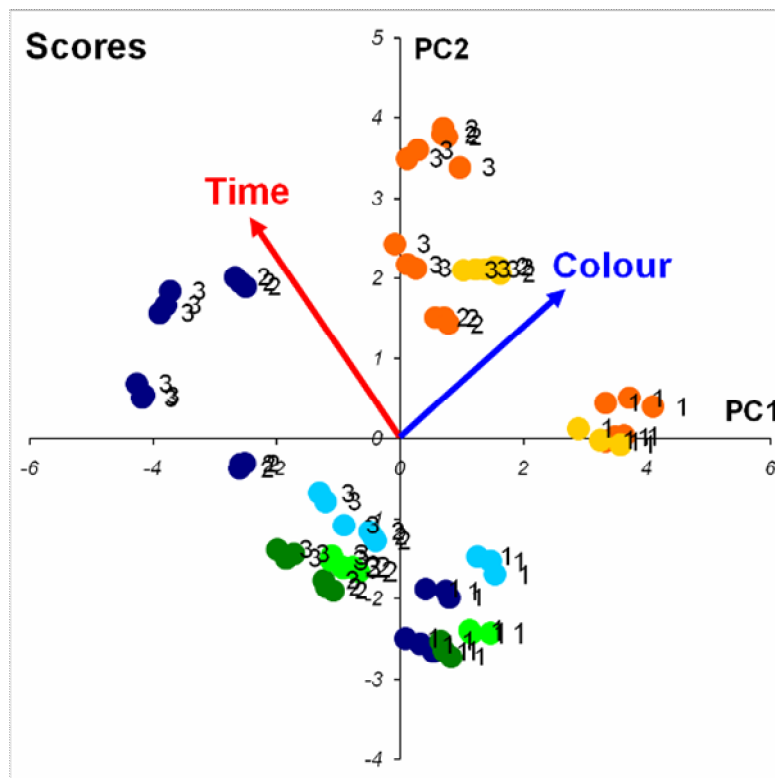


Fig. 10.3. PCA Scores

The goal of regression modeling is prediction of material operation quality. The following variables comprise the X-block (predictors): Conditions of aging time of ageing, thickness (Fig.10.4). In X block are the optical properties color features; shine; limiting wetting angle; optical density; Y-block - responses consists of mechanical properties: tensile; the module of the elasticity; elongation; creep.

Predictor matrix X				Response matrix Y			
Aging conditions				Mechanical properties			
Time	Thickness	Loading	Intensity	Tensile	Modulus	Elongation	Creep
	X1				Y training		
Optical properties							
Color	Shine	Angle	Opt. Density				
					Y new		
	X2						

Fig 10.4. Predictor and response matrixes (X, Y)

Test validation. All samples of light blue material now are in the test set.

A new model for a shorten dataset was established (Fig. 10.5). These are the results of calibration and these are for prediction. The similar results for the elongation at break. Thus, the established PLS models allow us to predict the mechanical characteristics using physical properties of a material.

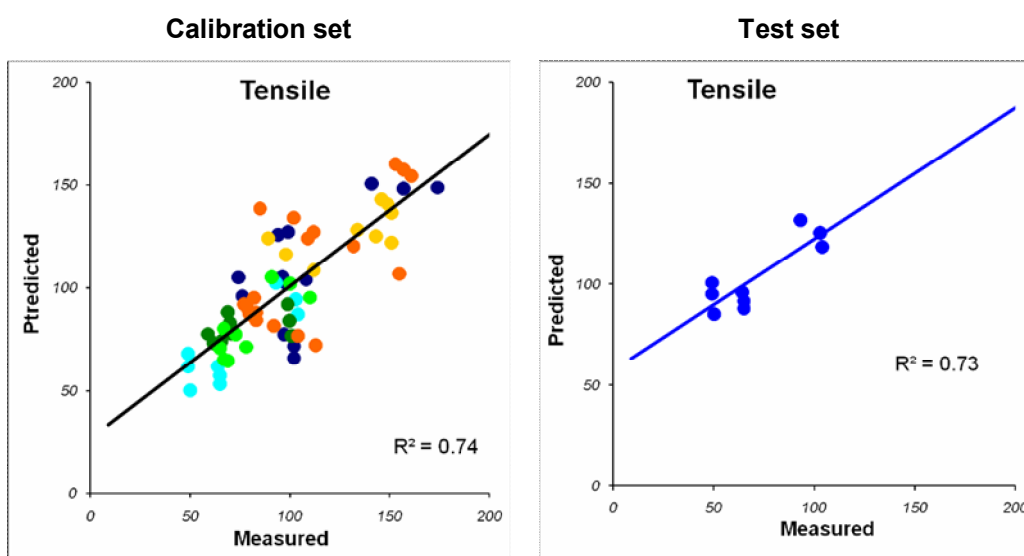














Fig. 10.5. Tensile strength prediction

A similar procedure was conducted to predict the changes in color of a material. The table 10.1 shows the changes in color of light green samples after 500 and 1000 hours of aging in the climatic chamber.

Table 10.1. Changes in color of light green samples after 500 and 1000 hours of aging in the climatic chamber

Sample	Measured				Predicted			
	L	A	B	Color	L	A	B	Color
GL21	43	-41	33		45	-32	22	
GL22	44	-42	33		46	-31	24	
GL23	44	-41	31		46	-33	21	
GL31	43	-41	33		46	-30	30	
GL32	41	-40	32		46	-30	27	
GL33	42	-41	32		45	-29	29	

Analyzing the PLS regression coefficients in Fig.6 we can state that a mechanical property cannot be predicted using only one physical factor. In each case, the result is a combination of several physical factors.

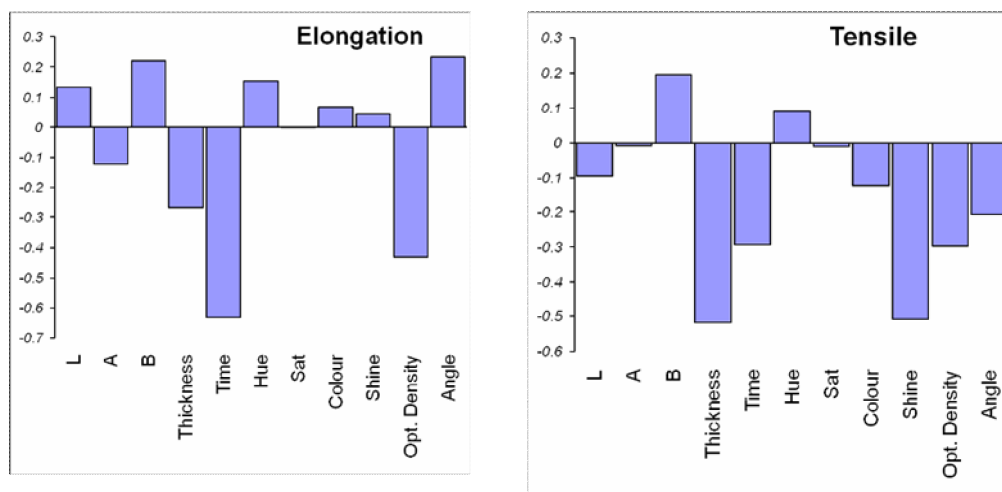


Fig.10.6. Regression coefficients Bw

10.4. CONCLUSIONS

Multivariate Approach enables us to analyze jointly the material properties in the course of its aging. The presented approach substitutes the mechanical testing by a rapid and non-destructive measurement of such physical properties as color, shine and others.

This is just an example of chemometric application. I am sure that the soft modeling approach will be of benefit in various areas of civil engineering.

REFERENCES

- [ESB00] Esbensen K.: *Multivariate Data Analysis – in Practice*. An introduction to Multivariate Data Analysis and Experimental design, CAMO ASA, 2000
- [HOS07] Hoskuldsson A., Rodionova O.Ye., Pomerantsev A.L.: *Chemom. Intell. Lab. Syst.*, v. 88, p.84–99, 2007
- [MUT06] Muteki K., MacGregor J.F., Ueda T.: *Ind. Eng. Chem. Res.*, v.45, p. 4653–4660, 2006
- [POM06] Pomerantsev A.L., Rodionova O.Ye., Hoskuldsson A.: *Chemom. Intell. Lab. Syst.*, v.81, p. 165–179, 2006
- [YAC04] Yacoub F., MacGregor J.F., *Chemom. Intell. Lab. Syst.*, v.70, p. 63–74, 2004

Creation of the User-Oriented Requirements Specification of the Data Intensive Web Application Basing on WebML

Nowadays application development concentrates on the web. The growing need for the new web based applications causes a demand for the methods of rapid web-app development and for the professionals knowing how to implement them. Consumers and developers expect the fast realization of their requests while keeping in mind application's security, scalability, usability, efficiency etc.

WebML is described by its developers as "a visual notation for specifying the content, composition, and navigation features of hypertext applications, building on ER and UML" [CER03]. It can be seen as a high level visual description of a MVC organized web application.

Besides requirements specification WebML consists of three models:

- Data Model,
- Hypertext Model,
- Presentation Model.

Data model describes application data design and is similar to a standard ERD model used in conceptual database design.

Hypertext model concentrates on a composition and navigation of the web-app. It arranges application into views containing areas and pages. Pages contain units responsible for the data presentation. The business logic is projected by the operation units placed either within or outside the pages. The set of units representing the

commonly used operations is made available. Non-standard operation can be introduced by defining the custom units. Pages and units are connected by the links representing information flow between them. Links can carry the activation and be rendered on the page, or be in so called “transport” mode, representing the inner data flows inside application. The hypertext model is independent from the data model, however it uses the names of the entities and attributes described by it [CER03].

The third layer is the presentation model describing the layout of the data presentation units on the pages, the pages layout and the layout of the whole application. It imports the standard DOM/CSS design and is partially left for the developers consideration making it a bit fuzzy, on the other hand leaving the freedom to use the presentation technique of developers choice.

All of described models (excluding requirements specification) are implemented by the WebRatio. IDE out of the box provides: designer tool, Apache Derby Database, Apache Tomcat (Application Server). It transfers WebML concepts to the real web application development and enables automatic application generation basing on the WebML model aided by a few own refinements. Resulting application is output in a JSP code and follows Apache Struts MVC pattern.

WebRatio handles the construction of a WebML Data model as well as the connection and creation of the corresponding database via Hibernate. It also allows to import an existing database into the model. The Hypertext and Presentation model is unified into the Site or the Service view. Both of these allow to visually construct a site Hypertext model starting from views, through areas and pages, ending with units, their attributes and connections. The Presentation model is handled by layout definitions of pages, as well as single units. The units organization on the page is handled by the Grid feature, resulting in the old way, table organized web page. It is also possible to use layers layout but it requires to manually setup needed areas [CER03].

11.1. PROCESS OF THE REQUIREMENTS SPECIFICATION

During requirements specification, information about expected functionality and purpose of the application is acquired and analyzed. Basing on specified functional and nonfunctional requirements, easy to understand and percept documents are produced. It is especially important for people without appropriate knowledge and technical skills (the supervisors and software purchasers). They can monitor the effects of developers' work and decide about the shape of the final product.

The process of requirements specification can be divided into two phases: collection of information and its analysis. The goal of the first one is to identify the purpose of the application. During this phase interviews with client's representatives are held and existing documentation is read. The final product of this phase is the information about actors, application background, application functionality and nonfunctional requirements [CER03].

Data collected in the first phase is analyzed, effecting in semiformal specification containing [CER03]:

1. The list of user groups that will access the application, their preliminary hierarchy and privileges.
2. The most important use cases showing the interactions between application and its users.
3. The list of the most relevant information objects in the application domain.
4. Informal specification of the areas of the application that allows to accomplish identified use cases.
5. Preliminary look and feel specification, including mockups.
6. Nonfunctional requirements of the application.

WebML was intended to cooperate with existing methodologies of developing data intensive web applications. It allows the usage of proven methods and tools like diagrams in UML and other notations. It also systematize some areas and add new elements of specification, to make it as much user-oriented as it could be.

This article concentrates on the new and modified issues in this area, that are presented on the exemplary data intensive web application called “Eagle 7”, which supports the work of educational institution.

11.2. SPECIFICATION OF THE USER GROUPS

The exemplary application can be accessed by unregistered and registered users. The latter ones are divided into administrators, students and teachers. To make specification clearer and more complete, designer can extend it by a diagram showing user groups hierarchy (Fig.11.1). Hierarchy adapted by WebML is called IS-A and does not allow multiple inheritance.

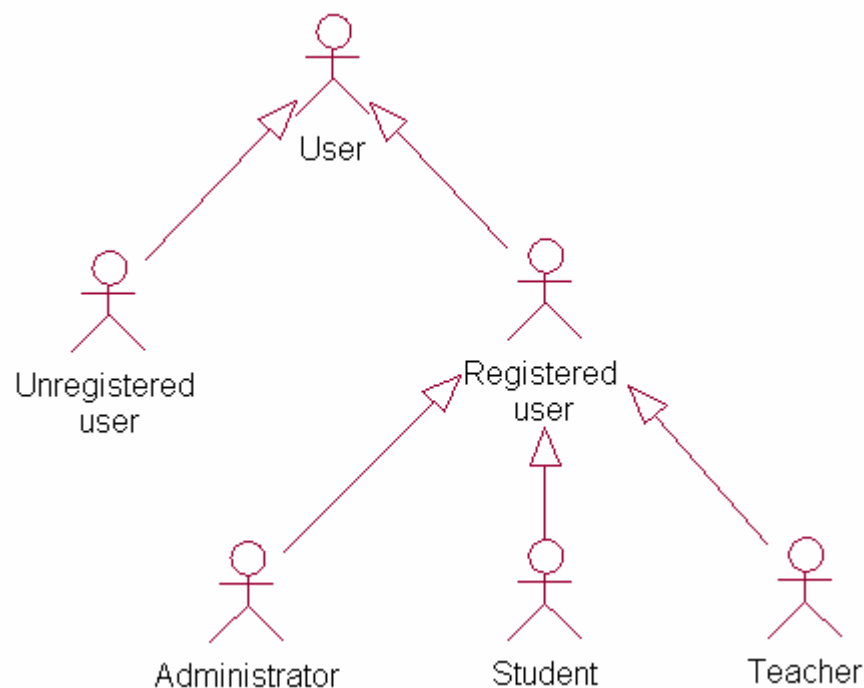


Fig. 11.1. Hierarchy of the user groups accessing the application „Eagle 7”

Each of the identified groups should be described by a specification sheet containing the following elements [CER03]:

1. Name – name of the user group.
2. Description – short description and criteria of clustering members.
3. Profile data – attributes characterizing the members of the group.
4. Super-group – the group that generalizes described one.
5. Sub-groups – a set of groups specializing described one.
6. Relevant use cases – preliminary list of use cases that can be realized by the group members.
7. Access rights – a set of permissions granted to members, regarding content and functionality of the application.

An exemplary specification sheet for the group called Student is presented in the Fig. 11.2.

Many groups may be difficult to percept, especially when it comes to dependencies between them and inheritance hierarchy. Therefore the specification sheets can be extended by the diagram showing hierarchy. Designer can also make additional diagrams in other notations, suitable for the needs of development process.

Name	Student
Description	People studying at university. All types of studies are considered.
Profile data	Personal Data: surname, name, father's name, date and place of birth, Social Security Number Contact data: phone number, address, e-mail Data about studies: credit book number, subject of studies, semester, students group, graduate?, active? Login data: login, password
Super-group	Registered user
Sub-groups	None
Relevant use cases	Modify the account, Login, Count average of grades, Display classes, Display grades, Display attendances, Display announcements, Send message, Display teachers
Access rights	Read: subjects, announcements, classrooms, exams, grades, attendances, reservations of classrooms, publically accessible information about teachers, personal data Read/Write: subscriptions for exams and classes

Fig. 11.2. Specification sheet for the Student user group

11.3. SPECIFICATION OF USE CASES

When it comes to the use cases specification, WebML takes advantage of use case diagrams known from UML (Fig. 11.3).

Interactions between users and application can be clearly presented by this kind of diagram. It is also facilitated by scenarios consisted of elements [CER03]:

1. Name – name of use case.
2. Description – short description of the functionality presented by the use case.

3. Pre-conditions – conditions that have to be fulfilled to perform the use case.
4. Post-conditions – conditions that are fulfilled after performing the use case.
5. Workflow – the chain of activities to be performed for successfully executing the use case.

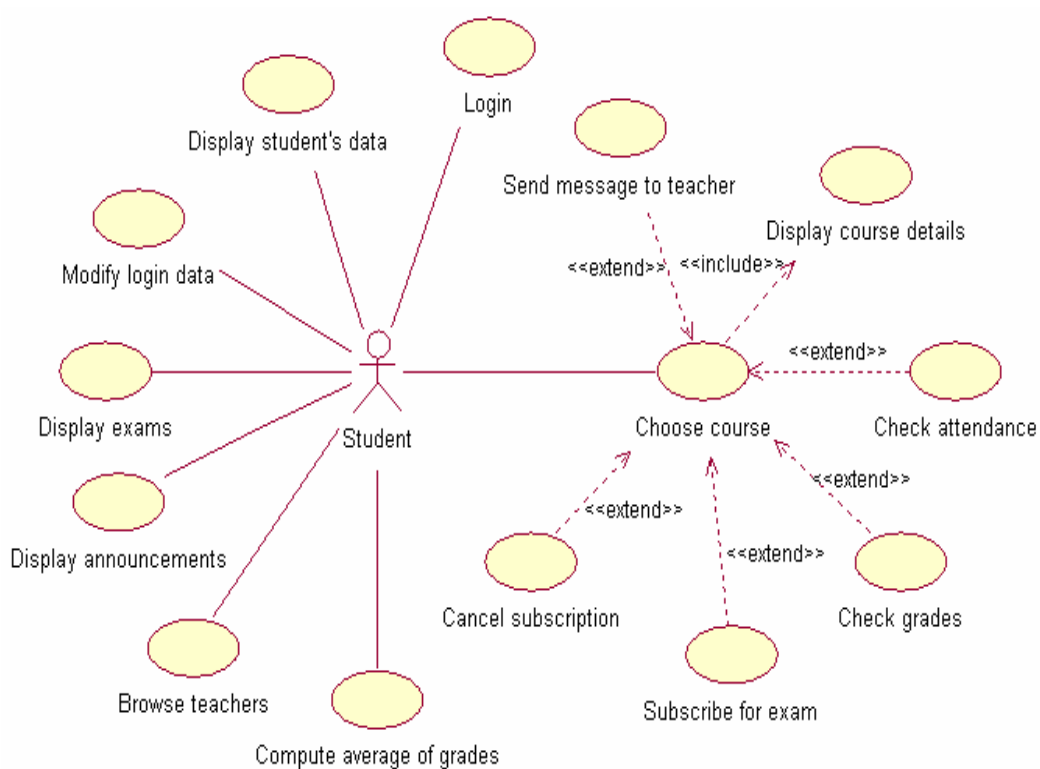


Fig. 11.3. Use case diagram showing interactions between Student and application „Eagle 7”

Exemplary scenario is presented in the Fig. 11.4.

Name	Subscribe for exam
Description	It allows to sing up on the list of students to examine. It is accessible when students are divided into few groups to examine in different time.
Pre-conditions	Student is logged in and the course is chosen.
Post-conditions	Student is signed up on the appropriate list.
Workflow	<ol style="list-style-type: none"> 1. Student receives a list of available terms. 2. Student chooses the term. 3. Student receives the number of available seats and information about classroom, date and time of the exam, students on the list. 4. When the list is not full, student signs up on the list. 5. If no error occurred, confirmation is displayed.

Fig. 11.4. Scenario regarding „Subscribe for exam” use case

The main disadvantage of proposed by WebML scenarios is the lack of separate sections devoted to alternative workflows and unexpected situations like hardware malfunction. Workflow presents only one way of successfully executing use case, so the user does not know how the software will react in other situations. Fortunately there are no objections to extend scenarios specification by the sections mentioned earlier. Furthermore, depending on the complexity and size of developed application, use case specification can also be extended by the activity diagrams.

11.4. SPECIFICATION OF DATA DICTIONARY

Data dictionary is a list of the most relevant information objects in the application domain. Listed objects depend on the main purpose of the application and the way of choosing them is strictly intuitive. One of the guidelines says: “The object is important, when it helps to reach the goal of the application” [CER03].

Data dictionary specification sheet consists of the following elements [CER03]:

1. Name – name of the object.
2. Synonyms – alternative names for the object used in context of the application.
3. Description – short description of the meaning of the object in context of the application.
4. Sample instances – exemplary instances of the object helping to understand its role in the application domain.
5. Properties – the most important attributes of the object with short description.
6. Components – a list of the most important internal components of the described object.
7. Relationships – a list of the most important relationships between other objects.
8. Super-concept – concept that generalizes the features of the object.
9. Sub-concepts – concepts that specialize the features of the object.

The following information objects can be identified in a domain of “Eagle 7”. The goal of exemplary application is also to present some data about actors – students and teachers, therefore they are included in the data dictionary (Fig. 11.5).

This kind of specification can be extended by the UML class diagram. It is helpful in case of many complex objects (containing many components) connected with many dependencies.

Name	Student
Synonyms	Undergraduate, pupil
Description	Object holds detailed information about student attending courses conducted by the teachers at the institute.
Sample instances	Adam Nowakowski, son of Bożydar, born in Szczecbrzeszyn on 23.09.1984, Social Security Number 84092345013, phone number 693121212, living in Lublin, Chemiczna Street 23/8, 20-623 Lublin, e-mail address franek@lidel.org, studying IT at Lublin University of Technology, credit book number D545691, group ID8.4, semester 8, not graduate yet, owner of the active account with login 'adam' and pass phrase 'lubelskie'
Properties	<ul style="list-style-type: none"> • surname – student's surname • name – student's name • father – father's name • birth_date – date of birth • place – place of birth • SSN – Social Security Number • cb_number – number of student's credit book • subject – subject of studies • studies_type – type of studies in terms of the way of attending classes • semester – semester of studies • group – students group • graduate – indicates if the student finished his studies • active – indicates if the personal account is active • login – username • pass – password phrase
Components <i>Contact Data</i>	Properties: <ul style="list-style-type: none"> • phone – phone number • city – place of living (part of the address) • street – street, where student lives (part of the address) • number – house or flat number (part of the address) • postal_code – postal code (part of the address) • post – place where the post office is placed (part of the address) • e_mail – e-mail address
Relationships <i>StudentToCourse</i> <i>StudentToExam</i>	Connects student with courses conducted by a teacher. Connects student with his exams.
Super-concept	Registered users
Sub-concepts	Graduates, students attending to courses

Fig. 11.5. Specification of information object representing a Student

11.5. SPECIFICATION OF SITE VIEWS

Site view “serves the use cases associated to one or more groups, and offers access or content management functions over selected data elements” [CER03]. The input to its specification is specification of user groups, use cases and data dictionary.

The specification sheet of a site view consists of elements [CER03]:

1. Name – name of the site view.
2. Description – short description of served functionality.
3. User groups – the list of user groups that are allowed to access the site view.
4. Use cases – use cases served by the site view.
5. Site view map – table describing areas (effect of division of the site view), which serves smaller parts of site view’s functionality.
 - 5.1. Area name – name of the area.
 - 5.2. Area description – short description of functionality provided by the area.
 - 5.3. Accessed/Managed objects – a list of objects from the data dictionary managed in the area.
 - 5.4. Priority level – numerical or symbolic value that indicates the importance of the area and establishes the order of areas design and development.

During requirement specification of the exemplary application, four site views were identified. They were assigned for students, teachers, administrators and unregistered users. Exemplary specification sheet (assigned for students) is presented in the Fig. 11.6.

Name	Students		
Description	Site view dedicated for students. It allows to access information about courses, exams and personal data.		
User groups	Student		
Use cases	Display student's data, Modify login data, Display exams, Display announcements, Browse teachers, Compute average of grades, Choose course, Send message to teacher, Display course details, Check attendance, Check grades, Subscribe for exam, Cancel subscription		
Site view map			
Area name	Area description	Objects	Pri- ority
Courses	Default page (home page of the area) contains the list of courses and gives ability to choose one of them. Another page displays the details about chosen course – classroom, date and time, subject, teacher. It also allows to gain access to the pages presenting student's attendance, grades and form that sends message to the teacher. Another page allows to make subscription for an exam.	Read: teacher, student, subject, classroom, course, term of exam	High
Settings	This area allows student to browse all of his data and manage login data. Default page contains detailed data regarding logged in student. Student is able to reach the pages designed to change the login data. All changes are sent via e-mail.	Read: student Read/Write: login data	High
Exams	This area allows to browse information about student's exams and compute some statistics (like average of grades). Default page of the area displays courses and exams. When the course is chosen, the page that shows details about exam and grades is displayed. The default page allows to compute average of grades received by the logged in student.	Read: exam, teacher, course	Low
Tea- chers	This area allows to browse publicly available information about teachers (like name, surname, science degree, work phone, e-mail) and send e-mail to the chosen one. Default page displays a list of teachers and allows to choose one of them. Another page displays detailed information about teacher and form for sending an e-mail.	Read: teacher, student	Low

Fig. 11.6. Specification sheet for site view dedicated for Student

In some cases it is important to point out properties of some objects (like teacher object) accessible by other groups of users. For example student should not know about home address of the teacher or his salary. It is not explicitly specified by the WebML notation, where to put this information. Designer has to decide about that on his own.

During this part of specification, reusable modules can also be identified. It is especially significant in further designing process, when it comes to hypertext modeling. Such modules allows to replace sets of components with only one component acting like a black box. It influences the readability of model, which becomes less complicated (less number of components and connections between them). After analyzing specification of the exemplary application, some potential modules can be identified: pages displaying details about exams and courses, pages displaying data about teachers and students, and pages used for sending messages.

11.6. SPECIFICATION OF STYLE GUIDELINES AND NONFUNCTIONAL REQUIREMENTS

When it comes to look and feel of the application, WebML advices to use mockups, presenting the position and size of areas showing content and providing some functionality (Fig. 11.7). All of commonly used guidelines are also applicable [CER03].

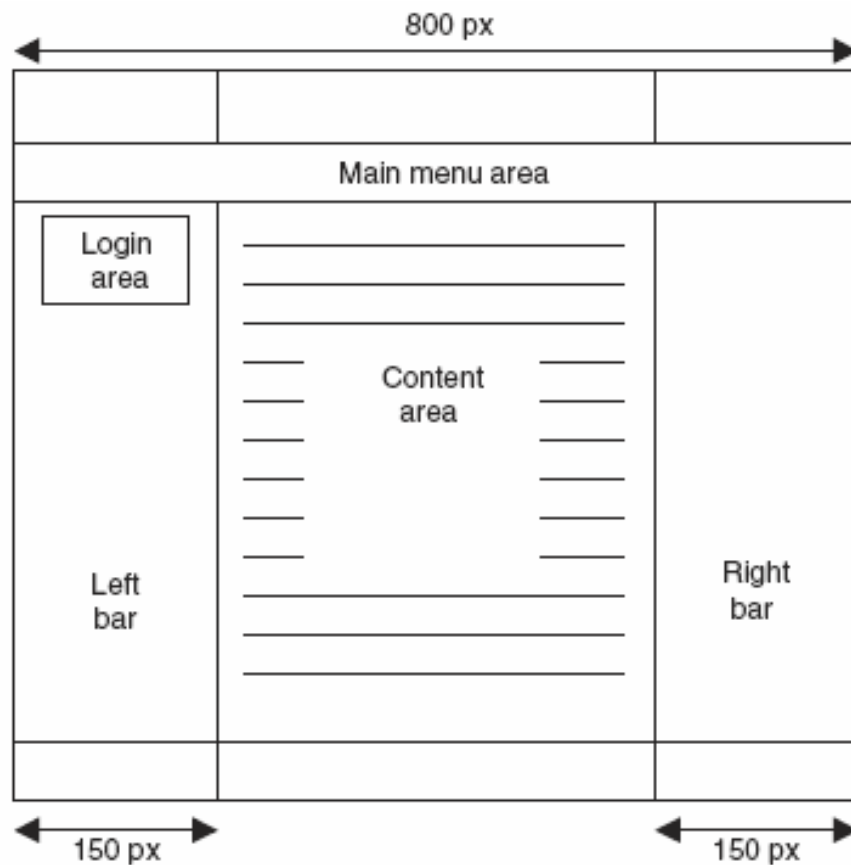


Fig. 11.7. An exemplary mockup [CER03]

Similarly, when it comes to nonfunctional requirements specification, all of commonly used guidelines are applicable. Although it is a good habit to describe issues like [CER03]:

1. Performance – the most suited and efficient hardware platform, which resources will be used by users and the application.
2. Scalability – ability of increasing the performance of the services in response to the growing volume of requests.
3. Availability – ways of assuring minimization of errors and failures.

4. Security – protection of integrity, confidentiality and privacy of the stored information, safe authentication of the users, protection of the information passing by the services.
5. Maintainability – ease of services maintenance and ability to personalize or adapt services to the specific users needs.

11.7. SUMMARY

Specification adapted by the WebML may be time consuming, although it is very precise. In an easy to read way describes all of the details of application, that will be developed.

It is based on descriptions that in the further steps of the WebML application development will be precised and formalized. It is worth to point out, that this specification holds a lot of necessary data regarding the application and it can be the basis for diagrams in UML or WebML style. It can be also used as a start for other application development methodologies.

The WebML notation concentrates on:

- Objects and entities on which the application operates
- The content, objects and functions distribution on the pages of application
- The control and processing flow between application pages
- Interaction between user and application

It considers all the technical aspects as implicit, which in case of need can be described by the appropriate supporting diagrams

Such a way of the requirements specification facilitates the creation of a documentation readable and understandable not only for the designers and coders but also for the unexperienced people involved in the creation of the application. It creates an user-oriented specification.

REFERENCES

- [CER03] Ceri S., Fraternali P., Bongio A., Brambilla M., Comai S., Matera M.:
Designing Data – Intensive Web Applications. Elsevier Science, 2003

Experiences in Building the Information Society in Poland

IT sector may be a force of fast development of the economy. In 1997 the Central Bank announced, that one dollar invested in the ICT causes investments of 10 dollars in other sectors of economy [WYB97]. It means new possibilities, reducing the unemployment rate, improving living conditions and making societies wealthy and prospering. Authors of so called Bangemann Report [BAN94] put the following statement: "The first countries to enter the information society will reap the greatest rewards." But they also include a warning: "By contrast, countries which temporise, or favour half-hearted solutions, could, in less than a decade, face disastrous declines in investment and a squeeze on jobs." No wonder that many countries have undertaken a great effort to enter into the path of evaluation towards the Information Society.

12.1. POLISH WAY TO THE INFORMATION SOCIETY

Also the authorities of Poland considered the idea of the Information Society and Knowledge-Based Economy to be the unique opportunity. The beginnings were promising. In 1994 First Polish Computer Science Congress took place. This was the meeting of about 500 participants from government, parliament, academic, users of IT applications, training companies and companies producing and selling IT products.

As the result the report 'Strategy of Information Technology Development in Poland' [STR94] was released. It was the first document containing outlines of building the Information Society and Knowledge-Based Economy in Poland. The report ended with recommendations concerning:

- Legislative actions – especially demonopolization of the ICT market, ensuring the free competition, adjusting the already existing and future IT systems to EU standards and privacy policy.
- Ensuring an access to ICT – especially development of Internet backbone, equipping schools in computers and Internet connection especially in the rural areas.
- Changing the system of education – introducing IT to the teaching content and training the youth in computer science.

These goals were extended in numerous documents appearing later. Academic environment was actively involved into planning the information society. In 2000 the document 'Objectives and directions of development of information society in Poland' was released. It contains results of seven expertise published under the collective title: 'Global Information Society in the conditions of Polish accession to the European Union'. This document contained goals of Polish authorities which should be achieved [KBN00]:

1. Ensuring access to telecommunications services for all citizens.
2. Preparing of Polish society to the technological evolution, social and economic problems relating to the formation of the information society.
3. Preparing of Polish society to the new labour market conditions and new ways of working: taking benefits and exploit the opportunities related to the changes, to cope with unemployment created by the restructuring of industry and agriculture.
4. Adaptation of regulatory requirements and rapid technological progress of information society era and using of ICT to develop new acts.
5. Adaptation of the national economy to the requirements of global electronic commerce by introducing regulations on electronic commerce.

6. Strengthening the public procurement system in order to facilitate the implementation of telecommunications systems for public sector entities. Improving the public procurement system through the use of ICT networks.
7. Establishing transparent and citizen-friendly structures of public administration as an open information society by using ICT tools and streamlining of the administration through the wider use of ICT.
8. Creating conditions for sustainable and balanced regional development with emphasis on modern ICT and its monitoring.
9. Developing of modern industries and increasing its innovativeness in order to improve the competitiveness of the Polish economy to the global electronic economy.
10. Providing support to the science sector for electronic commerce and the information society in order to better exploit opportunities and minimize risks. developing infrastructure for science.

Unfortunately, in some areas Poland today is far from bright vision outlined in numerous informatization strategies. It was caused especially by permanent changed plans, inconsistent actions and lack of good will to introduce new solutions. Completion of many programs was continually postponed. For instance the IT system for Social Security (ZUS) is still under construction after 13 years and about 1 billion \$ spent.

It seems that it is a question of lack of competence Polish administration and, the most important, political issues. There is no one, coherent plan, no proper delegation of powers and no permanent control. These three factors are crucial to build efficient e-government and e-administration.

12.2. EDUCATION AND EXPERTS

The case of the IT system for Social Security (ZUS) shows the role of proper expert support. In 2008 ZUS planned to buy 130.000 3.5" floppy discs to store data, which is perceived as out-of-date solution for probably every user of ICT (except people responsible for tenders in ZUS) [BLA08]. Main argument for this decision was the temporary nature of changes. 'Temporary changes' however, often exist longer than planned solutions and often are caused by no ideas to solve the problem. This leads us to questions: who is responsible for decisions? Is he or she prepared? Is he or she administrator or manager? Is he or she responsible for not breaking rules or results of actions taken?

Of course we cannot change dramatically the mentality or way of the administration functioning, but it can be supported by qualified experts. This solution brings savings of time and public money. As an example of necessity of such a support, is the case of use EU funds by local administration. Only 4% of 190 million \$ of funds for telecommunication infrastructure and 8% of 121 million \$ for purchasing computers and payments for the Internet connection were acquired. The money must be allocated by the end of 2013, and spend by the end of 2015. Three years left. So weak result, is caused by too complicated procedures, which became an impassable barrier for local administration without expert support [POZ10]. Overregulation affects also private sector, but entrepreneurs have more motivation to deal with red tape.

In the other hand Poland has great achievements in education of the youth. Polish computer scientists regularly gain high positions in international IT competitions. It is due to informatization of schools and academic support.

Education however, must be continuously adjusted to situation and cannot be treated as one-time fixed. It is crucial for society to obtain useful skills, especially in the area of new technology [SKI10]. It means learning programs containing subject preparing to collect, process, analyze and use data. It also means new form of education (e.g. e-learning) and involving citizens in long life learning process.

Investing in new generation leads to acquisition of skills or its improvement by older people in household with pupils or students. Older people possess experience and skills, which often can better be utilized, by using ICT.

12.3. LEGISLATION

Monopoly leads to stagnation. One of the barriers of development in Poland was the position of TPSA (Polish Telecom) – monopolist on the telecommunication market. The enterprise for many years tried to hold its position, but finally started to share the market with other companies. It resulted in price reductions which increased demand and makes the market blooming.

Ensuring proper law protection of sensitive data in Poland has been a success. Polish law was in line with Directive of European Parliament and Council 95/46/EC of 24 October 1995. The Directive introduces general principles [BAR09]:

- notification - the person whose data is collected must be informed of this fact and about the way in which the data will be used;
- selection - the person whose data is collected must be able to resign from the data collection process;
- transfer of data - the data can be transferred only to entities that provide an equally high level of protection of personal data;
- Security - data must be protected from loss
- data integrity - data must be reliable and relevant to the purpose for which it was collected
- access - the person whose data are collected must have access to the collected data, including their correction or deletion if they are inaccurate
- implementation - these rules must be effectively implemented.

It was very important to develop financial sector, which now takes benefits from new technologies and may be perceived as developed at a high level.

Unfortunately, law concerning the introduction the information society in different areas, especially in administration is changing very often, which becomes an unnecessary burden for individual citizen, entrepreneurs and administration. It also makes finishing plans uncertain or difficult.

12.4. INFRASTRUCTURE

There is no Information Society and knowledge-based economy without proper infrastructure. Case of Poland shows, that it is important to build flexible plans. This flexibility is necessary to adjust implementation to rapidly changing technology. There should be possibility of changing assumptions (e.g. change of the capacity from 512 Kb to 2 Mb) without delaying or abandoning the project. It is important at the beginning of the project to look further and plan to use as new technologies as possible. During implementation all assumption should be regularly checked.

Building new connection lines requires large expenditures. Money may come from public or private investors. Advantages and disadvantages of both solutions are shown below (Tab. 12.1).

Table 12.1. Chosen advantages and disadvantages of public and private investments in IT infrastructure

	Advantages	Disadvantages
Entrepreneurs	<ul style="list-style-type: none"> • development regulated by the market • better adjusting to needs of users • low costs for end user (as the result of competition) 	<ul style="list-style-type: none"> • it is regarded as risky investment (high exit barriers)
Public funds	<ul style="list-style-type: none"> • bigger investments (sum of money) • development of not profitable areas (distant or rural) 	<ul style="list-style-type: none"> • low flexibility and adjusting to the market needs and changing situation

The perfect solution is the public – private partnership. To decrease the investment risk, clear, not complicated and stabile legal act should be proposed. It will encourage entrepreneurs to invest more funds in infrastructure. The main role of the government will be to stabilize and stimulate the market (improving competition conditions and facilitate spreading new technologies to increase demand). Academics and experts will provide solutions and ensure proper information flow. It gives us triangle of partnership: authorities, entrepreneurs and academics.

There is also a question who should build the Information Society. There are three options:

1. Government chooses best directions of development and provides grants.
2. The market is the only creator of the Information Society development
3. There is private – public partnership which means cooperation between authorities and private companies.

Bottom-up initiative and public-private partnership can contribute significant value to development of the Information Society. It is also a proper base of setting actions by the Polish authorities. It is probably the most promising direction, because government can contribute by creating right climate (e.g. investor-friendly law or reducing red tape). On the other hand, entrepreneurs know their clients and their needs, as well as possibility of implementation. They would probably choose the most effective technology (high value of product) and they would try to decrease costs to achieve high incomes.

12.5. CASE STUDY - E-GOVERNMENT IN POLAND

E-government was the initiative of using ICT to serve citizens. Depending on the use of technology there are four levels of e-government service maturity [JEZ06]:

1. Information level – offices publish information in the Internet. Citizens obtain necessary information searching on official web sites or in special information kiosks.

2. Interactional level – citizens can communicate with particular offices using the Internet, but offices do not have to answer.
3. Transactional level – citizen can communicate with offices, and application of offices give response.
4. Integration level – citizen can communicate with chosen office, application of offices give response, and applications of one office is integrated with applications of other offices. It means, that citizen is able to obtain necessary information (e.g. forms), make a transaction (send forms, fulfill form on-line, pay taxes etc.) and get the decision, certificate or statement. The information about transaction will be available for every office which is concern.

To achieve the most desirable level four, the administration should take into account areas of implementation the solutions [OLS08]:

- Internal administration communication.
- External communication with EU.
- Communication with citizens and enterprises.
- Informing the society.

Building new system should provide interoperability, which means that there should be one coherent set of procedures methodologies, standards and specifications recommended for state government entities. It is very difficult especially in conditions of permanent law changes. Number of law act amendments is very high. For example, during preparing the system Payer for ZUS, there were at least 230 changes in act concerning Social Security [OLS08].

The second problem is the compatibility of software. Offices are equipped with different systems provided by different software providers (in 2008 there were 220 systems used in the financial sector [OLS08]). In fact it is an organizational/law problem. If public tenders on IT systems concerned not particular municipal office, but specification of winning system should have been introduced in all municipal offices,

there would not be problem with double working, delays (objections) and missed specifications.[GRY09].

No wonder, that the launch of e-government platforms were being postponed and fraught with numerous errors. Finally we can take benefits from the e-government platforms of particular offices and central platform: ePUAP (ang. *Electronic Platform of Public Administration Services*) - Fig.12.1.



Fig. 12.1. ePUAP homepage [EPU10]

The platform is user-friendly built. It contains particular services (receiving ID, voting, permissions & concessions etc.) available via the Internet, grouped in categories (citizen matters, enterprise, automotive and transport etc.). In the right

there are three sections providing information about ePUAP (*what is this page for?*, *want more?* and *help needed?*). At the bottom there is news for citizens concerning platform and its usability, and information for offices (using system of ePUAP to local service). The platform was not widely used due to the necessity of confirming identity of citizens [ZAW09].

Since April of 2009 e-Declaration (pol. E-Deklaracje) platform was launch (Fig. 12.2). According to plans first declaration should have been send via Internet since 2006. Complete system should have been implemented to 2008 [ZAW06].



Fig. 12.2. E-declaration platform [EDE10]

Now, there is no necessity of use expensive electronic digital signature (possessed device/software) by individuals using e-Declarations. Identity is confirmed by correct data: PESEL, NIP, date of birth and name [SKW09].

Since 2011 the new way of using ePUAP is available Trusted Profile account. It is a kind of digital identity, confirming identity of citizen. To create the Trusted Profile

citizen must create an account on the ePUAP platform, and confirm own identity in office (e.g. regional office, social security, tax office etc.). If data in profile is consisted with data from ID documents of citizen, the profile is confirmed and it becomes Trusted Profile, freely used to communication with offices via ePUAP or to confirming forms.

In the future as the most probable way of communication with offices, the new ID and PESEL2 would be used.

It seems, that e-government in Poland is really welcomed and awaited, especially by entrepreneurs, but also ordinary people. It is win-win situation: lower costs of business running, lower cost of administration and enormous savings of time. Enterprises are still obliged to use electronic digital signature. Nonetheless entrepreneurs use the Internet to handle official matters more and more often (Fig. 12.3).

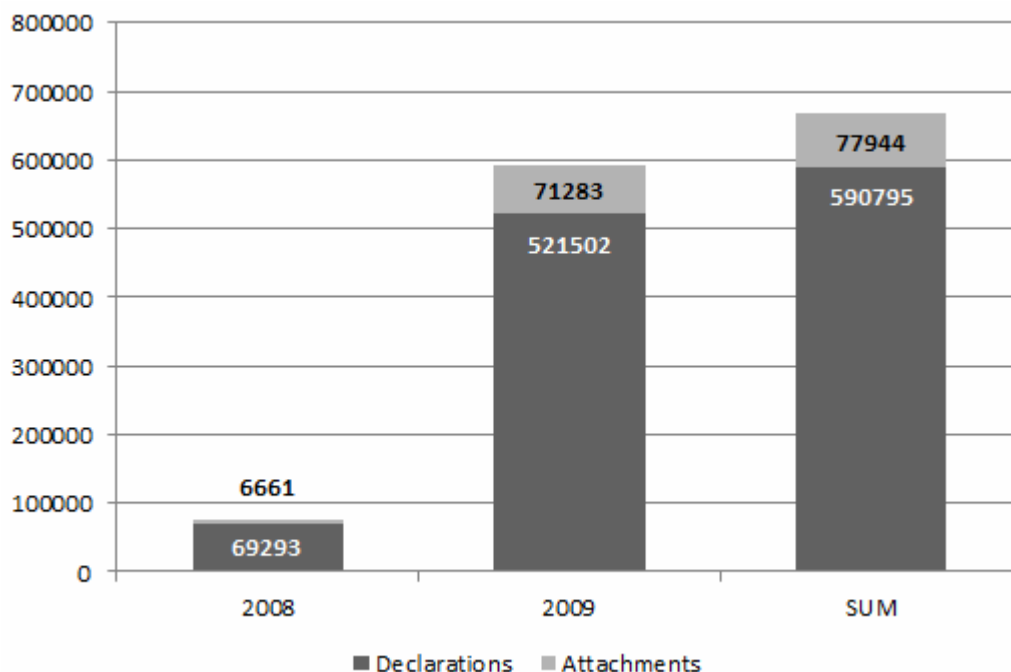


Fig. 12.3. Number of electronic documents submitted by companies 01.01.2008 – 22.10.2009 [ZAW09]

As the next goal there should be the attempt to integrate all systems (the fourth level of e-government development). It would reduce the number of citizen duties and probably reduce the time of setting new business. And that would make Polish economy more flexible and active.

Poland since 2004 is a part of EU, so there is a need of communication with offices in other languages, or at least putting information on web sites in foreign languages to encourage investors and supply them with proper information.

12.6. CONCLUSIONS

Poland is changing. ICT are widely used by citizen, but there is a lot to do, especially in case of good, stabile law. Approaching to the information society is not only about using ICT, it is also about reducing costs and enabling citizens to taking benefits from new technologies in private life or in their work. ICT can be also used to involving society in democratic processes. Well skilled citizens, developed infrastructure and commonly used new technologies allow transform Polish economy in truly knowledge-based economy.

REFERENCES

- [BAN94] *Europe and the global information society*. Recommendations to the European Council, (Bangemann Report), Brussels, 26.05.1994, p. 6
- [BAR09] Bartosiak J.: *O polityce prywatności*. (<http://prawo.vagla.pl/node/8419>), 28.03.2009
- [BLA08] Błaszczkiewicz D.: *ZUS ogłasza przetarg na 130 tysięcy dyskietek 3,5" (1,44 MB)*, (<http://www.idg.pl/news/174141/ZUS.oglasza.przetarg.na.130.tysiecy.dyski.etek.35.144.MB.html>) 14.11.2008
- [EDE10] e-Deklaracje (<http://www.e-deklaracje.pl/>), 10.10.2010
- [EPU10] ePUAP (<http://epuap.gov.pl/wps/portal>), 10.10.2010
- [GRY09] Gryniewicz T., Poznański P.: *Wymiećmy spod koldry brudy informatyzacji*. (<http://wyborcza.biz/biznes/1,101562,7063982,Wymiecmyspodkoldrybrudyinformatyzacji.html>), 21.09.2009
- [JEZ06] Jezierska A.: *System informatyczny w urzędzie*. Gazeta prawna 082/2006
- [KBN00] *Cele i kierunki rozwoju społeczeństwa informacyjnego w Polsce*. Komitet Badań Naukowych, Warszawa, 28.02.2000 r.
- [OLS08] Olszak C.M, Bilewicz G.: *Wybrane problemy rozwoju administracji publicznej w warunkach gospodarki elektronicznej*. (<http://www.swo.ae.katowice.pl/>), 16.03.2008
- [POZ10] Poznański P.: *Inwestycje w internet utknęły w MSWiA. Ministerstwo preferuje powodziań*, Gazeta Wyborcza, (<http://wyborcza.biz/biznes/1,101562,8050684,InwestycjewinternetutknelywMSWiAMinisterstwo.html>) 22.06.2010
- [SKI10] *Skills and jobs experts call for action now*, (<http://ec.europa.eu/social/main.jsp?langId=en&catId=89&newsId=697&furtherNews=yes>), 04/02/2010

- [SKW09] Skwirowski P.: *Wreszcie rusza e-PIT-37*. Gazeta Wyborcza Biz (http://wyborcza.biz/biznes/1,101562,6482237,Wreszcie_rusza_e_PIT_37.htm) 08.04.2009
- [STR94] *Strategy of Information Technology Development in Poland*. Report of First Polish Computer Science Congress, 1994
- [WYB97] *Wybrane problemy polskiej teleinformatyki*, Informatyka, nr 10/1997, p. 8–10
- [ZAW06] Zawadzka A.: *PIT wysłamy przez internet*. (<http://manager.money.pl/news/artukul/pit;wyslemy;przez;internet,55,0,148279.html>), 2006-03-27 14:28
- [ZAW09] Zawadzka A.: *Tylko 0,5 proc. podatników złożyło PIT przez internet*. (<http://msp.money.pl/wiadomosci/podatki/artukul/tylko;0;5;proc;podatnikow;zlozylo;pit;przez;internet,243,0,550131.html>), 26.10.2009

Author Index

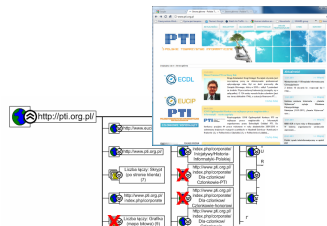
Brazhkina Ariana	Tupolev State Technical University Kazan, Russia
Dzieńkowski Mariusz	Lublin University of Technology Faculty of Management Department of Management Lublin, Poland
Juszczyk Marta	Lublin University of Technology Faculty of Management Department of Management Lublin, Poland
Kaczorowska Anna	University of Łódź Faculty of Management Department of Computer Science Łódź, Poland
Kęsik Jacek	Lublin University of Technology Computer Science Institute Lublin, Poland
Laskowski Maciej	Lublin University of Technology Computer Science Institute Lublin, Poland
Marciniak Radosław	College of Enterprise and Administration in Lublin, Lublin, Poland

Miłosz Elżbieta	Lublin University of Technology Faculty of Management Department of Management Lublin, Poland
Miłosz Marek	Lublin University of Technology Computer Science Institute Lublin, Poland
Pamuła Anna	University of Łódź Faculty of Management Department of Computer Science Łódź, Poland
Plechawska-Wójcik Małgorzata	Lublin University of Technology Computer Science Institute Lublin, Poland
Rakhimova Diana	al-Farabi Kazakh National University Almaty, Kazakhstan
Sirazetdinov Rifkat	Tupolev State Technical University Kazan, Russia
Suleymanov Alfred M.	Kazan State University of Architecture and Engineering Kazan, Russia
Tukeyev Ualsher	al-Farabi Kazakh National University Almaty, Kazakhstan
Wit Bogdan	Lublin University of Technology Faculty of Management Department of Management Lublin, Poland
Zhumanov Zhandos	al-Farabi Kazakh National University Almaty, Kazakhstan
Żyła Kamil	Lublin University of Technology Computer Science Institute Lublin, Poland

Edited by Marek Miłosz

ICT

- from Theory to Practice



This book is the result of international cooperation between Lublin University of Technology (Poland) and universities from from Lublin and Łódź (Poland), Kazan (Russian Federation) and Almaty (Kazakhstan). The main area of this cooperation is Information and Communication Technology (ICT).

Due to the relatively large variety of themes, the book is addressed to various groups of readers. A student, a researcher and a computer specialist from industry can find here something interesting for him/herself. A special recipient of the book is a Computer Science student, who wants to select the topic of the master thesis. The book will allow him/her to find potential issues of the thesis and to recognize the degree of their complexity.



ISBN 978-83-931710-1-9