Proceedings IIS

International Conference on Intelligent Information Systems

The Conference is dedicated to the 50th anniversary of the Institute of Mathematics and Computer Science

Institute of Mathematics and Computer Science
August 20-23, 2013, Chisinau, Moldova

CZU 004.8(082)

I-58

INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE
5, Academiei street, Chisinau, Republic of Moldova, MD 2028
Tel: (373 22) 72-59-82, Fax: (373 22) 73-80-27,
E-mail: imam@math.md
WEB address: http://www.math.md

Editors: Prof. C.Gaindric, Prof. S.Cojocaru.

Authors are fully responsible for the content of their papers.

Descrierea CIP a Camerei Naționale a Cărții

International Conference on Intelligent Information Systems: The Conference is dedicated to the 50th anniversary of the Institute of Mathematics and Computer Science, 20-23 aug. 2013, Chişinău, Moldova: Proceedings IIS / ed.: C. Gaindric, S. Cojocaru. – Chişinău: Institute of Mathematics and Computer Science, 2013 (Tipogr. "Valinex SRL"). – 274 p.

Antetit.: Inst. of Mathematics and Computer Science, Acad. of Sciences of Moldova . – Referințe bibliogr. la sfârşitul art. – 100 ex.

ISBN 978-9975-4237-1-7.

004.8(082)

ISBN 978-9975-4237-1-7

This issue is supported by the Information Society Development Institute

Part 1

INVITED PAPERS

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Steering the Information Society Development as a Vision-Based, Multi-Stakeholder Process: a Point of View

Horatiu Dragomirescu, Florin Gheorghe Filip

Abstract: The steering of the information society development at country level is a desideratum for the smooth unfolding of the respective societal change - nowadays a key issue worldwide for scientific research and public governance. Some possible guidelines are presented on how to come up with a vision shared by stakeholders and how to ensure concertation among them along the way.

1 Introduction

Information society is nowadays counted among those select issues that are top-ranked on the research agenda worldwide, as well as on the policy-making agenda of states and international organisations, such as the United Nations or the European Union. Interestingly, the tendency towards more and more diverse and refined ways of theorising the information society is paralleled by the pragmatic quest for synergy in blueprints and initiatives within and across countries. Progressing towards the information society is a priority held in common by both developed and developing countries, although their respective approaches significantly differ. Developed countries generally focus on advancing R&D in the field of ICTs and the subsequent boost of market supply of more and more sophisticated goods and services. In turn, developing countries, faced with sharp digital divide, due to their lagging behind in terms of computer skills and infrastructures, are chiefly concerned with taking catching-up steps and obtaining international assistance.

The issue of information society development at large, including the particular aspect of its steering, calls for a pluralist stance being adopted in thinking and acting alike, in light of a consistent series of arguments:

- theoretical: the multitude of scientific disciplines concerned with studying the information society (sociology, anthropology, informatics etc.) the encounter of which turned the respective matter into a standalone interdisciplinary subject;
- technological: the interplay of several key technologies of the digital age (computing, telecommunications networks, broadcasting) that came up to convergence;
- actional: the development of the information society became an endeavour that is subject to enactment at various levels (national, regional, global), a wide range of stakeholders being involved therein (the state, academia, business companies, the civil society, communities with professional, socio-cultural or territorial grounds, individual citizens).

The multiple ways in which the information society was defined along the timeline differ, depending of the perspective adopted, which can be either economical, social, technological, occupational or cultural (Webster, 2006, p. 8). Out of these, Kellerman (2002, p. 10) privileges the economical and the cultural ones; the information society is characterised, from the economic point of view, by the key role of information resources and information industries while, from the cultural one, it features a flourishing creativity and a rising intellectual standing of individuals.

At the second round of the World Summit on the Information Society, convened by the United Nations, in Tunis, in November 2005, there were consensually stated the "desire and commitment to build a people-centred, inclusive and development-oriented Information Society" (WSIS, 2005). Such an option is in line with the bolder emphasis on "I" term (information), within the "IT" construct (Davenport, 2000), and also with the reinforced anthropocentric and collaborative orientation in designing next-generation information systems.

2 The path of the information society development

The information society development at country level can be understood as a process of evolving towards upper maturity stages of this type of

societal system; the knowledge society represents the fully mature stage attainable by an information society over time.

At global scale, as Himanen (2004) remarked, "[T]he first phase of the information society focused on the development of technology, such as network connections. In the second phase, which has now begun, technological development will continue; however, the focus will shift to larger social matters and the main focus will be on changing the ways in which we operate."

The information society development path at country level can be mapped in several modalities. The qualitative ones are generally based upon scales of successive development stages, considered either prospectively or in retrospect. Each stage is assigned specific features the actual presence of which reveals the attainment of a certain maturity degree, with a view to assessing progresses achieved or projecting the way ahead.

Miles (2002, p. 163) proposed a range of maturity stages defined in metaphorical terms: islands, archipelagos, continents, ecosystem; this approach is useful in that it renders the bottom-up information society development path intelligible, although in rather intuitive terms. Accordingly, its steering process is meant to foster the consolidation of punctual instantiations into an integrated whole.

A scale with a more analytical format was proposed by Rao (2005), in which the informational society development stages range from the incipient ones (disarticulated, embryonic), followed by those where maturation is still underway (development, concertation, intermediate), up to the top ones (mature, advanced, world leader); the respective stages are distinguishable against 8 characteristics (connectivity, content, communities, commerce, culture, capacities, cooperation, capital).

Quantitative approaches, in turn, involve the use of specific metrics aimed at visualising the development degree reached by the information society at different levels (country, region, multi-country). Dedicated composite indexes became more popular lately and are widely used, especially by international bodies; the reasons are their transparent computation methodologies and the easiness of their assessment by policy-makers, as well as by the public at large. Their shortcomings include the fact that they are merging several dimensions, each of them

having a standalone significance that could be disguised through statistical consolidation; moreover, if countries in upper ranks are perceived as de-facto standards, a race for higher scores is triggered to the detriment of cooperation. For monitoring purposes, specific indicators can also be used in their elementary forms, but they become more informative if appropriately selected and assembled into meaningful scorecards that allow for synoptic grasp.

3 The need for shared vision and concertation among stakeholders

Although the passage from the industrial society to the information society was recognised as a global mega-trend, this shift does not take place by default; an overall steering is required to ensure its pertinent orientation, sustained pace and beneficial systemic impact. Such a steering should be exerted as a vision-based, multi-stakeholder process.

According to Afsarmanesh and Msanjila (2010, p. 62), "[A] vision is a deeply held picture of where a person, a group of people, an organization, or a society, wants to reach in the future." It is meant to lay down the main directions towards meeting the final goal, without setting specific targets and deadlines. Foresight studies are always useful in preparing such visions of information society development at country level (Dragomirescu and Filip, 2008).

In order for stakeholders to keep a sustainable engagement, it is essential that the respective vision be generated collaboratively; further on, a regular consultation framework should be established for updating and progress evaluation purposes.

The shared vision, once validated, allows for harmonising the unfolding of individual stakeholders' self-managed initiatives as well as their participation in joint actions. Concertation among stakeholders is a challenging endeavour though, as they have different logics and also maintain their autonomy. Cultivating autonomy in terms of choosing means and practical solutions is, nevertheless, fully compatible with cooperation on the grounds of the shared vision on information society development.

It may occur that, especially in developing countries, state bodies take lead in exerting the steering process. Understandably, the state always plays a key role, as it is in charge with designing and implementing public policies, also administering public spending; besides, the public sector is a major information producer. However, state bodies, taking the lead in steering the information society development, can be perceived by the other stakeholders as hegemonic, reverse effects being thus likely to occur as pluralism is weakened. A viable alternative in this respect consists of the lead being taken by independent, highly representative and forward-thinking bodies such as countries' academies of science (Dragomirescu and Filip, 2008). Consequently, governmental agencies set up to supervise the development of the information society should rather focus on operational aspects, acting on behalf of the state as one of the several stakeholders concerned.

4 Conclusions

Steering the development of the information society at country level involves balancing top-down and bottom-up approaches, incremental and leapfrog types of dynamics, and wisely matching means to the goals pursued. Adopting a systemic mindset, a collaborative format of the relationship among stakeholders, and smart methods of consultation and coordination are among the key success factors in performing the steering as a win-win enterprise.

References

[1] H.Afsarmanesh, S.Msanjila. (2010). *ePAL 2020 Vision for Active Ageing of Senior Professionals.* In: Camarinha-Matos, L. M., Boucher, X. and Afsarmanesh, H. (eds.), "Collaborative Networks for a Sustainable World". Proceedings of the 11th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2010, St. Etienne, France, October 11 - 13, 2010, Springer, Boston, pp. 60 - 72

[2] T. Davenport. (2000). *Putting the I in IT.* In: D. Marchand and T. Davenport (eds.), Mastering Information Management, Pearson Education, London, pp. 5 - 9

[3] H.Dragomirescu, F. G. Filip. (2008). *The Knowledge Society Agenda in Romania: from Experts' Vision to Public Perception.* In: M. D. Lytras et al. (eds.), "The Open Knowledge Society. A Computer Science and Information

Systems Manifesto". Proceedings of the 1st World Summit for the Knowledge Society. Communications in Computer and Information Science, vol. 19, Springer, Berlin, Heidelberg, pp. 24 - 31

[4] P.Himanen. (2004). *Challenges of the Global Information Society.* Committee for the Future, Parliament of Finland, Helsinki. Available at: http://web.eduskunta.fi/dman/Document.phx?documentId=br11307103930385&cmd=download (accessed 08.07.2013)

[5] A. Kelerman. (2002). *The Internet on Earth: A Geography of Information*, Wiley, London and New York

[6] B.Martin. (2005). *Information Society Revisited: from Vision to Reality*. Journal of Information Science, Vol. 31, No. 1, pp. 4 -12

Horatiu Dragomirescu[1], Florin Gheorghe Filip[2]

[1]Bucharest University of Economic Studies - ASE
E-mail: ffilip@acad.ro

[2]Romanian Academy
E-mail: dragomir@ase.ro

NEW VISION and FUTURE of INFORMATICS
Extended abstract

Jozef Gruska

Currently dominating perception of computer science has its origin in a very cleverly written, and much influential, paper of *Newel, Simon and Perlis*, published in Science in 1967, that well captured the perception of the field at that time.

The basic ideas presented in their paper were:

" *Whenever there are phenomena there can be a science dealing with these phenomena. Phenomena breed sciences. Since there are computers, there is computer science. The phenomena surrounding computers are varied, complex and rich.*"

Since that time there have been numerous attempts to modernize such a view of computer science. However, such a computer-centric view of computer science still dominates.

There are nowadays a variety of reasons why such a computer-centric view of the field should be seen as very obsolete and actually damaging the development of the field. They will be discussed in details in the talk. Here are some of them.

- An understanding starts to be developed that information processing plays key role both in physical and biological nature. For example, quantum, DNA, molecular information processing.
- Natural sciences are increasingly seen as being to a large extend information processing driven.
- It starts to be clear that in the future any very significant innovation will use advanced informatics tools, methods and paradigms.

All that requires that a much broader and deeper view of the field should be developed.

A new perception of the informatics presented in the talk will see the field as consisting of four much interleaved components: (a) scientific informatics; (b) technological informatics; (c) new methodology; (d) application informatics. As a scientific discipline of a very broad scope and deep nature, Informatics has many goals. Its main task is Main tasks of scientific informatics are to discover, explore and exploit in depth, the laws, limitations, paradigms, concepts, models, theories, phenomena, structures and processes of both natural and virtual information processing worlds.

To achieve its tasks, scientific Informatics concentrates on new, information processing based, understanding of universe, evolution, nature, life (both natural and artificial), brain and mind processes, intelligence, creativity, information storing, processing and transmission systems and tools, complexity, security, and other basic phenomena of information processing worlds.

One way to illustrate such a broad and deep perception of scientific informatics will be in the talk through presentation and analysis of its grand challenges. The same will be for technological informatics and applied informatics.

Of a key importance for a new perception of informatics is an understanding that informatics, as a symbiosis of a scientific and a technology discipline, develops also basic ingredients of a new, in addition to theory and experiments, third basic methodology for all sciences, technologies and society in general.

This new, informatics-based, methodology provides a new way of thinking and a new language for sciences and technologies, extending the Galilean mathematics-based approach to new heights.

Informatics-driven methodology subsumes and extends the role and improves tools mathematics used to play in advising, guiding and serving other scientific and technology disciplines and society in general.

Power of new methodology will be discussed in the paper in details. Here are only few of the reasoning:

- It brings new dimension to both old methodologies;
- It brings into new heights an enormous power of modeling, simulations and visualisation;
- It utilises an enormous exploratory and discovery power of automata, algorithms and complexity considerations.
- It utilizes enormous discovery and exploratory power of the correctness and truth searching considerations and tools.
- It utilizes an enormous potential that the study of virtual worlds brings for understanding of the real world.

Because of its enormous guiding power for practically all areas of science, technology and the whole society and an enormously powerful tools informatics offers, we can see informatics as a new queen and at the same time a new powerful servant for all of society.

In particular informatics is expected to play the key role in dealing with two main megachallenges of current science, technology and society. Namely:

- **To beat natural human intelligence.** More exactly, to create super-powerful non-biological intelligence and its merge with biological intelligence.
- **To beat natural human death**. More exactly, to increase much longevity for human bodies and to achieve uploading for human minds. In more details, to fight natural death as another disease and to find ways to upload human mind to non-biological substrate.

There starts to be enough reasons to see the above megachallenges as being currently realistic enough. Here are some of them.

- Since information processing keeps developing exponentially we can assume to have soon (2045?) laptops with information processing power and capacity larger than of all human brains.

- Exponential scaling up in genetic and nanotechnologies and AI create a basis for making two megachallenges as already feasible ones.

- Exponential developments of information processing technologies are believed to imply enormous speed up developments in science and technology.

- Tools and efforts to reverse engineering brains keep also developing exponentially and so we can assume to have quite soon ways to simulate functionality of human brains.

- Society starts to put enormous effort to develop genome engineering, to model human brains and minds as well as to vastly extend human longevity.

- A vision starts to be accepted to see the development of superintelligent machines as the next stage of evolution.

Some related food for thoughts.

- There is nothing in biology found yet that indicates the inevitability of death.

  Richard Feynman

- It seems probable that once the machine thinking method had started, it will not take long to outstrip our feeble power. They would be able to converse with each other to sharpen their wits. At some stage therefore, we should have to expect machine to take control.

  Alan M. Turing

- Let an ultraintelligent machine be defined as a machine that can far surpass all intellectual activities of any man, however clever. Since the design of machines is one of intellectual activities, an ultraintelligent machine could design even better machines; there

would then unquestionably be and "intelligent explosion" and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man needs ever make.

I. J. Good, 1965, a British mathematician

- Since there is a real danger that computers will develop intelligence and take over we urgently need to develop direct connections to brains so that computers can add to human intelligence rather than be in opposition.

Stephen Hawking

Jozef Gruska

Faculty of Informatics, Masaryk University, Brno, Czech Republic
E-mail: gruska@fi.muni.cz

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Swarm Intelligence Algorithms Search Capabilities

Milan Tuba

Abstract

Swarm intelligence is an important branch of nature inspired algorithms where collective intelligence of different species like ants, bees, cuckoos, etc. is simulated. Even though swarm intelligence algorithms mimic very different phenomena, they all start with a collection of random points in the space of potential solutions and then search for better solutions, keeping balance between exploitation and exploration. We have recently successfully modified and improved several swarm intelligence algorithms. Here we present some of the techniques that we used for these modifications and improvements.

Keywords: nature inspired algorithms, swarm intelligence, optimization metaheuristics.

1 Introduction

Very interesting class of hard combinatorial and continuous optimization problems, with many practical applications, cannot be solved by deterministic, classical mathematical methods because of high irregularity and/or enormous number of local optima. Such problems have recently been successfully solved by nature inspired metaheuristic algorithms. The main idea of these algorithms is to guide random, Monte-Carlo search by simulating some successful systems from the nature. Swarm intelligence is an important branch of nature inspired algorithms where collective intelligence of different species like ants, bees, cuckoos,

 Ministry of Science of the Republic of Serbia has supported the research for this paper, Grant No. III-44006.

fireflies, bats, fish, birds, krill etc. is simulated. Even though swarm intelligence algorithms mimic very different phenomena, they all start with a collection of random points in the space of potential solutions and then search for better solutions keeping balance between exploitation and exploration. Exploitation utilizes already found good solutions to improve convergence speed, while exploration moves to new areas in order to avoid being trapped in local optima. Exploitation and exploration are performed by simulation of different natural processes: pheromone deposit and evaporation in case of ants (Ant Colony Optimization – ACO), honey bee's dance in the hive that communicates information about food source quality (Artificial Bee Colony – ABC), light attraction of fireflies (Firefly Algorithm – FA), Levi flight of cuckoo birds (Cuckoo Search – CS) etc. However, mathematically both exploitation and exploration are always determined by the way in which new candidate solution is generated and evaluated. We have recently successfully modified and improved several swarm intelligence algorithms and here we present some of the techniques that we used for these modifications and improvements.

2 Swarm Intelligence Algorithms Improvements

We introduced specific pheromone correction strategy for avoiding stagnation in the local optima to the ACO algorithm and improved results for the Minimum Weight Vertex Cover Problem (MWVCP) [1], for the Minimum Connected Dominating Set Problem [2] and for the Traveling Salesman Problem [3]. The idea is to use the information about the best-found solution to perform some corrections on the pheromone trail. The concept of this hybridization is suspicion that some elements of the global best solution are not good. In our hybridization we direct to new search areas that are less suspicious i.e. with less undesirable properties. Directing the search is done by greatly decreasing the pheromone trail values at suspicious points. A big drawback of minimum pheromone threshold strategy (MPTS) is that added vertices are

chosen just for having low values of pheromone which is a relatively random process. As a result, vertices that do not belong to good solutions are often reintroduced in the search. In our hybridization however, we do not add vertices to the "popular" group but rather remove vertices with suspicious properties, making the group smaller. In the following iterations ants will first select vertices from the popular set, and when none are left, vertices with better properties. This way we direct to new search areas with less undesirable properties.

We introduced different modifications to the ABC algorithm to improve balance between exploitation and exploration and improved results for engineering optimization problems [4], and general constrained problems [5]. In our upgraded artificial bee colony (UABC) algorithm [4] in the onlooker phase (exploitation) we reduce the diversity, as may be expected, by reducing the space where new generated solution can be. In the original ABC algorithm new solution is generated from the current solution and one random solution so that new solution is inside the hyper-cube where these two old solutions are located at the opposing points of the cube's diagonal, while in our UABC algorithm only points on the diagonal, not from the whole cube, are permitted as new solutions. On the other hand, original ABC algorithm lacks some diversity (or courage to search towards infeasible points) because in order to handle the constraints of the problem it employs Deb's rules and consequently, infeasible points are heavily suppressed. Additionally, in the ABC algorithm, probability for selection by onlookers is calculated in such a way that all feasible solutions have that probability above 0.5 and all infeasible solutions below 0.5. Our UABC algorithm also employs Deb's rules, but we found that algorithm works better if mentioned discrimination is removed and infeasible solutions get better chance of being selected by being judged only on their fitness value, not inversely proportional to their constraint violation values. This increases diversity in onlooker phase, but only in that specific direction of less discriminating infeasible solutions, which in no way contradicts previously mentioned reduced diversity by allowing new solutions only on the diagonal of the hyper-cube; just the opposite: both modifica-

tions work together towards faster convergence.

We also improved Seeker optimization algorithm (SOA) by hybridization with FA for better performance [6].

References

[1] R. Jovanovic, M. Tuba. *An ant colony optimization algorithm with improved pheromone correction strategy for the minimum weight vertex cover problem*. Applied Soft Computing 11 (8) (2011) 5360–5366.

[2] R. Jovanovic, M. Tuba. *Ant colony optimization algorithm with pheromone correction strategy for the minimum connected dominating set problem*. Computer Science and Information Systems (ComSIS) 10 (1) (2013) 133–149.

[3] M. Tuba, R. Jovanovic. *Improved ant colony optimization algorithm with pheromone correction strategy for the traveling salesman problem*. International Journal of Computers, Communications & Control 8 (3) (2013) 409–417.

[4] I. Brajevic, M. Tuba. *An upgraded artificial bee colony algorithm (ABC) for constrained optimization problems*. Journal of Intelligent Manufacturing, published Online First (2012) DOI:10.1007/s10845–011–0621–6.

[5] N. Bacanin, M. Tuba. *Artificial bee colony (ABC) algorithm for constrained optimization improved with genetic operators*. Studies in Informatics and Control 21 (2) (2012) 137–146.

[6] M. Tuba, I. Brajevic, R. Jovanovic. *Hybrid seeker optimization algorithm for global optimization*. Applied Mathematics and Information Sciences 7 (3) (2013) 867–875.

Milan Tuba

Graduate School of Computer Science, Megatrend University of Belgrade, SERBIA
E-mail: tuba@ieee.org

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Wiki-Translator: Multilingual Experiments for In-Domain Translations

Dan Tufiş, Radu Ion, Ştefan Daniel Dumitrescu

Abstract

The benefits of using comparable corpora for improving translation quality for statistical machine translators have been already shown by various researchers. The usual approach is starting with a baseline system, trained on out-of-domain parallel corpora, followed by its adaptation to the domain in which new translations are needed. The adaptation to a new domain, especially for a narrow one, is based on data extracted from comparable corpora from the new domain or from an as close as possible one. This article reports on a slightly different approach: building an SMT system entirely from comparable data for the domain of interest. Certainly, the approach is feasible if the comparable corpora are large enough to extract SMT useful data in sufficient quantities for a reliable training. The more comparable corpora, the better the results are. Wikipedia is definitely a very good candidate for such an experiment. We report on mass experiments showing significant improvements over a baseline system build from highly similar (almost parallel) text fragments extracted from Wikipedia. The improvements, statistically significant, are related to what we call the level of translational similarity between extracted pairs of sentences. The experiments were performed for three language pairs: Spanish-English, German-English and Romanian-English, based on sentence pairs extracted from the entire dumps of Wikipedia as of June 2012. Our experiments and comparison with similar work show that adding indiscriminately more data to a training corpus is not necessarily a good thing in SMT.

Keywords: comparable corpora, extraction of parallel sentences, language model, statistical machine translation, translation models.

1 Introduction

The research on domain adaptation based on comparable corpora has been motivated by the scarce parallel data for most of the language pairs or by the scarcity of (narrow) domain specific parallel data. The standard approach is to start with a baseline system, trained on as much as possible out-of-domain parallel corpora, followed by its adaptation to the domain in which new translations are needed. To date, OPUS[1] (Tiedemann, 2012) is the largest **online** collection of parallel corpora, comprising juridical texts (EUROPARL and EUconst), medical texts (EMEA), technical texts (e.g. software KDE manuals, PHP manuals), movie subtitles corpora (e.g. OpenSubs), translated transcribed talks (e.g. TED) or news (SETIMES) but these corpora are not available for all language pairs nor their sizes are similar with respect to the domain. Another example of a large collection of aligned parallel texts (in 22 languages) is JRC-Acquis (Steinberger et al., 2006), the total body of European Union (EU) law applicable in the EU Member States.

The adaptation to a new domain, especially for a narrow one, is based on data extracted from comparable corpora from the new domain or from an as close as possible one.

This article[2] reports on slightly different approach: building an SMT system entirely from comparable data for the domain of interest. The approach is feasible if the comparable corpora are large enough to extract SMT useful data in sufficient quantities for a reliable training. If the corpora, from which the translation-useful data are searched for, are strongly comparable, the outcomes may be surprisingly good.

[1] http://opus.lingfil.uu.se/

[2] A preliminary version of the results have been described in (Tufiş et al, 2013); here we bring more and new experimental results and comments.

Wikipedia is definitely a very good candidate for such an experiment. Wikipedia is not a real parallel corpus, but a strongly comparable multilingual corpus with many documents in different languages representing translations from (mainly) English. More often than not, the documents in one language are shortened or adapted translations of documents from other (not always the same) languages and this property of Wikipedia together with its size makes it the ideal candidate of a strongly comparable corpus from which parallel sentences can be mined.

SMT engines like Moses[3] produce better translations when presented with larger and larger parallel corpora. In this context, large and good quality parallel corpora extracted from Wikipedia for different language pairs, can serve three purposes:

1. provide in-domain training data for aiding automatic translation of English (or other languages) Wikipedia articles into other languages thus paving the way to growing for poorer foreign Wikipedia sites;

2. provide in-domain training data for aiding automatic translation of Wikipedia non-English articles thus helping the dissemination of other nations' cultural and scientific contributions;

3. add a new domain (for many language pairs), the encyclopedic domain, to the list of domains for which parallel data already exist;

The structure of the article is as follows: we begin with a short review of related research (Section 2), continue with an informal description of the tool that was used to collect the parallel sentences from Wikipedia (Section 3). In Section 4, we describe the two-steps methodology for data extraction and provide quantitative data about the obtained parallel corpora for the English-Spanish, English-German and English-Romanian language pairs. We also provide BLEU evaluation of SMT

[3] http://www.statmt.org/moses/

using extracted data. Next, in Section 5 we compare our experiments with similar ones. The last section draws some conclusions and presents future plans.

2 Related work

Adafre and Rijke (2006) were among the first to attempt extraction of parallel sentences from Wikipedia. Their approach consists of two experiments: 1) the use of a MT system (Babelfish) to translate from English to Dutch and then, by word overlapping, to measure the similarity between the translated sentences and the original sentences and 2) with an automatically induced (phrase) translation lexicon from the titles of the linked articles, they measure the similarity of source (English) and target (Dutch) sentences by mapping them to (multiple) entries in the lexicon and computing lexicon entry overlap. Experiments were performed on 30 randomly selected English-Dutch document pairs yielding a few hundred parallel sentence pairs.

Mohammadi and GhasemAghaee (2010) continue the work of Adafre and Rijke (2006) by imposing certain limits on the sentence pairs that can be formed from a Wikipedia document pair: the length of the parallel sentence candidates must correlate and the Jaccard similarity of the lexicon entries (seen as IDs) mapped to source (Persian) and target (English) must be as high as possible. As with Adafre and Rijke, the work performed by Mohammadi and GhasemAghaee does not actually generate a parallel corpus but only a couple hundred parallel sentences intended as a proof of concept.

Another experiment, due to Smith et al. (2010), addressed large-scale parallel sentence mining from Wikipedia. They automatically extracted large volumes of parallel English-Spanish (almost 2M pairs), English-German (almost 1.7M pairs) and English-Bulgarian (more than 145K pairs) sentences using binary Maximum Entropy classifiers (Munteanu and Marcu, 2005). The work of Smith et al. (2010) is the only one we are aware of, which extracted parallel corpora of similar sizes to ours. They released their Wikipedia test sets for English-

Spanish (500 pairs) and for English-German (314 pairs), an inescapable opportunity for a direct comparison between our results and theirs. This comparison is documented in the Section 5.

3 Extracting bilingual comparable translation units

The EU project ACCURAT[4] (2010-2013) collected from the web very large sets of comparable documents and classified them (using a specially designed metrics) into different comparability classes: strongly comparable, comparable, weakly comparable and unrelated documents (see the public site for detailed reports and the associated data). From different comparability classes, various text mining systems, developed within the project, extracted useful MT data (highly similar cross-lingual sentences and parallel terms and name entities) which subsequently were used for assessing the impact on the translation quality of the existing baseline systems (Skadia et al., 2012; Tufiş, 2012).

For the experiments described in this article we used one of the ACCURAT text miners, namely LEXACC (Ştefănescu et al., 2012). It is a fast parallel sentence mining algorithm from comparable corpora, developed to handle large amounts of comparable corpora in a reasonable amount of time. Unlike most text-miners based on binary classifiers (e.g. Munteanu and Marcu (2005)) which do not make the distinction between truly parallel sentences, partial parallel sentences, strongly comparable or weakly comparable sentences (or other, finer degrees of parallelism), LEXACC uses a similarity metrics allowing for ranking translation candidate pairs according to their similarity scores (with values continuously ranging from a very low number assigned to unrelated sentences to a very high number assigned to truly parallel sentences).

In order to dramatically reduce the search space, LEXACC uses

[4]http://www.accurat-project.eu/

Lucene[5] to index the entire collection of target sentences (storing the document pair ID with each sentence). Using CLIR techniques the candidate sentence pairs are subject to several restricting filters (e.g. the length of the source and target sentence candidates must correlate, a high proportion of the source sentence content words must have a translation in the target candidate, etc.). In order to extract more informative sentence pairs, LEXACC filters out titles or short sentences (with less than 3 words). Once this fast initial filtering is finished, a second step, computationally more expensive, generates the final similarity ranking of the translation pairs and leaves out all the pairs with a score below a pre-established threshold.

The translation similarity measure is a weighted sum of feature functions that indicate if the source piece of text is translated by the target. Given two sentences, s in the source language and t in the target language, then the translation similarity measure $P(s,t)$ is

$$P(s,t) = \sum_i \theta_i f_i(s,t) \tag{1}$$

such that $\sum_i \theta_i = 1$. Each feature function $f_i(s,t)$ returns a real value between 0 (s and t are not related at all) and 1 (t is a translation of s) and contributes to the overall parallelism score with a specific fraction θ_i that is language-pair dependent and that is automatically determined by training a logistic regression classifier on existing parallel data in both directions: source-target and target-source. It follows that the translation similarity measure possible values are between 0 (s and t are not related at all) and 1 (t is a translation of s).

Some of the features used by the translation similarity measure in equation 1 are as follows (for a detailed description of these features, the reader is directed to (Ştefănescu et al., 2012)):

- the content words translation strength (i.e. the score of the best alignment between content words of s and t);

[5] http://lucene.apache.org/

- the functional words translation strength (i.e. the score of the best alignment of functional words near a strong alignment link of content words);

- alignment obliqueness (i.e. the score of a content word alignment whose links do not cross is larger than the score of an alignment with crossing links);

- the *sentinel* translations feature: we noticed that, more often than not, two parallel sentences begin and end with strongly related content words even if words in the middle are not found in the lexicon.

In order to use LEXACC for mining useful MT sentence pairs one needs a translation lexicon. Ideally, this lexicon should be domain-specific, with a large lexical coverage of the search space. However, this requirement, difficult to meet, may be avoided using a boosting technique: use any available bilingual lexicon or extract a translation lexicon from whatever parallel corpora; run LEXACC on the in-domain comparable corpus and use the mined sentence pairs for extracting better in-domain lexicons; redo the sentence pair extraction. The boosting procedure may be repeated a number of times until no improvements are observed. Yet, one has to consider that the entire chain of processing is highly computational intensive, and depending on the size of the search space (as is the case of large Wikipedias) it may take several weeks.

4 Mining Wikipedia

Among the 285 language editions of Wikipedia (http://meta.wikimedia.org/wiki/List_of_Wikipedias[6]) created under the auspices of Wikimedia Foundation, the English, German and Spanish ones are listed in the best populated category, with more than 1,000,000 articles:

[6]Consulted on May 22^{nd} 2013:

English is the largest collection with 4,238,043 articles, German is the second largest with 1,587,660 articles while Spanish is the 6^{th} in the top Wikipedias with 1,017,938 articles. Romanian Wikipedia is in the medium populated category, and with 226,004 articles is the 25^{th} largest collection. For our experiments we selected three very large Wikipedias (English, German and Spanish) and a medium sized Wikipedia (Romanian) and performed SMT experiments for three language pairs: English-German, English-Spanish and English-Romanian.

With these monolingual Wikipedias selected for parallel sentence mining, we downloaded (December 22^{nd}, 2012) the "database backup dumps"[7] for which Wikipedia states that they contain "a complete copy of all Wikimedia wikis, in the form of wikitext source and metadata embedded in XML". Parsing the English XML dump, we kept only the proper encyclopedic articles which contain links to their corresponding articles in the Spanish, German or Romanian. Thus, we removed articles that were *talks* (e.g. Talk:Atlas Shrugged), *logs* (e.g. Wikipedia:Deletion log), *user related articles* (e.g. User:AnonymousCoward), *membership related articles* (e.g. Wikipedia:Building Wikipedia membership), *manuals* and *rules related articles*, etc.

For each language, the retained articles were processed using regular expressions to remove the XML mark-up in order to keep only the raw, UTF-8 encoded text, which was saved into a separate file. The non-textual entries like images or tables were stripped off. Each text document was then sentence-split using an in-house freely available[8] sentence splitter based on a Maximum Entropy classifier.

Table 1 lists the number of sentence-split Wikipedia comparable document pairs (by following the inter-lingual links) for each considered language pair (see (Ştefănescu and Ion, 2013) for further details).

Looking at the size ratio of the linked documents for each language pair it is apparent that Romanian documents are much shorter than the linked English ones. The size ratios for other language pairs are

[7] http://dumps.wikimedia.org/backup-index.html

[8] http://nlptools.racai.ro/nlptools/index.php?page=ssplit

Table 1. Linked documents for three language pairs

Language pair	**Document pairs**	**Size on disk**	**Size ratio (L1/L2)**
English-German	715,555	2.8 Gb (English)	1,22
		2.3 Gb (German)	
English-Romanian	122,532	778.1 Mb	3,91
		198.9 Mb	
English-Spanish	573,771	2.5 Gb	1,66
		1.5 Gb	

more balanced, coming closer to expected language specific ratio for a parallel corpus (see below).

We applied the boosting procedure as follows:

1. We used the JRC-Acquis parallel corpora to extract initial translation lexicons for English-Romanian and English-German language pairs. For English-Spanish pair we used the corresponding parallel sub-part of EUROPARL. We run GIZA++ (Gao and Vogel, 2008) and symmetrized the extracted translation lexicons between the source and target languages. The Romanian-English lexicon extracted with GIZA++ was merged with an in-house dictionary generated from our wordnet (Tufiş et al., 2013) aligned to Princeton WordNet. With these lexicons we performed the first phase of LEXACC extraction of comparable sentence pairs from the respective Wikipedias. Let us call this data-set, for a language pair L1-L2, as Wiki-Base (L1, L2). The experiments with Wiki-Base for three language pairs are described in Section 4.2;

2. From the most MT useful parts of Wiki-Base(L1, L2), as resulted from the first step, we extracted new translation lexicons used for a second phase (the boosting) of LEXACC (symmetrized) extraction, thus getting a new and larger data set which we refer

to as Wiki-Train (L1,L2). The most useful parts of Wiki-Train were identified based on their impact on the BLEU score for the test set as described in Section 4.3 and used for the training of the Wiki-Translators.

4.1 Building Wiki-Base (L1,L2)

Table 2 lists, for different similarity scores as extraction thresholds, the number of MT useful sentence pairs (P) found in each language pair dataset, as well as the number of words (ignoring punctuation) per language (English Words, German Words, Romanian Words, Spanish Words) in the respective sets of sentence pairs. Obviously, data extracted with a given Similarity score threshold was a proper sub-set of any data extracted with a lower Similarity score threshold.

Table 2: Wiki-base: number of parallel sentences and words for each language pair, for a given threshold

Similarity score	**English-Romanian**	**English-German**	**English-Spanish**
0.9	Pairs: 42,201 English Words: 0.814 M Romanian Words: 0.828 M	Pairs: 38,390 English Words: 0.554 M German Words: 0.543 M	Pairs: 91,630 English Words: 1.126 M Spanish Words: 1.158 M
0.8	Pairs: 112,341 English Words: 2.356 M Romanian Words: 2.399 M	Pairs: 119,480 English Words: 2.077 M German Words: 2.010 M	Pairs: 576,179 English Words: 10.504 M Spanish Words: 11.285 M

Continuation of Table 2

Similarity score	**English-Romanian**	**English-German**	**English-Spanish**
0.7	Pairs: 142,512 English Words: 2.987 M Romanian Words: 3.036 M	Pairs: 190,135 English Words: 3.494 M German Words: 3.371 M	Pairs: 1,219,866 English Words: 23.730 M Spanish Words: 25.931 M
0.6	Pairs: 169,662 English Words: 3.577 M Romanian Words: 3.634 M	Pairs: 255,128 English Words: 4.891 M German Words: 4.698 M	Pairs: 1,579,692 English Words: 31.022 M Spanish Words: 33.706 M
0.5	Pairs: 201,263 English Words: 4.262 M Romanian Words: 4.325 M	Pairs: 322,011 English Words: 6.453 M German Words: 6.186 M	Pairs: 1,838,794 English Words: 36.512 M Spanish Words: 39.545 M
0.4	Pairs: 252,203 English Words: 5.415 M Romanian Words: 5.482 M	Pairs: 412,608 English Words: 8.470 M German Words:8.132 M	Pairs: 2,102,025 English Words: 42.316 M Spanish Words: 45.565 M
0.3	Pairs: 317,238 English Words: 6.886 M Romanian Words: 6.963 M	Pairs: 559,235 English Words: 11.797 M German Words: 11.353M	Pairs: 2,656,915 English Words: 54.932 M Spanish Words: 58.524 M

Depending on the similarity threshold, the extracted pairs of sentences may be really parallel, may contain real parallel fragments, may be similar in meaning but with a different wording, or lexically unrelated in spite of domain similarity. That is, the lower the threshold, the higher the noise.

By random manual inspection of the generated sentence pairs, we saw that, in general, irrespective of the language pair, sentence pairs with a translation similarity measure higher than 0.6 are parallel. Based on the number of words in each language side of the parallel extracted sentences, one can easily compute an expected average length ratio for the three considered language pairs. Those pairs with a translation similarity measure of at least 0.5 have extended parallel fragments which an accurate word or phrase aligner easily detects. Further down the threshold scale, below 0.3, we usually find sentences that roughly speak of the same event but are not actual translations of each other. The noisiest data sets were extracted for the 0.1 and 0.2 similarity thresholds and we drop them from further experiments.

If we consider the extraction rate ExtR as the ratio between the number of parallel sentences (those with similarity score higher or equal to 0.7) and the number of linked documents we get the following figures:

$$\mathrm{ExtR}(\mathrm{En}-\mathrm{Ro}) = 1.16; \quad \mathrm{ExtR}(\mathrm{En}-\mathrm{De}) = 0.26;$$

$$\mathrm{ExtR}(\mathrm{En}-\mathrm{Es}) = 2.12.$$

The striking differences may have several explanations. The first one is that the Spanish documents linked to English documents are more literary translated, while the German documents are more distant from the English documents to which they are linked. Romanian documents are somewhere in between. Another partial explanation might be the quality of the dictionaries LEXACC used for each language. Augmenting the Romanian-English lexicon extracted by GIZA++ from JRC-Acquis with the data from wordnet resulted in a cleaner (although smaller) dictionary than the German-English extracted also from JRC-Acquis. In case of Spanish-English extraction rate (higher than for other two language pairs) we hypothesize that the GIZA++ dictionary

extracted from EUROPARL has a better covering of the Wikipedia vocabulary. The experimental results described in the following sections strongly support these hypotheses.

4.2 SMT experiments with Wiki-Base

In order to select the most MT useful parts of Wiki-Base for the three considered language pairs, we built three baseline Moses-based SMT systems using only parallel sentences, that is those pairs extracted with a similarity score higher or equal to 0.7 (see Table 2). We incrementally extended the training data by lowering the similarity score threshold and, using the same test-set, observed the variation of the BLEU score. The purpose for the evaluation of the SMT systems was only to indicate what would be the best threshold for selecting the training set from the Wiki-Train for building the Wiki-Translators. As the standard SMT system we chose Moses surface-to surface translation and lexical reordering model with parameters wbe-msd-bidirectional-fe, with phrase-length of maximum 4 words, and the default values for the rest of parameters.

The **language model** (LM) for all experiments was trained on all monolingual, sentence-split English Wikipedia after removing the administrative articles as described in Section 3. The language model was limited to 5-grams and the counts were smoothed by the interpolated Knesser-Ney method.

Since we experimentally noticed that the additional sentence pairs extracted for a threshold of 0.6 were almost as parallel as those extracted for higher thresholds we included this interval too in the sampling process for **test set** design. Thus, we proceeded to randomly sample 2,500 sentence pairs from similarity intervals ensuring parallelism ([0.6, 0.7), [0.7-0.8), [0.8, 0.9) and [0.9-1]). We obtained 10,000 parallel sentence pairs for each language pair. Additionally, we extracted 1,000 sentence pairs as development set (**dev set**). These 11,000 sentences were removed from the training corpora of each language pair. When sampling parallel sentence pairs, we were careful to obey the Moses' filtering constraints: both the source and target sentences must have

at least 4 words and at most 60 words and the ratio of the longer sentence (in tokens) of the pair over the shorter one must not exceed 2. The duplicates were also removed.

Further on, we trained **seven translation models** (TM), for each language pair, over cumulative threshold intervals beginning with 0.3: TM_1 for [0.3, 1], TM_2 for [0.4, 1] ..., TM_7 for [0.9, 1]. The resulting eight training corpora have been filtered with Moses' cleaning script with the same restrictions mentioned above. For every language, both the training corpora and the test set have been tokenized using Moses' tokenizer script and true-cased. The quality of the translation systems is measured as usual in terms of their BLUE score (Papineni et al., 2002) on the same test data.

We have to emphasize that the removal of the sentences in the test and development sets from the training corpora does not ensure an unbiased evaluation of the BLEU scores since their context still remained in the training corpora. This requires some more explanations. For each extracted sentence pair, LEXACC stores in a book-keeping file, the ID of the document-pair out of which the extraction was done. This information allows for elimination from the training set of all the pairs coming from the same documents from which the development and evaluation sets were selected. However, due to the nature of the Wikipedia article authoring, this strategy of filtering the development and evaluation sets does not ensure an unbiased evaluation. The Wikipedia contributors are given specific instructions for authoring documents[9] and by observing these instructions, inherently one could find in different documents almost identical sentences except for a few name entities. Indeed we found examples of such sentence pairs in the train set similar, but not identical, to sentences in the test set, yet coming from different document-pairs. Certainly one could build a tough test-set by removing from train set all similar (pattern-based) sentences, but we did not do that because it would have been beyond the purpose of this work. As we mentioned before, this evaluation was meant only for estimating most useful extraction level for the second

[9] http://en.wikipedia.org/wiki/Wikipedia:Translation

phase of training the WIKI-Translators.

Table 3 summarizes the results of this first step experiment, the bold characters identifying the most MT useful parts of Wiki-Base (L1,L2). We considered TM $_{[0.7,1]}$ (the shaded line in Table 3) as the baseline for all language pairs.

Table 3. Comparison between SMT systems trained on various parts of Wiki-Base

TM based on Wiki-Base	**BLEU SCORE Romanian –> English**	**BLEU SCORE German –> English**	**BLEU SCORE Spanish –> English**
$TM_{[0.3,1]}$	37.24	39.16	**47.59**
$TM_{[0.4,1]}$	37.71	39.46	47.52
$TM_{[0.5,1]}$	**37.99**	**39.52**	47.53
$TM_{[0.6,1]}$	37.85	39.5	47.44
$TM_{[0.7,1]}$	37.39	39.24	47.28
$TM_{[0.8,1]}$	36.89	38.57	46.27
$TM_{[0.9,1]}$	32.76	34.73	39.68

4.3 Building Wiki-Train (L1,L2)

The experiments on Wiki-base revealed that the most useful training data has been extracted by using LEXACC with 0.5 similarity score for German-English and Romanian-English language pairs and 0.3 for Spanish-English pair (see Table 3). We re-run GIZA++ on these subsets of Wiki-Base to extract new in-domain lexicons.

The new lexicons were merged with the initial ones and the LEXACC extraction was repeated with the resulted mined comparable sentence-pairs denoted as Wiki-Train.

As the experiments on the Wiki-base showed that for a similarity threshold less than or equal to 0.2 LEXACC delivers not very useful

data, we started the second step of mining using the similarity scores of at least 0.3.

Table 4 shows the results of the boosted extraction process. As one can see the extracted data, at each similarity score level, is significantly increased for the English-Romanian and English-German language pairs. For English-Spanish, except for the similarity scores 0.8 and 0.9 the number of sentence pairs is smaller than in Wiki-Base. The reason is that in this round we detected several identical pairs with those in the training and development sets and several duplicated pairs in the training set. Anyway, the English-Spanish Wiki-Train was the largest train-set and containing the highest percentage of fully parallel sentence pairs.

Table 4: Wiki-Train: number of parallel sentences and words for each language pair, for a given threshold

Similarity score	**English-Romanian**	**English-German**	**English-Spanish**
0.9	Pairs: 66,777 English Words: 1.077 M Romanian Words: 1.085 M	Pairs: 97,930 English Words: 1.069 M German Words: 1.042 M	Pairs: 113,946 English Words: 1.164 M Spanish Words: 1.193 M
0.8	Pairs: 152,015 English Words: 2.688 M Romanian Words: 2.698 M	Pairs: 272,358 English Words: 3.695M German Words: 3.552 M	Pairs: 597,992 English Words: 9.733 M Spanish Words: 10.510 M
0.7	Pairs: 189,875 English Words: 3.364 M Romanian Words: 3.372 M	Pairs: 434,019 English Words: 6.201 M German Words: 5,929 M	Pairs: 1,122,379 English Words: 19.941 M Spanish Words: 21.821 M

Continuation of Table 4

Similarity score	**English-Romanian**	**English-German**	**English-Spanish**
0.6	Pairs: 221,661 English Words: 3.961 M Romanian Words: 3.970 M	Pairs: 611,868 English Words: 8.944 M German Words8.532 M	Pairs: 1,393,444 English Words: 25.068 M Spanish Words: 27.411 M
0.5	Pairs: 260,287 English Words: 4,715 M Romanian Words: 4,722 M	Pairs: 814,041 English Words: 12.361 M German Words: 11.792M	Pairs: 1,587,276 English Words: 28.987 M Spanish Words: 31.567 M
0.4	Pairs: 335,615 English Words: 6.329 M Romanian Words: 6.324 M	Pairs: 1,136,734 English Words: 18,089 M German Words: 17.306 M	Pairs: 1,807,892 English Words: 33.619 M Spanish Words: 36,369 M
0.3	Pairs: 444,102 English Words: 8.712 M Romanian Words: 8.700 M	Pairs: 1,848,651 English Words: 31.405 M German Words: 30.175 M	Pairs: 2,288,163 English Words: 44.021 M Spanish Words: 47.180 M

4.4 SMT experiments with Wiki-Train

The Wiki-Train corpora were used with the same experimental setup as described in Section 4.2. The training of each translation system was followed by the evaluation on the respective test sets (10,000 pairs) in both translation directions. To make the comparison between the

translation qualities we did the translations without MERT optimization of the parameters. The results are presented in Table 5.

Having much more training data, in case of the Romanian –> English and German –>English the BLEU scores significantly increased (with 3.1 and 2.58 points respectively). For Spanish-English the decrease of number of sentences in Wiki-Train as compared to Wiki-Base negatively impacted the new BLEU score, which is 1.31 points lower. It would be interesting to see what would happen with a higher threshold training set, for instance $TM_{[0.5,1]}$, as used for the other language pairs.

As expected, the translations into non-English languages are less accurate due to a more complex morphology of the target language (most of the errors are morphological ones), but still the BLEU scores are very high, better than most of the results we are aware off (for in-domain experiments).

Table 5. Best translation SMT systems, trained on Wiki-Train[10]

TM based on Wiki-Train	**$TM_{[0.5,1]}$ Romanian –> English**	**$TM_{[0.5,1]}$ German –> English**	**$TM_{[0.3,1]}$ Spanish –> English**
BLEU SCORE	**41.09**	**40.82**	**46.28**
	$TM_{[0.5,1]}$ English –> Romanian	**$TM_{[0.5,1]}$ English –> German**	**$TM_{[0.3,1]}$ English –> Spanish**
BLEU SCORE	**29.61**	**35.18**	**46.00**

[10]For a fair comparison with data in Table 3 we did not used here the MERT optimization

5 Comparison with other works

Translation for Romanian-English language pair has also been studied in (Boroş et al., 2013; Dumitrescu et al., 2012; 2013) among others. In these works we had explicit interests in experiments on using in-domain/out-of-domain test/train data, and various configurations of the Moses decoder in surface-to-surface and factored translation. Out of the seven domain-specific corpora (Boroş et al., 2013) one was based on Wikipedia. The translation experiments on English-Romanian, similar to those reported here, were surface based (t0-0, m0) with training on parallel sentence pairs extracted from Wikipedia by LEXACC at a fixed threshold: 0.5 (called "WIKI5"), without MERT optimization. A random selection of unseen 1,000 Wikipedia Romanian test sentences[11] has been translated into English using combinations of:

- a WIKI5-based translation model (240K sentence pairs)/WIKI5-based language model;
- a global translation model (1.7M sentence pairs)/global language model named "ALL", made by concatenating all specific corpora.

Table 6 gives the BLEU scores for the Moses configuration similar to ours.

Table 6. BLEU scores on 1000 sentences Wikipedia test set of Dumitrescu et al. (2013)

	WIKI5 TM	**ALL TM**
WIKI5 LM	**29.99**	29.95
ALL LM	29.51	29.95

Boroş et al.'s results confirm the conclusion we claimed earlier: the ALL system performs worse than the in-domain WIKI5 system. The

[11] The test-set construction followed the same methodology described in this article

large difference between the herein BLEU score (41.09) and 29.99 in (Boroş et al., 2013) may be explained by various factors. First and more importantly, our current language model was entirely in-domain for the test data and much larger: the language model was built from entire Romanian Wikipedia (more than 220,000 documents) while the language model in (Boroş et al., 2013) was built only from the Romanian sentences paired to English sentences (less than 240,000 sentences). Our translation model was built from more than 260,000 sentence pairs versus 234,879 sentence pairs of WIKI5). Another explanation might be the use of different Moses filtering parameters (e.g. the length filtering parameters) and different test sets. As suggested by other researchers, Wikipedia-like documents are more difficult to translate than, for instance, legal texts. The BLEU scores on JRC-Acquis test sets (with domain specific training) reported in (Boroş et al., 2013) is almost double than those obtained on Wikipedia test sets.

The most similar experiments to ours have been reported by Smith et al. (2010). They mined for parallel sentences from Wikipedia producing parallel corpora of sizes even larger than ours. While they used for training all the extracted sentence pairs, we used only those subsets that observed a minimal similarity score. We checked to see if their test sets for English-Spanish (500 pairs) and for English-German (314 pairs) contained sentences in our training sets and, as this was the case, we eliminated from the training several sentence pairs (about 200 sentence pairs from the English-Spanish training corpus and about 140 sentence pairs from the English-German training corpus). We retrained the two systems on the slightly modified training corpora. Since in their experiments they used MERT-optimized translation systems, we optimized, also by MERT, our new $TM_{[0.5,1]}$ for German–>English and new $TM_{[0.3,1]}$ for Spanish–>English translation systems, using the respective dev-sets (each containing 1,000 sentence pairs).

Their test sets for English-Spanish and for English-German were translated (after being true-cased) with our best translation models and also with Google Translate (as of mid-February 2013).

Table 7 summarizes the results. In this table, "Large+Wiki" denotes the best translation model of Smith et al. which was trained on many corpora (including Europarl and JRC Acquis) and on more than 1.5M parallel sentences mined from Wikipedia. "$TM_{[0.4,1]}$" and "$TM_{[0.5,1]}$" are our Wiki-Train translation models as already explained. "Train data size" gives the size of training corpora in multiples of 1,000 sentence pairs.

Table 7. Comparison between SMT systems on the Wikipedia test set provided by Smith et al. (2010)

Language pair	**Train data size (sentence pairs)**	**System**	**BLEU**
Spanish-English	9,642K	Large+Wiki	43.30
	2,288K	$TM_{[0.4,1]}$	**50.19**
	–	Google	44.43
German-English	8,388K	Large+Wiki	23.30
	814K	$TM_{[0.5,1]}$	24.64
	–	Google	21.64

For Spanish-English test set of Smith et al. (2010) our result is significantly better than theirs, in spite of almost 4 times less training data. For the German-English pair, the difference is larger between $TM_{[0.5,1]}$ and Large+Wiki systems, and one should also notice that our system used 10 times less training data (but, presumably, much cleaner).

However, our $TM_{[0.5,1]}$ for German-English performed on the new test set much worse than on our test-set (24.64 versus 40.82[12] BLEU points) which was not the case for the Spanish-English language pair. We suspected that some German-English translation pairs in the Smith et al. (2010) test set were not entirely parallel. This idea was supported

[12]Note that this value for our $TM_{[0.5,1]}$ was obtained on a very different and much larger test set and also without MERT optimization. Yet, the difference is large enough to raise suspicions on the test set used for this comparison.

by the correlation of the evaluation results between our translations and Google's for Spanish-English and German-English. Also, their reported results on German-English were almost half of the ones they obtained for Spanish-English.

Therefore, we checked the German-English and Spanish-English test sets (supposed to be parallel) by running the LEXACC miner to see the similarity scores for the paired sentences. The results confirmed our guess. The first observation was that the test sets contained pieces of texts that looked like section titles (e.g. BT : Contaminación – BT : Pollution; Segunda clase – Second class; Autoengano-Self-deception, in Spanish – English test-set or Städte – Cities and towns; 1956 Armagnac – 1956 Armagnac; Produkte – Products; Geschwindigkeitsrekorde - Speed records; Geschichte – History in the German-English test-set). Such short sentences were ignored by LEXACC. While out of the considered sentence pairs (ignoring the sentences with less than 3 words), for Spanish-English LEXACC identified more than 92% as potentially useful SMT pairs (with a similarity score higher than or equal to 0.3 – this was the extraction threshold for Spanish-English sentence-pairs), for German-English LEXACC identified only 35% potentially useful SMT pairs (a similarity score higher than or equal to 0.5 – this was the extraction threshold for German-English sentence-pairs). Even if the threshold for German-English was lowered to 0.3, only 45% passed the LEXACC filtering. As for parallelism status of the sentence pairs in the test-sets (i.e. similarity scores higher than 0.6 for both languages) the percentages were 78% for Spanish-English and only 29% for German-English. Without ignoring the short sentences (easy to translate) these percentages would have probably been a little bit higher (80.8% for Spanish-English and 32.82% for German-English).

These evaluations outline also that LEXACC is too conservative in its rankings: we noticed almost parallel sentences in the test-set for Spanish-English even for a similarity score of 0.1 while in the German-English the same happens for similarity scores lower than 0.3. The most plausible explanation was that one of the LEXACC's parameters (cross-linking factor) strongly discourages long-distance reordering

(which was quite frequent in the German-English test set and has also a few instances in the Spanish-English test set).

Table 8 shows some examples of sentence pairs in the German-English and Spanish-English test sets showing low level of parallelism (inappropriate for translation quality evaluation) but also some examples of sentence pairs which were conservatively lower ranked by LEXACC.

Table 8: Examples of sentence-pairs in the German-English and Spanish-English test sets used by Smith et al. (2010)

Similarity	**German source sentence**	**English reference translation**
1	**2**	**3**
< 0.1	Zuletzt stand sie für Robert Dornhelms Historienfilm Kronprinz Rudolf als Mary von Vetsera und in Le Ragazze di San Frediano als Mafalda vor der Filmkamera.	Puccini 's role as Mafalda in the 2007 Rai Uno miniseries Le ragazze di San Frediano cast her among many other well-known Italian actresses, including Martina Stella, Chiara Conti, and Camilla Filippi.
< 0.1	Unter anderem ist es für die Durchführung der Volkszählung zuständig.	Every 10 years, this organisation conducts a national census.
< 0.1	Daraufhin nahm sich Nikolaus, der es mit der ehelichen Treue schon mehrfach nicht so genau genommen hatte, eine Mätresse, Alexandras Hofdame Barbara Nelidowa.	Nelidova went with them, and though Alexandra was jealous in the beginning, she soon came to accept the affair, and remained on good terms with her husband's mistress.

Continuation of Table 8

1	**2**	**3**
0.12	Im Unterschied zu Cognac wird Armagnac in einem kontinuierlichen Brennverfahren nur einmal destilliert, also nicht rektifiziert.	Armagnac is traditionally distilled once, which results initially in a less polished spirit than Cognac, where double distillation usually takes place.
0.29	Die 64,5 Prozent , welche die SPD unter seiner Führung erzielte, waren das höchste Ergebnis, welches je eine Partei auf Bundeslandsebene bei einer freien Wahl in Deutschland erzielt hatte.	In the election that was conducted in the western part of Berlin two months later, his popularity gave the SPD the highest win with 64.5 % ever achieved by any party in a free election in Germany.
Similarity	**Spanish source sentence**	**English reference translation**
Misaligned	En febrero de 1988, a 12 UA del Sol, el brillo de Quirón alcanzó el 75 % Este comportamiento es típico de los cometas pero no de los asteroides.	In February 1988, 12 AU from the Sun, Chiron brightness reached 75%.
0.1	Sin embargo, el museo, llamado no fue terminado sino hasta el 10 de abril de 1981, dos días antes del vigésimo aniversario del vuelo de Yuri Gagarin.	However, it took until April 10, 1981 (two days before the 20th anniversary of Yuri Gagarin's flight) to complete the preparatory work and open the Memorial Museum of Cosmonautics.

6 Conclusions

Wikipedia is a rich resource for parallel sentence mining in SMT. Comparing different translation models containing MT useful data ranging from strongly comparable, to parallel, we concluded that there is sufficient empirical evidence not to dismiss sentence pairs that are not fully parallel on the suspicion that because of the inherent noise they might be detrimental to the translation quality. On the contrary, our experiments demonstrated that in-domain comparable data are strongly preferable to out-of-domain parallel data. However, there is an optimum level of similarity between the comparable sentences, which according to our similarity metrics (for the language pairs we worked with) is around 0.4 or 0.5.

Additionally, the two step procedure we presented, demonstrated that an initial in-domain translation dictionary is not necessary, it can be constructed subsequently, starting with a dictionary extracted from whatever out-of-domain data.

We want to mention that it is not the case that our extracted Wikipedia data is the maximally MT useful data. First of all, LEXACC may be improved in many ways, which is a matter for future developments. For instance, although the cross-linking feature is highly relevant for language pairs with similar word ordering, it is not very effective for language pairs showing long distance re-ordering. We also noticed that a candidate pair for which its two parts contained different numerical entities (numbers, dates, times) was dropped from further consideration. Thirdly, the extraction parameters of LEXACC were not re-estimated for the Wiki-Train construction. Additionally, we have to mention that LEXACC evaluated and extracted only full sentences: a finer-grained (sub-sentential) extractor would likely generate more MT useful data. Also, one should note that the evaluation figures are just indicative for the potential of Wikipedia as a source for SMT training. In previous work it was shown that using factored models for inflectional target languages (Boroş et al, 2013) and cascading translators (Tufiş and Dumitrescu, 2012) may significantly improve (several

BLEU points) the translation accuracy of an SMT system. Some other techniques may be used to improve at least translations into English. For instance, given that English adjectives and all functional words are not inflected, a very effective way, for a source inflectional language would be to lemmatize all words in these categories. Another idea is to split compound words of a source language (such as German) into their constituents. Both such simplifications are, computationally, not very expensive (and for many languages appropriate tools are publicly available) but may significantly reduce the number of out-of-vocabulary input tokens.

The parallel Wiki corpora (before and after the boosting step), including the test sets (containing 10,000) and the dev–sets (containing 1,000 sentences) are freely available on-line[13].

Acknowledgments. This work has been supported by the EU under the Grant Agreements no. 248347 (ACCURAT) and no. 270893 (METANET4U).

References

[1] Sisay Fissaha Adafre and Maarten de Rijke. 2006. *Finding similar sentences across multiple languages in Wikipedia.* In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006), April 3-7, 2006. Trento, Italy, pp. 62–69.

[2] Tiberiu Boroş, Stefan Daniel Dumitrescu, Radu Ion, Dan Ştefănescu, Dan Tufiş. 2013. *Romanian-English Statistical Translation at RACAI.* În E. Mitocariu, M. A. Moruz, D. Cristea, D. Tufiş, M. Clim (eds.) Proceedings of the 9th International Conference "Linguistic Resources and Tools for Processing the Romanian Language", 16-17 mai, 2013, Miclăuşeni, Romania, 2013. "Alexandru Ioan Cuza" University Publishing House. 197 p. ISSN 1843-911X. pp. 81–98, 2013.

[13] http://ws.racai.ro:9191/repository/search/

[3] Ştefan Dumitrescu, Radu Ion, Dan Ştefănescu, Tiberiu Boroş, Dan Tufiş. 2013. *Experiments on Language and Translation Models Adaptation for Statistical Machine Translation.* In Dan Tufiş, Vasile Rus, Corina Forăscu (eds.) *TowardsMultilingual Europe 2020: A Romanian Perspective,* pp. 205–224, 2013.

[4] Ştefan Dumitrescu, Radu Ion, Dan Ştefănescu, Tiberiu Boroş, Dan Tufiş. *Romanian to English Automatic MT Experiments at IWSLT12.* In Proceedings of the International Workshop on Spoken Language Translation, December 6 and 7, 2012, Hong Kong, pp. 136–143.

[5] Qin Gao and Stephan Vogel. 2008. *Parallel implementations of a word alignment tool.* In Proceedings of ACL-08 HLT: Software Engineering, Testing, and Quality Assurance for Natural Language Processing, June 20, 2008. The Ohio State University, Columbus, Ohio, USA, pp. 49–57.

[6] Philipp Koehn. 2005. *Europarl: A Parallel Corpus for Statistical Machine Translation,* In Proceedings of the tenth Machine Translation Summit, Phuket, Thailand, pp. 79–86.

[7] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondej Bojar, Alexandra Constantin, and Evan Herbst. 2007. *Moses: open source toolkit for statistical machine translation.* In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions (ACL '07). Association for Computational Linguistics, Stroudsburg, PA, USA, pp.177–180.

[8] Mehdi Mohammadi and Nasser GhasemAghaee. 2010. *Building bilingual parallel corpora based on Wikipedia.* In Computer Engineering and Applications (ICCEA 2010), Second International Conference on Computer Engineering and Applications, Vol. 2, pp. 264–268. IEEE Computer Society Washington, DC, USA.

[9] Dragoş Munteanu, Daniel Marcu. 2005. *Improving machine translation performance by exploiting comparable corpora.* Computational Linguistics, 31(4), pp. 477–504.

[10] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. 2002. *BLEU: A method for automatic evaluation of machine translation.* In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), July 2002. Philadelphia, USA, pp. 311–318.

[11] Inguna Skadiņa, Ahmet Aker, Nikos Glaros, Fangzhong Su, Dan Tufiş, Mateja Verlic, Andrejs Vasiljevs, Bogdan Babych, 2012. *Collecting and Using Comparable Corpora for Statistical Machine Translation.* In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), May 23-26, 2012. Istanbul, Turkey, ISBN 978-2-9517408-7-7.

[12] Jason R. Smith, Chris Quirk, Kristina Toutanova. 2010. *Extracting parallel sentences from comparable corpora using document level alignment.* In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 403–411. © Association for Computational Linguistics (2010).

[13] Ralf Steinberger, Bruno Pouliquen, Anna Widiger, Camelia Ignat, Tomaž Erjavec, Dan Tufiş. 2006. *The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages.* In Proceedings of the 5^{th} LREC Conference, Genoa, Italy, 22-28 May, 2006, ISBN 2-9517408-2-4, EAN 978-2-9517408-2-2.

[14] Dan Ştefănescu, Radu Ion, Sabine Hunsicker. 2012. *Hybrid parallel sentence mining from comparable corpora.* In Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT 2012), pp. 137–144, Trento, Italy, May 28-30, 2012.

[15] Jörg Tiedemann. 2012. *Parallel data, tools and interfaces in OPUS.* In Proceedings of the 8th International Conference on Lan-

guage Resources and Evaluation (LREC 2012), May 23-26, 2012. Istanbul, Turkey, ISBN 978-2-9517408-7-7.

[16] Dan Tufiş. *Finding Translation Examples for Under-Resourced Language Pairs or for Narrow Domains; the Case for Machine Translation.* 2012. In Computer Science Journal of Moldova, Academy of Sciences of Moldova, Institute of Mathematics and Computer Science, ISSN 1561-4042, vol.20, no.2(59), pp. 227–245.

[17] Dan Tufiş, Verginica Barbu Mititelu, Dan Ştefănescu, Radu Ion. 2013. *The Romanian Wordnet in a Nutshell.* Language and Evaluation, Springer, Vol. 47, no. 2, 2013, ISSN 1574-020X, DOI: 10.1007/s10579-013-9230-7

[18] Dan Tufiş, Radu Ion, Ştefan Dumitrescu, Dan Ştefănescu. 2013. *Wikipedia as an SMT Training Corpus.* In Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP 2013), Hissar, Bulgaria, September 7-13, 2013.

[19] Dan Ştefănescu, Radu Ion. 2013. *Parallel-Wiki: A Collection of Parallel Sentences Extracted from Wikipedia.* In Proceedings of 14th International Conference on Intelligent Text Processing and Computational Linguistics, pp. 137–144, Trento, Italy, March 24-30, 2013, Samos, Greece.

Dan Tufiş[1], Radu Ion[2], Ştefan Daniel Dumitrescu[3]

Institute for AI, Romanian Academy, Bucharest, Romania
[1] E-mail: tufis@racai.ro
[2] E-mail: radu@racai.ro
[3] E-mail: danstef@racai.ro

Part 2

REGULAR PAPERS

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Visual Analytics of Movement: a short introduction

Gennady Andrienko, Natalia Andrienko

Abstract

The paper gives a short introduction to visual analytics of movement.

Keywords: visual analytics, movement data, mobility.

1 Introduction

The mission of Visual Analytics [1] is to find ways to fundamentally improve the division of labour between humans and machines so that the computational power could amplify the human perceptual and cognitive capabilities. The term Visual Analytics stresses the key role of visual representations as the most effective means to convey information to humans mind and prompt human cognition and reasoning. Visual Analytics is defined as the science of analytical reasoning facilitated by interactive visual interfaces. It combines automated analysis techniques with interactive visualizations so that to support synergetic work of humans and computers.

In many areas of peoples life and activities it is important to understand movement behaviours and mobility patterns of people, animals, vehicles, or other objects. Thanks to the recent advent of inexpensive positioning technologies, data about movement of various mobile objects or agents are collected in rapidly growing amounts. There is a pressing need in adequate methods for analysing these data and extracting relevant information.

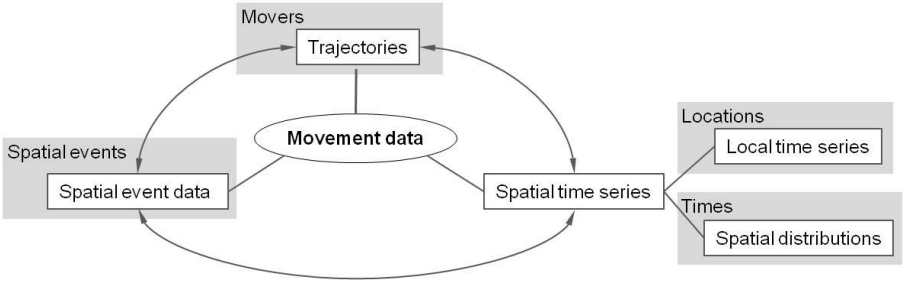


Figure 1. Movement data.

Movement data are inherently complex as they involve moving objects, events, geographic space and time (Figure 1). In addition to their own intrinsic complexities, these components are interdependent, which multiplies the overall complexity. As a result, movement data cannot be adequately modelled (at least at the present time) for a fully automatic analysis. At the same time, movement data, which are mostly acquired by automatic position tracking, are usually very poor semantically. The records basically consist of time stamps and coordinates. Semantic interpretations must emerge as a result of exploration and analysis where a human analyst plays the key role. Appropriate visual representations of movement data and outcomes from automated analysis procedures are paramount for this process.

The presentation gives an overview of how to analyse such data. For selected tasks, we propose scalable visual analytics methods [2,3]. The work of the methods is illustrated using several examples of real-world datasets significantly differing in their properties. For example, Figure 2 illustrates several methods for analysis of car traffic data. We analyse to what extent these and other existing methods cover the space of the types of movement data and the possible analysis tasks, identify the remaining gaps, and outline the directions for the future research.

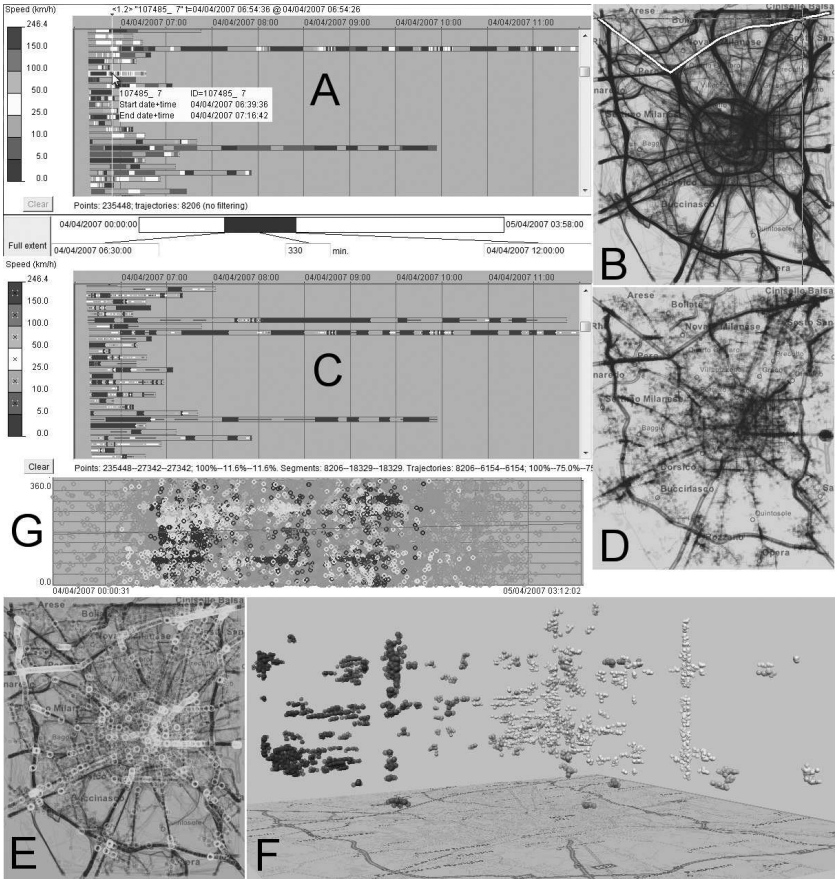


Figure 2. A: A time bars display shows the speeds by colour-coding. Mouse-pointing highlights the trajectory and marks the pointed position in a map (B). C: Trajectory segments are filtered according to the speed values. D: Only the segments satisfying the filter are visible on the map. E: Low speed events have been extracted from the trajectories according to the segment filter. F: Density-based spatio-temporal clusters of the low speed events are shown in a space-time cube. G: A scatterplot shows the times (horizontal dimension) and movement directions (vertical dimension) of the low speed events.

References

[1] J. Thomas, K. Cook. *Illuminating the Path. The Research and development Agenda for Visual Analytics.* IEEE Computer Society, 2005.

[2] N. Andrienko, G.Andrienko. *Visual Analytics of Movement: an Overview of Methods, Tools, and Procedures.* Information Visualization, vol. 12, no 1 (2013), pp. 3–24.

[3] G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Wrobel. *Visual Analytics of Movement.* Springer-verlag, 2013.

Gennady Andrienko[1], Natalia Andrienko[2]

[1]Fraunhofer Institute IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
Email: gennady.andrienko@iais.fraunhofer.de

[2]Fraunhofer Institute IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
Email: natalia.andrienko@iais.fraunhofer.de

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Similarities Identification in Logo Images

Silviu Ioan Bejinariu, Mihaela Costin,
Adrian Ciobanu, Svetlana Cojocaru

Abstract: Automatic trademarks identification is important in order to easily highlight possible resemblance or dissimilarity in design. The paper proposes a new method for fast identification of similarities in logo images. As a measure of similarity we used the number of matching keypoints. The keypoints are selected using the SIFT algorithm. It presents scale and rotation invariant features used to robustly describe digital images. The proposed method was validated using an image database prepared by authors.

Keywords: image database, invariant features, SIFT algorithm, trademarks, similarity identification.

1 Introduction

In order to ensure trademarks "uniqueness" and avoiding confusions, we propose a procedure for automatic detection of similar patterns in logos. The imposed conditions are: (a) two logos can't have identical forms and colors; (b) two logos can't have identical forms with different colors and to represent different structures (or societies); (c) a logo can't be part of another logo in order to represent different companies; (d) two logos may contain common elements only if these are characters, exterior edges of the emblems, or other geometrical forms, not representing a specific trademark of a company. The issue we study is not on logos recognition or identification; it is focusing instead on the common elements (patterns) identification. Numerous methods for logo detection or recognition were developed. Some are based on polygonal approximation of objects, physical modeling, Principal Component Analysis or use the spatial information. Several methods are based on keypoints detection: SIFT,

Harris, SUSAN [1], [2], [3], [4]. Other methods use statistical features as moments, texture and color matching (histogram intersection). Some hybrid methods based on the weighted combination of different image features have good results in logo and trademark recognition. The proposed method is based on SIFT keypoints matching. Even if the SIFT algorithm is not very fast, it is used in this paper only to find similarities between logos, not for a complete recognition.

2 Features extraction, SIFT algorithm

The SIFT algorithm (Scale Invariant Feature Transform) is a relatively new method employed to select distinctive features, used in pattern recognition, localization, 3D mapping, tracking, image mosaicing, etc. It allows scale and rotation invariant features detection, with good results for affine distortions. SIFT features were proposed by Lowe in [3], [4], [5] and [6]. The algorithm has 4 distinctive stages: extrema detection in the scale space of the image, keypoints selection and localization, keypoints orientation assignment and description generation. The identified features have to be distinctive.

2.1. Scale-space extrema detection

Keypoint candidates selection is performed by finding the extrema of the Difference of Gaussians (DOG) function computed as the difference of two scaled images separated by a multiplicative factor k.

$$\begin{aligned} D(x,y,\sigma) &= L(x,y,k\sigma) - L(x,y,\sigma) = \\ &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y), \end{aligned} \tag{1}$$

where $L(x,y,\sigma)$ is the scale space of the image $I(x,y)$ obtained by convolving it with the Gaussian kernel $G(x,y,\sigma)$. Extrema points depend on the frequency sampling in the scaled space and the initial value of σ.

2.2. Keypoints localization

Keypoints are selected from the most stable and accurately localized candidates. Keypoint candidates which have low contrast or strong edge response in one direction only are removed. Because the candidates obtained in higher scales correspond to several pixels in the original

image, an exact localization is required. This is performed by computing the extrema points of the Taylor expansion up to quadratic terms of the scale space function $D(x, y, \sigma)$ [4].

2.3. Orientation assignment

To make the keypoint description invariant to rotation, its orientation is computed using the orientation histogram of local gradients of the closest smoothed image $L(x, y, \sigma)$. The gradient magnitude and orientation are computed using pixel differences:

$$m(x, y) = \\ = \sqrt{(L(x-1, y) - L(x+1, y))^2 + (L(x, y-1) - L(x, y+1))^2} \quad (2)$$

$$\theta(x, y) = arctg \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (3)$$

Each point is added to the histogram weighted by the gradient magnitude $m(x, y)$ and by a circular Gaussian. To obtain a more accurate orientation, the dominant peaks in the histogram are interpolated with their neighbors.

2.4. Keypoint descriptor computing

The keypoint descriptor contains $128 = 4 \times 4 \times 8$ values obtained using 16 orientation histograms computed in a 4×4 grid. Each histogram contains 8 orientation bins. The descriptor is computed in a support window of 16×16 pixels around the keypoint [5].

3 Image matching procedure

Our procedure imposes: (1) building and structuring a feature database (stages 1.a., 1.b.) and (2) identifying images in the database that include similarities to a specified image.

1.a. A collection of images (models) is created and preprocessed. To reduce computing time, images converted in grey levels, are scaled to a standard dimension, 256 pixels width, with the aspect ratio preserved. It

should be noted that these changes do not affect the processing results. Filters might be applied for noisy, blurred images.

1.b. A feature database is created for the selected images. SIFT algorithm is applied to extract features (keypoints). In the database we store information related to position, orientation, dimension of the significant neighborhood, filter response (the power), the octave from which the keypoint was selected, and the 128 SIFT coefficients. The identification decision will afterward rely only on these coefficients.

(2) During the identification stage, the same scaling operations will be applied on the source image, to be processed therefore: 256 grey levels conversion, SIFT features extraction, computing the similarity coefficient with each model image from the data base. Establishing the similarity score is realized by detecting keypoints that match. We compute the Euclidian distance between SIFT descriptors of each keypoint from source images and those of the model image. By sorting the computed values for source images keypoints, a match is established when the minimum computed distance is less than a certain percentage (30%) from the second distance.

Models list is sorted upon the matching coefficient with the source image. We obtain a list of images which contain similarity elements with the source image.

a. b. *(green)* c. d. e. f.

g. h. i. j. k. l. *(red)*

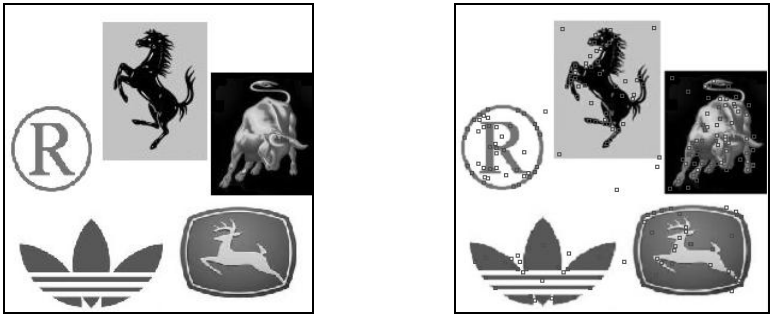
m. n. o. p. q. r.

Figure 1. Logo images database (partial)

4 Experiments

Our similarity identification procedure was validated using for tests a model images database with 100 different items including only a few logos for the same company (thus presenting similarity elements). Figure 1 illustrates 18 images from the model database. In order to validate this procedure a test image (source) was created including component elements of more logos (b, d, I, o, r) from the initial database. Further denoted as "test", this image is presented in Figure 2.

Table 1 presents the results of applying our procedure for similarity detection, on some images from Figure 2. It highlights the first 3 matches from the point of view of matching keypoints. As each image is also included in the test set, always the most important match situation happens when compared to itself. However, the second match coefficient is obtained for the "test" image excepting when: (p) the image doesn't contain common elements with the test image, for (q) and (r), images that are relatively similar and the "test" image matching is the third in line. Table 2 presents the results of comparing the "test" image with all the images in the database, on the decreasing matching number. All images from list are those containing common elements with the test image. Neither of the results was containing false positive matching.



a. Original image b. Detected keypoints (322)

Figure 2. Test logo image

Conclusion

Our method allows similarity identification in images, being suitable in establishing logos for new conceived companies. This procedure was implemented as a Windows application in C++ language. We used the OpenCV free library [7] to detect SIFT Transform keypoints and to work with images. As we obtained promising results, we will develop this direction, applying new similarity criteria for a qualitative optimization of the results.

Table 1. Keypoints match results, original images against all images

Source image	Key points	1st match		2nd match		3rd match	
		Image	Matches	Image	Matches	Image	Matches
b	84	b	84	test	22	l	14
d	144	d	144	test	6	other	4
i	127	i	127	test	23	other	16
l	63	l	63	test	10	b	8
n	218	n	218	test	23	o	22
o	115	o	115	test	46	n	20
p	484	p	484	other	55	other	4
q	218	q	218	r	215	test	50
r	215	r	215	q	110	test	72

Table 2. Keypoints match results, test image against all images

Source image	Key points	Matched images					
		Image	Matches	Image	Matches	Image	Matches
test	322	test	322	r	72	q	50
		o	47	i	24	b	21
		n	21	d	10	l	9

References

[1] Oliver Sidla, Michal Kottmann, Wanda Benesova (2011), Real-time pose invariant logo and pattern detection, *Proc. SPIE 7878, 78780C.*

[2] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, (2005) "A comparison of affine region detectors , International Journal of Computer Vision, 65(1/2):43–72.

[3] Lowe, D. (1999). Object Recognition from Local Scale-Invariant Features. *Proceedings of the 7th IEEE International Conference on Computer Vision*, *2*, pg. 1150-1157.

[4] Matthew Brown and David G. Lowe (2002). Invariant features from interest point groups. In British Machine Vision Conference, BMVC 2002, Cardiff, Wales (September 2002), pp. 656-665.

[5] Lowe, D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60*(2), 91-110.

[6] Lowe, D. (2004). Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image, US Patent 6,711,293 (March 23, 2004), Asignee: The University of British Columbia.

[7] Bradski, G., & Kaehler, A. (2008). Learning OpenCV. Computer Vision with the OpenCV Library. O'Reilly Media, Inc.

Silviu Bejinariu[1], Mihaela Costin[2], Adrian Ciobanu[3], Svetlana Cojocaru[4],

[1]Institute of Computer Science, Romanian Academy, Iasi Branch
E-mail: silviu.bejinariu@iit.academiaromana-is.ro, silviu.bejinariu@gmail.com

[2]Institute of Computer Science, Romanian Academy, Iasi Branch
E-mail: mihaela.costin@iit.academiaromana-is.ro, mccostin@gmail.com

[3]Institute of Computer Science, Romanian Academy, Iasi Branch
E-mail: adrian.ciobanu@iit.academiaromana-is.ro, adrian.ciobanu@yahoo.com

[4]Institute of Mathematics and Computer Science Academy of Sciences of Moldova
E-mail: Svetlana.Cojocaru@math.md

Utilization of the Computational Resources Provided by HP-SEE Project

Peter Bogatencov, Boris Rybakin,
Grigore Secrieru, Nicolai Iluha

Abstract

Scientific Computing Infrastructure, related technologies and services have begun developing for R&D communities of Moldova due to the support of a series of international and national projects. HP-SEE project linked existing HPC facilities in South East Europe in a common infrastructure and allowed to provide access to modern HPC resources to wide range of regional research teams. In Moldova, the project activated new research communities and promoted joint research activities at national, regional and European levels.

Keywords: European eInfrastructures, Scientific Computing Infrastructure, HPC resources.

1 Introduction

The international cooperation in the field of Scientific Computing (HPC, Grids, Cloud computing) represents an important factor for developing the area of scientific research and perspectives of the European future for research community of Moldova. In the last years, Moldova as a part of South-East Europe (SEE) actively participated in a number of targeted initiatives funded by the European Commission, focused on the creation of new user communities, and enabling collaborative research across various fields in South-East Europe. Although the necessary initial contributions in the region were done, the computational facilities available now in SEE region are in general less developed than in Western Europe.

2 Organization of the access to the regional HPC resources

In the field of High Performance Computing, the European Commission supports a series of initiatives to provide access to HPC facilities to leading European researchers. To cover the permanently rising needs of researchers in SEE region the regional eInfrastructure development project "High-Performance Computing Infrastructure for South East Europe's Research Communities" (HP-SEE) was launched in 2010. The aim of the South-East Europe HPC initiative (http://www.hp-see.eu/) is equal participation of all countries of the region in European eInfrastructure development trends.

In the project there are two categories of participants, which form the project consortium – resource providers and partners-beneficiaries. Beneficiary countries like Moldova receive preferences from gaining access to resources available in the region. RENAM Association (National Research and Educational Network of Moldova) and the Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova (IMI ASM) are involved in the project from Moldova.

HP-SEE links existing and upcoming HPC facilities in the region in a common infrastructure. The available regional HPC infrastructure integrates the most powerful HPC clusters and supercomputers provided by the main infrastructure partners from the six countries, participating in the project: Greece, Bulgaria, Romania, Hungary, Serbia and FYROM. In Moldova two HPC clusters are operating – one installed in IMI ASM and the second in the State University of Moldova. They are in mid-size computational installations and used at present mainly for applications development, testing and debugging. These clusters are used also for training purposes and provide resources to regional SEE HPC training infrastructure.

European HP-SEE project offered ability of using regional HPC resources by national research teams. The signed by HP-SEE partners Memorandum allowed to provide access to modern HPC resources, enhanced use of computational infrastructures by national community of

Moldova and intensified complex application development.

3 Utilization of the HP-SEE project computational resources

All applications that are being developed in the project, are grouped in Virtual Research Communities (VRC). Among Computational Physics VRC applications that use HPC infrastructure there are two applications from Moldova: in IMI ASM – application AMR_PAR, and in the Institute of Applied Physics of ASM – application TMDC (crystallography models – electronic structure, optical and transport properties of layered transition-metal dichalcogenides and their intercalated compounds). AMR_PAR is considering a continuum mechanics problem, and namely the problem of modeling the explosion of a supernova type II and, for this example, there was created an algorithm and parallel program using the AMR method. Calculations method based on the use of AMR hierarchical grid cells can significantly improve quality of calculations in various fields of science and engineering. In addition, the most of applications that use AMR, is well parallelized. We used the programming language Fortran 90. For running AMR_PAR we used the following cluster configuration: operating environment – Linux, 36 nodes, each node has 2 Intel XeonX5560, 2,8 Ghz, 24 Gb RAM, in total 576 computing cores available. Benchmarking results are collected for the following dimensions: 128x128x128, 256x256x256. For arrays of dimension 128x128x128 with five levels of nesting the optimal number of cores for calculation is about four. Otherwise, the increase of cores number does not reduce wall-clock time of calculations, but the processing time increases dramatically (see Figure 1).

Estimation of necessary amount of resources for performing calculations for arrays of dimension up to 2048x2048x2048 cells is shown in Table 1. We see that for high dimension arrays calculations can be performed only on a supercomputer. For small grids (up to 384x384x384 cells) we continued using resources of HPC cluster. In addition to the mentioned above, two new applications are being prepared to run on HP-SEE resources – for computer added design of semiconductor de-

vices and decision-making processes modeling for economical systems. These applications are being developed by research teams from the State University of Moldova with support of specialists from IMI ASM and RENAM Association.

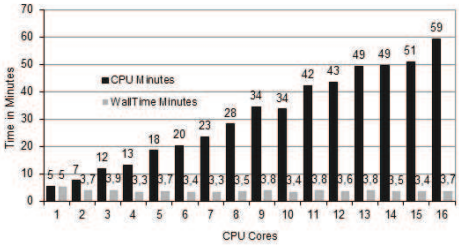


Figure 1. Acceleration dependences from number of CPU cores (HPC cluster, Linux)

Table 1. Benchmark demands for resources results and forecasting

Dimension	Layers	Cores	RAM Gb	CPU time	Wall Time
128x128x128	5	4	0,789	28 min	3,5 min
256256256	5	8	6,062	527 min	66 min
384x384x384	5	8	19,2	2110 min	270 min
448x448x448	5	8 - 16	37,7	$\sim$ 4500 min	$\sim$ 500 min
1024x1024x1024	5	16 - 32	$\sim$ 415	$\sim$ 2000 hours	$\sim$ 248 hours
2048x2048x2048	5	32 - 64	$\sim$ 3250	$\sim$ 1200 days	$\sim$ 154 days

Acknowledgments. This work is supported by EC HP-SEE project under contract number 261499 and by STCU project #5807 "Instrumental support for complex applications porting to the regional HPC infrastructure".

Peter Bogatencov[1], Boris Rybakin[2], Grigore Secrieru[3], Nicolai Iluha[1]

[1]RENAM Association, 5 Academiei str., Chişinău, Moldova
Email: bogatencov@renam.md, nick@renam.md

[2]MSU, Lenin Mountains, Building 1; SRISA RAS, Moscow, Russia
Email: rybakin@vip.niisi.ru

[3]The Institute of Mathematics and Computer Science of ASM,
Email: secrieru@cc.acad.md

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Language Technology and Resources for Cultural and Historic Heritage Digitalization

E.Boian, C.Ciubotaru, S.Cojocaru,
A.Colesnicov, L.Malahov, M.Petic

Abstract

This article describes digitization of old Romanian texts, problems at their recognition, and motivates the necessity to create specific electronic resources mirroring the history of the standard Romanian language. We propose a technology for creation of linguistic lexicon for Moldavian Cyrillic script of 1967–1989, starting from modern Romanian lexicon (standard Romanian). This technology is based on transliteration and alignment of parallel texts.

Keywords: digitization, Romanian linguistic resources, text recognition, language technology, Cyrillic script, transliteration, text aligning.

1 Introduction

The main directions of the cultural policy into zones when the Romanian language is spoken refer to study, evaluation and digitization of cultural and historic heritage. Process of heritage digitization needs many problems to be solved that are related to recognition, editing, translation, interpretation, circulation and reception of texts printed in Romanian and other modern languages. These problems became more complicated for Romanian as we need to consider the historic period when the source was printed, and we have several periods.

In its history, the Romanian language has passed through a long and rich evolution [1, 2]. The first book printed in the Romanian territory was the Church-Slavonic *Liturgy Book* (1508) edited by Serbian hieromonk Makarie. The first printed book in Romanian appeared in Brashov in 1535 [3]. It was *The Romanian Catechism* published by deacon Coresi.

The National Library of the Republic of Moldova possesses approx. 21,000 old and rare books. The collection contains approx.20 books printed in Romanian in the Romanian Cyrillic and transitional scripts in Bessarabia (Chisinau and Dubasari) [4].

Public libraries of Sankt-Petersburg keep important quantities of old Romanian books (the 16^{th}–19^{th} centuries). For example, there are 66 titles in *The Catalog of Cyrillic editions of Southern Slavonians and Romanians*. 45 volumes are of the Southern Slavonian origin, while 21 can be attributed to Romanian lands [5].

The existent studies explain aspects of development of main language components: alphabet, lexicon, and orthography, appearance of each vowel and consonant at each specific stage of the language evolution. This information is useful to construct linguistic resources and to use specific tools for a specific period of the language history.

Taking into account a specific period, we will propose a technology to obtain these components. In particular, we study the problem of digitization of books printed in Moldavian Soviet Socialist Republic (MSSR) in 1967–1989.

Our work is a long-term project that is in its beginning now. We implement it using the principle "from now in the depths of time".

2 State-of-the-arts in working with historical texts

The problem of digitization and preservation of historical linguistic heritage is a domain of priority in digital agenda for Europe.

The difficulties in digitization and preservation of this heritage lie in correct recognition of characters and in lack of adequate lexicons

corresponding to the periods of the texts printing. One of solutions of the lexicon problem could be aligning of old texts to contemporary linguistic norms [6].

Manuscript digitization and recognition is complicated because it requires additional operations, such as adjusting the contrast, cleaning the image, text segmentation. We also need to develop special algorithms of recognition and specialized lexicons. Further we limit ourselves with Romanian texts printed with Latin letters.

Process of digitization and recognition consists of the following stages:

- Digitization of the text resulting in its graphical electronic copy.
- Recognition by standard techniques, namely, using OCR (Optical Character Recognition) software, possibly, with its training. Without OCR, procedures of conversion using Artificial Intelligence techniques should be applied. Transliteration of the text is performed taking into account specific letters from the initial text.
- Verification of the recognized text is performed using lexical resources specialized for the corresponding period.

As to OCR of printed and handwritten Cyrillic characters, we can mention a paper [7] where both standard ABBYY FineReader and AI techniques are used, in particular, artificial neural networks. There exists an application of methods based on knowledge technologies to the digital archive and multimedia library for Bulgarian traditional culture and folklore [8]. Problems of transliteration caused by parallel use of two alphabets, Cyrillic and Latin, which appear at processing of written texts in modern Serbian, were solved applying monolingual and multilingual corpora and various e-dictionaries [9].

Some aspects of recognition of Romanian texts printed in the Cyrillic script are exposed in [10].

For Romanian historical linguistic heritage, solution for this problem confronts with specific difficulties: big number of periods of the

language evolution; relatively small number and big dispersion of deposited resources; big variety of used alphabets, in particular, several so-called "transitional" (mixed Latin-Cyrillic) alphabets.

There are two variants of the Cyrillic alphabet that were used to write Romanian texts. The first one was used for Romanian writing since 14^{th}–15^{th} centuries until 1862. The second is, in fact, an adaptation of the Russian Cyrillic alphabet to reproduce the Romanian phonetics by Russian orthographical norms that led to some weird orthographical effects.

This second variant based on the Russian alphabet was used in the Moldavian Autonomous Soviet Socialist Republic (MASSR) in 1930–1932 and 1938–1940, then in the MSSR since its formation in 1940, and until 1989. This alphabet is used till now in Transnistria. In 1932–1938 the Latin-based alphabet was used in the MASSR.

3 Processing of Romanian texts in the Cyrillic script in MASSR and MSSR

The initial period of Cyrillic writing (1924–1951) is associated with an extremely specific lexicon. It is characterized by:

- use of Russian words, for example, **совет**, **указ**, **словарь** (council, decree, dictionary) instead of their Romanian equivalents (**consiliu**, **decret**, **dicţionar**);
- deletion of Romanian neologisms that were claimed as "bourgeois";
- fixing of local (Transnistrian) lexicon;
- introduction of self-invented neologisms for abstract notions that cannot be found in the language of Bessarabian countrymen, for example, **амувремник** (**amuvremnic** = **contemporary**), instead of **contemporan**; the word was constructed from dialectal forms of words **now** and **time** with an adjectival suffix;

- fixing of peculiarities of local (Transnistrian) accent, like **ди** (**di)** instead of **de** (preposition), **мержи** (**merji**) instead of **merge** (go), **сунити** (**suniti**) instead of **sunete** (sounds), **кы** (**cî**) instead of **că** (how), etc.

We will refer below to 1967-1989 when the Cyrillic alphabet was used in the Moldavian SSR. To perform OCR of texts of this period, it is necessary to train the OCR system to recognize an additional letter **ӂ**, and to provide the corresponding lexicon. The lexicon characteristic for this period should be used to automate verification and validation of the recognized words at the OCR process. This lexicon can be created: a) manually; b) by transliteration Romanian words written in the Latin script to the Cyrillic script; c) by alignment of texts printed in parallel in the Cyrillic and Latin scripts.

In this context, transliteration is transcription of a Romanian word in its equivalent form written in the Cyrillic script conforming linguistic norms accepted in the MSSR in 1967–1989. The method of transliteration would be ideal for this case if we could formalize all rules of transcription. A preliminary study show that this process is vague and cannot be fully automated because of irregularities and discordance in phonology, morphology, and syntax between the linguistic norms of the Romanian language and those accepted in the MSSR. The process could be automated partially by formalizing rules of transliteration, manual intervention, and text alignment.

Obvious difficulties appear in transliterating words of foreign origin, and foreign proper names, if the original language uses a Latin-based script. In most cases, such words keep their writing in the Romanian language, whilst their Cyrillic transliteration conforms their pronunciation.

For example, ***design*** – ***дизайн, cowboy*** – ***ковбой, Watt*** – ***Ват, charleston*** – ***чарлстон***. These words can be transliterated only manually.

The transliteration can be partially automated for the original Romanian lexicon. The rules for transcription of letters and their pairs were developed. Several from these rules follow.

1. Rules for letter to letter transcription. For example, ***a*** **→*а*,** ***ă*** **→ *э*, *b* →*б*, *d*→ *д*, *f*→ *ф*, *l*→*л*, *m*→ *м*, *n*→ *н*, *r*→ *р*, *ş*→ *ш*, *t*→ *т*, *ţ* → *ц*, *v*→ *в*, *z*→*з* (*bardă* – *бардэ [axe]*, *zarvă* – *зарвэ [embroilment]*, *măr*→*мэр [apple]*).**

2. Transcription rules for letters ***î*** and ***â***. Combinations ***âi*** and ***îi*** are transcribed as ***ы*** in words ***mâine, pâine, câine*** and their derivatives (***mâine*→*мыне [arm]*, *pâine*→*пыне [bread]*, *câine*→*кыне [dog]*, *mîine*→*мыне*, *pîine*→*пыне*, *cîine*→*кыне***). Letter-to-letter rules ***â* →*ы*, *î* →*ы*** are applied in other cases (***român*→*ромын [Romanian]*, *întâi*→*ынтый [the first]***).

3. Transcription rules for letter ***c***.

   (a) ***c*→*к***, if ***c*** is followed by vovels ***a, â, î, o, u***, or a consonant different of ***h*** (***încreţit*→*ынкрецит [frizzled]*, *clocot*→*клокот [bubbling]*, *casă*→*касэ [home]*, *cucoş*→*кукош [cock]*, *câmp*→*кымп [field]***).

   (b) Combinations ***che***, ***chi*** are transcribed as ***ке*** and, correspondingly, ***ки*** (***cheltuială*→*келтуялэ [expense]*, *chihlimbar*→*кихлимбар[amber]*, *chibzui*→*кибзуи[consider]***).

   (c) If ***ce*** is not followed by ***a***, then the rule ***ce*→*че*** is apllied (***cercel* →*черчел [earring]*, *cep*→*чеп [pin]***).

   (d) ***cea* → *ча*** (***ceară*→*чарэ [request]*, *ceas*→*час [clock]*, *ceaţă*→*чацэ [fog]*, *ceaşcă*→*чашкэ[cup]***). Exception: demonstrative article ***cea (acea) →*чя*[≈*that*] (*ачя [≈this]*).**

   (e) If ***ci*** is not followed by vovels ***a***, ***o, u***, the the rule ***ci*→*чи*** is applied (***ciment*→ *чимент [cement]*, *ciclu*→*чиклу [cycle]*, *cimbrişor*→ *чимбришор [thyme]***). If a word is terminated by ***ci***, then one of the rules is applied: as an exception, ***ci*→*ч*** (***arici*→*арич [hedgehog]*, *beci*→*беч*, *prichici*→*прикич***); ***ci*→*чь*** for plural (***arici*→*аричь [hedgehogs]***, saci→***сачь [bags]***, maci→***мачь [poppies]***); ***ci*→*чи***, other cases (***răci*→*рэчи [cool]*, *înveşnici*→*ынвешничи [etrnalize]***).

(f) ***cio→чо*** (***ciorbă→чорбэ [soup]***, ***ciocârlie→чокырлие [lark]***, ***cioban→чобан [shepherd]***, ***cocioabă →кочоабэ [shanty]***).

(g) ***ciu→чу*** (***ciuperci→чуперчь [mushrooms]***, ***ciubotă→чуботэ [boots]***, ***bucium→бучум [horn]***).

4. Transcription rules for letter ***g***.

(a) If ***g*** is followed by a vowel ***a,â,î,o,u***, or by a consonant different from ***h***, or ***g*** is the last letter in the word, then ***g→ă*** (***grevă→гревэ [strike]***, ***gură→ гурэ [mouth]***, ***galben →галбен [yellow]***, ***gârlă→гырлэ [backwater]***, ***gară→ гарэ [station]***, ***meleag →меляг [land]***).

(b) ***ghea→гя*** (***gheaţă→ гяцэ [ice]***, ***dugheană→дугянэ [booth]***, ***tejghea→тежгя [counter]***).

(c) ***ghe→ге*** (***dulgher→дулгер [carpenter]***, ***gheţar→гецар [glacier]***, ***evanghelie→евангелие [gospel]***).

(d) ***ghi→ги*** (***ghiocel→гиочел [snowdrop]***, ***frânghie→ фрынгие [rope]***, ***ghiozdan→гиоздан [satchel]***).

(e) ***gea→йся*** (***geaba→йсяба [vain]***, ***geamăt→йсямэт [groan]***, ***săgeată→сэйсятэ[arrow]***).

(f) ***gia→йсиа*** (***norvegian → норвейсиан [Norwegian]***, ***refugiat→рефуйсиат [refugee]***).

(g) ***ge→йсе*** (***ger→йсер [frost]***, ***gen→йсен [genre]***, ***scurgere →скурйсере [drain]***).

(h) ***gi→йси*** (***gingaş→йсингаш [sweet]***, ***argintar→арйсинтар [silversmith]***, ***gimnastică→йсимнастикэ [gymnastics]***, ***milogi→милойси [beggars]***). Exceptions: if a word is terminated with ***gi*** and if is plural, or a verb in the 2^{nd} person, singular, then ***gi→йсь*** (***fulgi→фулйсь [flakes]***, ***fugi→фуйсь [run]***, ***mergi→мерйсь [go]***).

Using these rules, the transliteration process turns into passage through "sieves and screens". Starting from the contemporary lexicon of Romanian language [11,12], it establishes a set of filters, each filter with a priority coefficient that depends on the probability of obtaining a correct result in the application of that filter. Thorough selection of filters would exclude or minimize manual intervention. Filtered words are excluded from the lexicon, and the remaining words pass other filters. Unfortunately, certain manual intervention is necessary after all.

4 Conclusion

Our technology is oriented to successful solution, for each period of the language development, of two main problems: 1) developing of algorithms to recognize alphabets of a specific period; and 2) development of tools and interfaces needed to create the corresponding linguistic resources (lexicons). This would permit to recognize words and to align texts conforming contemporary linguistic norms.

As we move from one period to another, we can use previously elaborated tools and resources, thus implementing the principle "from now in the depths of time".

The proposed technology can be used in the formation and completion of specific linguistic resources with new words extracted from digitized materials and certified by language experts. It would allow construction of parallel corpora of different nature. Development of the proposed technology would provide opportunities to transliterate digitized text into modern Romanian, to customize graphics, to offer possibilities for corpora building, to preserve the original texts.

Specific electronic resources can be placed in Internet for public access contributing into development of the informational communicative media for the Romanian language. Moreover, these resources constitute an essential support for researchers, and conversions into modern standard text can be used as didactic materials at teaching.

References

[1] G. Ivănescu, *History of the Romanian language*, Iaşi, 1980. – In Romanian.

[2] Ştefan Munteanu, Vasile Ţâra. *History of the standard Romanian language*, Editura Didactică şi Pedagogică, Bucharest, 1978. – In Romanian.

[3] P. P. Panaitescu. *The beginning and the victory of the Romanian writing,* Bucureşti, 1965.

[4] *Books of Moldova (XVII – beg. of XX cent.).* Printing in Cyrillic script (XVII – beg. of XX cent.). General catalogue. Chişinău, 1992.

[5] *Bibliophils' values–2008.* Librarians' newspaper, June-July 2008, Nr. 6-7, p.1, http://87.248.191.115/bnrm/publicatii/files/3/93.pdf.

[6] M. Moruz, A. Iftene, A. Moruz, D. Cristea. *Semi-automatic alignment of old Romanian words using lexicons.* In: Proceedings of the 8-th International Conference "Linguistic resources and tools for processing of the Romanian language", Iaşi, Editura Universităţii, A.I. Cuza", 2012, pp. 119–125.

[7] Kornienko S.I. et al. *Program tools for recognition of handwritten and old-printed texts.* In: Informational Resources of Russia, 2011, nr. 1, pp. 35–37. – In Russian.

[8] Radoslav Pavlov, Galina Bogdanova, Desislava Paneva-Marinova, Todor Todorov, Konstantin Rangochev. *Digital archive and multimedia library for bulgarian traditional culture and folklore.* International Journal "Information Theories and Applications", Vol. 18, Number 3, 2011, pp. 276–288.

[9] Duško Vitas, Cvetana Krstev, Ivan Obradović, Ljubomir Popović, Gordana Pavlović-Lažetić. *An Overview of Resources*

and Basic Tools for the Processing of Serbian Written Texts, (http://poincare.matf.bg.ac.rs/∼cvetana/biblio/Solun03MATF.pdf).

[10] E.Boian, S.Cojocaru, C.Ciubotaru, A.Colesnicov, L.Malahov, M.Petic. *Electronic linguistic resources for historical standard Romanian.* In: Proceedings of the 9th International Conference "Linguistic Resources and Tools for Processing of the Romanian Language", May 16-17, 2013, Miclăuşeni-Iaşi, România, pp. 35–50.

[11] *Reusable Resources for Romanian:* http://www.math.md/elrr/.

[12] O.Burlaca, C.Ciubotaru, S.Cojocaru, A.Colesnicov, G.Magariu, L.Malahov, M.Petic, T.Verlan. *Applications based on reusable linguistic resources.* In Multilinguality and interoperability in language processing with emphasis on Romanian, Editors: D. Tufiş, C. Forăscu, Bucureşti, 2010, pp.461–476.

E.Boian, C.Ciubotaru, S.Cojocaru, A.Colesnicov, L.Malahov, M.Petic

Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova
E-mail: {lena,chebotar,Svetlana.Cojocaru,kae,mal,mirsha}@math.md

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

The optimal value of one monotone method divisor for PR voting systems

Ion Bolun, Sergiu Cebotari

Abstract: It is described the methodology and determined by simulation the optimal value of the divisor of the Mixed monotone method for decision making by voting with proportional representation – method that exceeds the widely applied Sainte-Laguë one.

Keywords: disproportion, divisor, index, monotone method, optimization, proportional representation, voting systems.

1 Introduction

When taking collective decisions, using voting systems with proportional representation (PR), to minimize the disproportion of deciders' will representation is required – disproportion caused by the character in integers of the number of deciders and that of alternative options. To minimize this disproportion, diverse methods (algorithms, "votes-decision" rules), including the Hamilton (Hare), Sainte-Laguë, d'Hondt and Huntington-Hill ones, are used.

To estimate this disproportion, each method applies an index, which may differ from one method to another. In [2], basing on a comparative multi aspectual analysis, the opportunity of using, in this aim, the Average relative deviation (ARD) index is argued. This index conveys the average relative deviation of the representation in the decision of deciders will from their mean value.

It has been proved [1, 2], that the optimum (minimum) value of ARD index is obtained when using Hamilton method. However, its use can lead, in some cases, to Alabama, of Population or of the New state

"paradoxes" [1]. The d'Hondt, Sainte-Laguë and Huntington-Hill [1] methods and the Mixed monotone one, proposed in [3], are immune to these paradoxes.

The Mixed monotone method, at parameter $c = 2$, exceeds the Sainte-Laguë one, in the sense that there are no cases, when the Sainte-Laguë method leads to a better solution in minimizing the ARD index, than the Mixed one, but there are cases, when the Mixed method leads to better solutions, than the Sainte-Laguë one. However, the optimal value of the parameter c, equal to 2, was obtained in [3], basing on some assumptions, which veracity was not strictly proved. In this paper aspects of determining, by simulation, the optimal value of the constant c, are investigated, using, as example of PR voting system, the elections in an elective body on party lists (blocks, coalitions).

2 Mixed monotone method

The problem of minimizing disproportionality in PR systems is formulated as follows [2]. Let: M – number of seats in the elective body; n – number of parties that have reached or exceeded the representation threshold; V – total valid votes cast for the n parties; V_i – total valid votes cast for party i; x_i – number of seats to be allocated to party i; I_d – index of disproportionality (Average relative deviation). Known quantities: M; n; V_i, $i = \overline{1,n}$ and

$$V_1 + V_2 + \ldots + V_n = V. \qquad (1)$$

It is required [2] to determine unknowns x_i ($i = \overline{1,n}$) – nonnegative integers, which will assure the I_d minimum value

$$I_d = \frac{\Delta d}{d}100 = \sum_{i=1}^{n}\left|v_i - m_i\right| = \frac{100}{V}\sum_{i=1}^{n}\left|V_i - Qx_i\right| \to \min, \qquad (2)$$

where $\Delta d = \frac{1}{V}\sum_{i=1}^{n} V_i\left|d_i - d\right|$, and $Q = V/M = 1/d$ is the simple quota, named the Hare one, too [1], $v_i = 100 \cdot V_i/V$ (%) and $m_i = 100 \cdot x_i/M$ (%), in compliance with the restriction:

$$x_1 + x_2 + \ldots + x_n = M. \qquad (3)$$

The Mixed monotone method for solving problem (2)-(3) is the following [3]:

1. Calculate the quantities

$$a_i := \lceil V_i/Q \rceil,\ i = \overline{1,n}, \tag{4}$$

where $\lceil z \rceil$ signifies the integer of number z. Afterwards, determine the number ΔM of still undistributed seats

$$\Delta M = M - (a_1 + a_2 + \ldots + a_n). \tag{5}$$

If $\Delta M = 0$, then the distribution has been completed and is proportional.

2. Otherwise, the ΔM seats, remaining undistributed after the first step, to assign, by one, to each of the first ΔM parties with the larger ratio $V_i/(ca_i + 1)$.

3 Aspects of RP systems simulation methodology

From the initial data of the problem (2)-(3), subject to simulation are, basically, only quantities $V_i,\ i = \overline{1,n}$, the sum of which, for each ballot, must be equal to V. Given this constraint, to generate values for $V_i,\ i = \overline{1,n}$, the following procedure is proposed:

1. Randomly generate n numbers $N_i,\ i = \overline{1,n}$ in the interval (0; 1).
2. To determine $w_i = \left\lfloor VN_i / \sum_{j=1}^{n} N_j \right\rfloor, i = \overline{1,n}$.
3. If, considering $V_i = w_i,\ i = \overline{1,n}$, the condition (1) occurs, then quantities $w_i,\ i = \overline{1,n}$ to order in decreasing, thus obtaining the expected values of quantities $V_i,\ i = \overline{1,n}$ (stop).
4. To determine $\Delta W_i = \left\lfloor VN_i / \sum_{j=1}^{n} N_j - w_i \right\rfloor i = \overline{1,n}$ and $\Delta W = \sum_{i=1}^{n} \Delta W_i$.
5. From the n values ΔW_i, to select ΔW highest values ΔW_i and for each of them to determine $W_i = w_i + 1$, and for the other $n - \Delta W$ cases to set $W_i = w_i$.
6. Quantities $W_i,\ i = \overline{1,n}$ to order in decreasing, thus obtaining the expected values of quantities $V_i,\ i = \overline{1,n}$. Stop.

For done sample size, the software application SIMRP calculates the average value $\bar{I}_d$ (c) of index $I_d(c)$. The uniform distribution and, separate, the normal one for numbers N_i, $i = \overline{1,n}$ in the interval (0, 1) are investigated.

4 Results of calculations for obtaining the value of *c*

The initial data, used for the application SIMRP, are: M = 20, 50, 100, 1000; n = 2, 4, 6, 8, 10; $V = 10^8$; sample size 20000. For each pair of values {M; n}, the quasi optimal value of parameter c, that assures the lowest value of the index $\bar{I}_d$, is calculated five times (five probes, $k = \overline{1,5}$).

Some results of the calculations of parameter c quasi optimal value, for $5 \cdot 5 \cdot 4 \cdot 2 \cdot 10^4 = 10^6$ ballots, are presented in Table 1 (normal distribution).

Table 1 Results of parameter *c* calculation (normal distribution)

M	n	Probes for determining c ($c_k, k = \overline{1,5}$)					c_{med}	δ,%
		1	2	3	4	5		
20	2	2	2	2	2	2	2	0
	4	2,04	2,03	2,03	1,99	2	2,018	0,9
	6	2,02	2,05	2,03	2,06	2,02	2,036	1,8
	8	1,99	2,01	2,05	2,04	2,02	2,022	1,1
	10	2,06	2,02	2,01	2,04	2,05	2,036	1,8
50	2	2	2	2	2	2	2	0
	4	1,98	1,95	1,99	1,96	1,95	1,966	-1,7
	6	1,99	2,02	1,97	1,97	1,98	1,986	-0,7
	8	1,98	2,01	2,02	2	2	2,002	0,1
	10	1,99	2	2,04	2,03	2,03	2,018	0,9
100	2	2	2	2	2	2	2	0
	4	2,04	2,01	2	2,04	2,03	2,024	1,2
	6	1,98	1,99	2,01	2,01	1,99	1,996	-0,2
	8	2	2	2,02	2	2,01	2,006	0,3
	10	1,99	1,99	2,02	2,04	2,01	2,01	0,5
1000	2	2	2	2	2	2	2	0
	4	2	2	2,01	1,97	1,99	1,994	-0,3
	6	1,98	1,98	2,03	2,02	1,99	2	0
	8	1,96	1,96	2,01	2	2,01	1,988	-0,6
	10	1,98	1,98	1,99	2	2	1,99	-0,5
Average							2,0046	0,23

From Table 1 one can see that at n = 2 takes place c_k = 2, $k = \overline{1,5}$. Such a situation results from relation [4]

$$c = n/\Delta M. \quad (6)$$

Indeed, at n = 2, taking into account that $\Delta M \in [1; n-1]$ (the case ΔM = 0 ensure the proportional distribution), one has ΔM = 1 and replacing in (6), we obtain c = 2.

Thus, the average size c_{med} of parameter c is 2,0046, at a relative deviation δ of c_{med} from 2, equal to 0,23%. Here, takes place:

$$c_{med} = \sum_{k=1}^{5} c_k / 5;\ \delta = (c_{med} / 2 - 1)100\%.$$

For the uniform distribution, similar calculations lead to: c_{med} = 2,005; δ = 0,25% – values falling into error of simulations.

5 Conclusion

The optimum value of parameter c, for the Mixed monotone method, proposed in [3], is equal to 2, both at the normal distribution and at the uniform one of the number of votes V_i, $i = \overline{1,n}$.

References

[1] Gallagher, M.; Mitchell, P. *The Politics of Electoral Systems*. London: Oxford University Press, 2008. 672 p.

[2] Bolun, I. *Algorithmization of optimal allocation of seats in PR systems*. In: Economica. 2011, vol. 77, nr.3. Chişinău: Editura ASEM, p. 137-152.

[3] Bolun, I. *"Votes-decision" monotone method in PR systems.* In: Economica, nr.4(78)/2011. Chişinău: Editura ASEM, p. 108-117.

[4] Bolun I. Disproportionality of some "votes-decision" rules in PR systems. In: ASEM: 20 years of ascension. - Chişinău: Editura ASEM, 2011 (Romanian).

Ion Bolun, Sergiu Cebotari

Academy of Economic Studies of Moldova
E-mail: bolun@ase.md

Modern developments of the automatic annotations of medical images

Lyudmila Burtseva

Abstract

In this work the trends of modern developments of the automatic annotations of medical images are presented and analyzed. The specific selection of discussed trends for creation and usage of medical image annotation in SonaRes system is presented.

Keywords: medical imaging, medical images annotation

1 Introduction

Medical imaging is at the heart of many modern diagnostic and treatment technologies. But, because of this prevalence, medical imaging meets the challenge of treating extremely large medical images databases. To obtain real assistance in diagnostics, the problem of supplying of every image in such database by annotation became more urgent and required as more automatic solutions as it is possible.

Based on given in the work [1] generalization of medical images annotation process, in this work the trends of modern developments of the automatic annotations of medical images are presented and analyzed.

2 Landmarks of automatic annotating path

As the summary of developments of previous decade, researchers assumed that medical images annotating process is based on relation between visual features and textual description. The acquired relations are usually saved in ontology or another type of knowledge base. Presented in the work [1] in the most general form, the components

of medical images annotation process are the following: extracting knowledge from image metadata; using image recognition to detect anatomical concepts and landmarks; reasoning the automatically detected relationships between visual and ontology-based representation of image.

One notable example is the Essence framework [2], where the process of extraction of visual abnormalities and pathologies is implemented by physician through definition of semantic terms using ontology.

Further development of automation was based on application of textual representation of visual content that is contained in DICOM headers. The structured learning framework presented in work [8] uses information from both DICOM headers and medical reports and applies for indexing medical concepts from the Unified Medical Language System meta-thesaurus. But, starting with the idea for automatic enrichment of knowledge, researchers finally provided only the framework for medics to acquire and to collect knowledge from medical cases.

Some interesting extension of description of the image semantics through ontology is given in work [3]. The employed ontologies usually contain information about anatomy and pathologies. Concerning information that is hidden in medical image, current work focused on imaging techniques (PET, CT, ultrasonography, etc.) used for medical diagnosis. The information about applied technique is saved in image modality, so the annotation, containing diagnostics rules, can be constructed automatically.

Finally, during the last two years the **controlled vocabulary** type was added to employed ontologies set. Controlled vocabulary RadLex [5] contains standardized terms for images and combined text, accompanied with content-based methods of semantic image retrieval. Recently the imaging signs were also incorporated into RadLex ontology [6]. The information will allow RadLex users to identify imaging signs by modality (e.g., ultrasound signs) and to find all signs related to specific pathology. The illustrative example of RadLex application is given by the author of mentioned above work [1], who proposed an incremen-

tal knowledge acquisition process for radiology images for automatic annotation application. The knowledge is structured by RDF Schema, employing: ontology(Foundational Model of Anatomy) for anatomical annotations; the ICD-10 (International Classification of Diseases) toolkit for disease annotations; Radlex to express visual features of a particular anatomical entity or disease. But actually, only the steps of extraction of metadata from the DICOM images and their representation according to the structured knowledge model are implemented automatically.

Considering research discussed above we can select the methods suitable for SonaRes system. The SonaRes decision support system [7] provides a second opinion for abdominal sonography specialists with necessary explanations and images that are similar to the currently examined case. To provide visual part of explanation, SonaRes system collects a set of **model** annotated images. In current SonaRes version the process of acquiring of the **model** ultrasound images is executed by the experts-physicians. But the annotating, in fact, is implemented semi-automatically through the association of visual artifacts (actually - ROIs) with textual knowledge base nodes. In SonaRes each image in retrieval results is supplied by its textual annotation and the list of other ROIs marked on current image. Textual annotation is used for explanation of diagnosis associated with presented visual content. Full list of ROIs is supplied as the tool for possible correction of pure visual content retrieval. In cases when the proposed diagnosis looks incorrect for medic-user, he can inspect the diagnoses, which are also associated with visually similar image.

3 Conclusion

The development of medical images annotation technique and tools reveals that annotation process mostly applied the ontology-like collection of terms and explanations. The requirements of current problem solution provide the extension of set of types of such collections by controlled vocabulary. In attempts to provide the automatic annotation process the step of establishment of relation between images (or ROIs)

and ontology nodes remains the unresolved problem. Although the problem of implementation of annotation process is still not completely solved, the number of useful techniques can be found in developments and be used for further development.

References

[1] D. Sonntag. *Intelligent Interaction and Incremental Knowledge Acquisition for Radiology Images,* Proceedings of SAMT, Springer (2010), http://www.dfki.de/~sonntag/samt2010-article.pdf.

[2] A. S. Barb, Chi-Ren Shyu, Y. P. Sethi. 2005. *Knowledge representation and sharing using visual semantic modeling for diagnostic medical image databases.* Trans. Info. Tech. Biomed. 9, 4 (December 2005), pp. 538–553.

[3] M. Moller, et al. *Medical Image Understanding through the Integration of Cross-Modal Object Recognition with Formal Domain Knowledge.* In Proc. of HEALTHINF 2008, vol. 1, pp. 134–141.

[4] L. Temal, M. Dojat, G. Kassel, B. Gibaud. *Towards an ontology for sharing medical images and regions of interest in neuroimaging.* J. of Biomedical Informatics 41, 5 (October 2008), pp. 766–778.

[5] http://radiopaedia.org/articles/radlex.

[6] M. W. Shore, D. L. Rubin, C. E. Kahn. *Integration of Imaging Signs into RadLex,* Journal of Digital Imaging, 2012, Volume 25, Number 1, pp. 50–55.

[7] L. Burtseva, S. Cojocaru, C. Gaindric, I. Secrieru, O. Popcova, D. Sologub. *"SONARES" a decision support system in ultrasound investigations,* Computer Science Journal of Moldova, vol.15, no.2(44), 2007, pp. 153–178.

[8] C. Lacoste, Joo-Hwee Lim, J. -P. Chevallet, D. T.H. Le. 2007. *Medical-Image Retrieval Based on Knowledge-Assisted Text and Image Indexing.* IEEE Trans. Cir. and Sys. for Video Technol. 17, 7 (July 2007), pp. 889–900.

Lyudmila Burtseva

Institute of Mathematics and Computer Science of Academy of Sciences of Moldova
E–mail: burtseva@math.md

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Knowledge Networking – a Promising Tool for Developing Moldova's R&D Potential

Igor Cojocaru

Abstract: Knowledge networking is a more and more popular research practice worldwide, being a key feature of the collaborative work regime that is nowadays prevailing in the creation and use of new scientific knowledge. This paper refers to the importance of knowledge networking to academic communities at large, and particularly to their involvement in collaborative research projects. In the Republic of Moldova's research, development and innovation (R-D-I) sphere, knowledge networking has become a priority given the country's integration in the European Research Area, fostered by its affiliation, in October, 2011, to the Framework Program 7 of the European Union. The ACADEMICA network is a dedicated ICT-based platform that can enable knowledge networking among key actors of Moldova's R-D-I system; the development path of the ACADEMICA network is presented, and its potential is highlighted with respect to supporting the necessary leapfrog from mere connectivity to sustained knowledge networking.

Keywords: knowledge networking, collaborative research, dedicated e-infrastructure, Moldova's academic community, information & communications technologies (ICT).

1 Introduction

In the 21st century, knowledge is recognised as a key driver of countries' economic competitiveness and sustainable development, (Seok, Noh and Filip, 2012) [1], as well as the prerequisite of rationality in tackling the challenges facing mankind, such as global warming, economic crises or demographic ageing. Human knowledge, in general, and the scientific and technological one, in particular, can be approached from a two-fold

perspective: stock and flow. The stock hypostasis is emphasizing the cumulative character of knowledge, as illustrated by the presence of think-tanks, idea repositories and pools of expertise; the main aspects relevant to this perspective are: expanding the amount of knowledge available and increasing its use. On the other hand, the flow hypostasis is centered on the transfer of knowledge within organizations or communities, as well as among them. While knowledge networking is involving both of the aforementioned hypostases, this paper focuses on the latter - knowledge as flow. After pointing out the role played by the respective practice within contemporary R-D-I systems, the example of pilot project ACADEMICA - the ICT-based network of Moldova's academic community - is briefly examined; the development path of this network is presented and reference is also made to its potential to support the necessary leapfrog from mere connectivity to sustained knowledge networking.

2 Knowledge networking – a promising practice in R-D-I systems

Knowledge networks are devoted to supporting information sharing and new knowledge creation, through enhancing the research and communication capacities among their members, be they individual or collective. Web 2.0 technological generation is a major enabler of knowledge networking, as it is collaboration-oriented by design, while the collaborative regime is currently prevailing in knowledge work (Razmerita, Kirchner and Sudzina, 2009) [2]. As pointed out by Lee and Lan (2007), "Through Web 2.0 platform, the traditional knowledge management with centralized knowledge repository has shifted into a more interactive conversational approach. Knowledge from specific disciplines is no longer provided and assessed solely by the domain experts, but by the peers who also possess the capabilities."

Within this context, connectivity appears to be a necessary, still insufficient condition for co-creating and sharing knowledge. The broader understanding of the notion of "networking" goes well beyond its technical layer, also encompassing behavioural features, such as openness towards collaborative research, mutual learning or open innovation.

From the point of view of ICT infrastructures available for knowledge networking, the distinction should be made between open-

access tools, such as the blogs, social networks or wikis versus platforms dedicated to certain projects or communities of users. The latter option appears to be relevant to the collaborative R&D projects unfolded by consortia or international thematic networks, such as the networks of excellence (NoE) gathering individual and/or institutional partners from member states of the European Union.

Getting engaged in knowledge networking requires partners sharing a common set of professional and ethical values, a common responsibility in managing their joint creative capacities and the portfolio of new ideas co-produced, as well as high degree of mutual trust and transparency of individual actions aimed at reaching the common goal. Given the As BRIC countries (Brasil, Russia, India and China) are of special interest from various points of view, including the dynamics of their information societies, it is worth mentioning the example of the National Knowledge Network (NKN) (www.nkn.in), developed under India's National grid Computing Initiative GARUDA (www.garudaindia.in) [3]; NKN is aimed at interconnecting all research, higher education and scientific institutions of the country, based upon an ultra high-speed backbone/data-network communication highway, thus encouraging sharing of knowledge and collaborative research (C-DAC, 2007). Almost at the same time, a scalable P2P platform for the Knowledge Grid was created in China (Zhuge et al., 2005) [4].

3 The experience with the Moldavian pilot project ACADEMICA

The ACADEMICA network was officially set up on 27 May, 2010, on the basis of the Decision no. 86 made by the Supreme Council for Science and Technological Development of the Academy of Sciences of Moldova. Its main objective was to provide a dedicated computerised infrastructure for accessing and sharing scientific and technological information by organisations within the research, development and innovation (R-D-I) sphere coordinated by the Academy of Sciences of Moldova (ASM).

The ACADEMICA network was meant to be designed and further developed in line with European and international standards of performance to date in providing ICT support for R-D-I activities.

The evolution of the ACADEMICA network in terms of its configuration, facilities provided and range of individual and institutional users is shown in Table 1.

Table 1. The development path of the ACADEMICA network

Year	Number of hubs	Types of services available	Number of organisations connected			Estimated number of users within beneficiary organisations
			Institutions in the R-D-I sphere		Educational institutions	
			R-D-I institutions	Coordination and support institutions		
2010	18	14	19	4	1	763
2011	21	14	19	6	2	925
2012	21	16	19	6	2	1014
2013	21	16	19	8	2	1030

Source: records of the Information Society Development Institute, Chisinau

Table 2 includes the main services supported by the ACADEMICA network that are relevant to knowledge networking.

Table 2. Main services supported by the ACADEMICA network

Types of services	2010	2011	2012
ICT HelpDesk	yes	yes	yes
Providing institutional e-mail accounts for members of the affiliated R-D-I institutions (number of accounts)	734	822	1012
Access to major international scientific publications databases	yes	yes	yes
Hosting websites of affiliated R-D-I institutions and their members	yes	yes	yes
Online platform for the submission of research project proposals competing for public funding (number of proposals uploaded)	374	232	278

Management of video records on domestic scientific events (number of events covered)	40	8	35
Videoconferencing among affiliated institutions	no	no	yes

It is worth mentioning that the ACADEMICA network plays a pioneering role in bringing together R-D-I institutions and educational ones, a kind of partnership that, despite its crucial importance, is still underdeveloped in Moldova; on this basis, the current weakness consisting of the insufficient research orientation of universities (Dragomirescu and Tighineanu, 2012) [5] could also be addressed.

4 From connectivity to knowledge networking - a necessary leapfrog in Moldova's R-D-I system

Based upon the experience acquired with the ACADEMICA network, the next logical step from the technical point of view would consist of rendering the same range of services available throughout the wider ICT-based network RENAM that includes all major universities of the country.

From the technical point of view, the connectivity ensured by the ACADEMICA network ranks high in terms of performance and reliability. But the capacity of the ACADEMICA is still underused, due to the low degree of sophistication of the applications currently run by the affiliated R-D-I institutions.

Knowledge networking can and should be used as a powerful tool for increasing the synergy of the country's R-D-I system, especially between the universities and the institutes coordinated by the Academy of Sciences of Moldova. Up to now, there was an emphasis on rendering research results available for use in innovation and industry; however, a priority for the near future is rendering these results available to universities, in order to update the scientific standing of course taught, also involving professors and students in research projects.

Research project consortia are operational patterns that not only allow, but also rely upon sustained knowledge sharing among partners. Therefore, another strategic axis in promoting knowledge networking in Moldova refers to boosting the involvement of domestic researchers in international partnerships, especially within the European Research Area.

5 Conclusion

As a pilot project meant to support the collaborative knowledge work within Moldova's R-D-I community, the ACADEMICA network can be considered a successful one. The experience acquired to date confirms the initial assumption that a dedicated network is the right technical solution that also best fits users' needs.

Undertaking the necessary leapfrog from mere connectivity to knowledge networking is conditioned by factors mostly non-technical. Cultivating the values of partnership for science and adopting management stakes higher than just survival are key challenges to the addressed in the near future.

References

[1] Seok, H., Nof, S. Y. and Filip, F. G. (2012). *Sustainability decision support system based on collaborative control theory*. Annual Reviews in Control, Volume 36, Issue 1, pp. 85–100.

[2] Razmerita, L., Kirchner, K. and Sudzina, F. (2009). *Personal knowledge management.* The role of Web 2.0 tools for managing knowledge at individual and organizational levels. Online Information Review, Vol. 33, No. 6, pp. 1021–1039.

[3] C-DAC (2007). *GARUDA, the nationwide computer grid, takes wing*. C-DAC press kit. Available at: http://www.cdac.in/html/press/4q07/spot619.aspx (Accessed June 17, 2013)

[4] Zhuge, H., Sun, X., Liu, J., Yao, E. and Chen, X. (2005). *A Scalable P2P Platform for the Knowledge Grid*. IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 12, pp. 1721–1740.

[5] Dragomirescu, H. and Tighineanu, I. (2012). *Risks Affecting the Development of the Information Society in the Republic of Moldova: Insights from a Delphi Survey*. In: M. Ivanovic and Z. Budimac (eds.), Proceedings of BCI '12 - Balkan Conference in Informatics, Novi Sad, Serbia, September 16 - 20, 2012, ACM, New York, pp. 82–87.

Igor Cojocaru

Information Society Development Institute
http://www.idsi.md/
E-mail: igor.cojocaru@idsi.md

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Formalization of a General SCD-based Parser for Dictionaries Using Parametrized Grammars

Neculai Curteanu, Alex Moruz, Svetlana Cojocaru

Abstract: This paper relies on previous results concerning the parsing technology of SCD (Segmentation-Cohesion-Dependency) *configurations* applied to six largest thesaurus-dictionaries, presents the project of a new, procedural DTD (Document Type Description) that extends the currently existing DTD for dictionaries, and outlines the architecture of a general SCD-based parser for dictionaries by computing *least upper bounds* (LUBs) on sense Dependency Hypergraphs (DHs) and on their formal representation as *parameterized grammars*.

Keywords: SCD parsing method for dictionaries, new DTD for dictionaries, least upper bounds on sense dependency hypergraphs, architecture of a general SCD-based parser for dictionaries.

1 Representing the Sense DHs with Parametrized Grammars

The aim of the paper is to survey preceding results and to outline the architecture of a general SCD-based (*Segmentation-Cohesion-Dependency*) parser relying on the new parsing technology of SCD configurations. In [1], [2], and similar papers we developed the dictionary-entry text version for the *parsing method* of SCD *configurations*, and applied this parsing technology to model and parse, with outstanding efficiency and portability, the following *six*, sensibly different, Romanian, French, German, and Russian largest thesaurus-dictionaries: **DLR** (The Romanian Thesaurus – new format), **DAR** (The Romanian Thesaurus – old format), **TLF** (Le Trésor de la Langue Française), **DWB** (Deutsches Wörterbuch – GRIMM), **GWB** (Göthe-Wörterbuch), and **DMLRL** (Dictionary of Modern Literary Russian Language). In the SCD method, the sense Dependency Hypergraph (DH) describes, for a dictionary, the specific pre-established dependency

relations between the sense marker classes of that dictionary. The DHs of a dictionary, on its SCD parsing configurations, are akin to its *fingerprint*.

In [3] we make a significant step forward, showing how to represent the sense DHs as parametrized grammars on the SCD configurations, *i.e.* parsing levels of the dictionaries. [3] proposes the project of a new, procedural DTD (Document Type Description) and of a general parser for dictionary entries, computed as LUBs on DHs and on the associated parametrized grammars for the dictionaries and their SCD parsing configurations. Two parametrized grammars for **DLR** are enclosed in [3], as typical samples from a larger package of combined grammars for the six considered dictionaries. This package should be constructed as the LUB of *all* the parametrized grammars written for the parsed dictionaries, on the SCD configurations / parsing levels. Such a grammar package should represent the new DTD description, thoroughly extending the current DTD in the standard TEI P5 (2007) [5], completed with the parametrized grammar of a SCD-based *parser* for the six (or more) dictionaries.

2 DH Optimization by Total-Ordering of Sense Marker Classes

In [4], we solved the following problem: how *to transform* the DHs with non-embedded recursive call cycles and non-connected hypernodes (see [4 :121, Fig. 1]), into *linearly recursive* DHs, with *completely embedded*

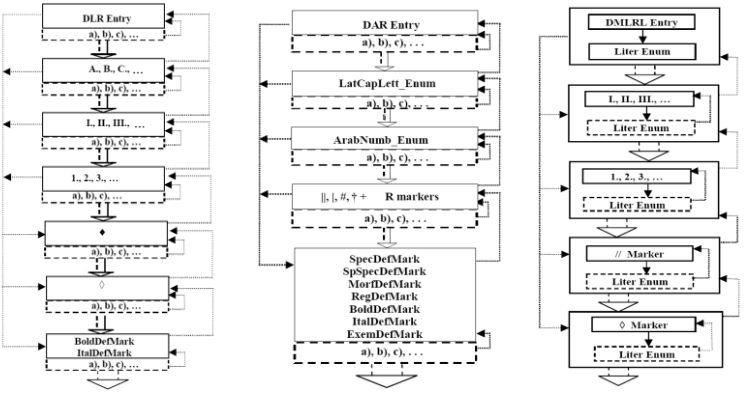


Figure 1.Optimized DHs for **DLR**, **DAR**, and **DMLRL** dictionaries [4]

call cycles, optimized DHs as in Fig. 1. These optimized DHs are suitable to LUB-computing (by unification-matching algorithms), the *LUB*(DH_i) being that DH in Fig. 2, whose parametrized grammar can represent the new, procedural DTD of the primary and secondary sense marker classes on the SCD-config2 parsing level in **DAR** thesaurus. [4] describes the *total ordering procedure* on the sense marker classes in DHs, transforming non-optimized DHs into optimized ones, as in Fig. 1.

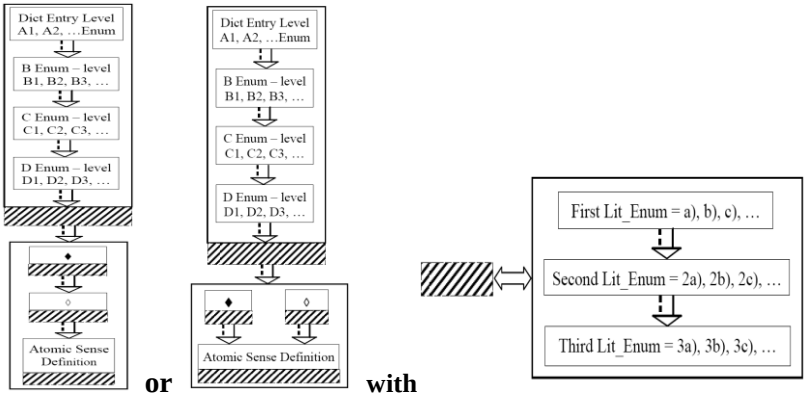


Figure 2.The two DHs as the LUB outcome of the three DHs in Fig. 1 [4]

3 New DTDs and Formal SCD-based Parser as Parametrized Grammars Attached to LUBs of Optimized DHs

The major breakthrough in [4] is the following: in the presence of non-optimized DHs, computing their parametrized grammars, and then their LUB pararametrized grammar, as in [3], is a quite intricate process. Solving the problem of DH optimization in [4] changes radically the solution for obtaining the general DTD and a formal dictionary parser: instead of computing the LUB of parametrized grammars from non-optimized DHs, we apply the optimization procedure to the DHs of the SCD configurations, compute their LUB DHs, and write their corresponding parametrized grammars, a much more efficient computational solution. We outline the following control grammar scheme for the SCD-based dictionary parser:

Table 1. Sample from the control grammar of the SCD general parser

//lexicographic segment parsing in DLR / DAR
entry → entryMarker entryRootSense entryBody entryTail
entryBody → S
S → Seg | Seg S
Seg → Mrk Root_sense Body_sense Tail_sense
Mrk → "" | depTreeNode_SCD1
Root_sense → "" | text | subSegMrk
Body_sense → "" | sense | MorphologicalPart
Tail_sense → ""
// sense tree parsing in DLR
sense → senseMarker definition sense_list
senseMarker -> depTreeNode_SCD2 //the rule that links the sense paring to SCD2 DH grammar
sense_list → sense sense_list | ""
	definition → defItem definition | defItem
	defItem → MorfDef | spSpecDef | specDef | regDef | defExemList
	defExemList → defExemPair defExemList | defExemPair
	defExemPair → quote sigle
	regDef → regDefPart regDef | regDefPart
	regDefPart → regDefPartComponent regDefPart | regDefPartComponent
	regDefPartComponent → gloss | reference | synonym | sigle | specDef | spSpecDef
	specDef → (specDefPart specDefRec) | (specDefPart)
	specDefRec → specDefPart specDefRec | specDefPart… … … … …

4 Conclusions

The new, procedural DTD and the general SCD-based parser for the largest thesaurus-dictionaries is a huge challenge since they make possible the direct comparison among the sense marker classes of such thesauri, among the lexical-semantics granularity of their lexicographic units, and provide effective formal means for the standardization of their parsing processes.

References

[1] N. Curteanu, A. Moruz, D. Trandabăț. *An Optimal and Portable Parsing Method for Romanian, French, and German Large Dictionaries*. CogAlex-2 Proceedings, The 2nd Workshop on Cognitive Aspects of the Lexicon, COLING-2010, Beijing, China, 2010, ISBN: 978-7-302-23456-2, pp. 38-47, http://www.aclweb.org/anthology-new/W/W10/W10-3407.pdf.

[2] N. Curteanu, Svetlana Cojocaru, Eugenia Burcă. *Parsing the Dictionary of Modern Literary Russian Language with the Method of SCD Configurations. The Lexicographic Modeling*. Computer Science Journal of Moldova, Academy of Sciences of Moldova, Vol. 20, No.1(58), 2012, pp. 42-81, http://www.math.md/files/csjm/v20-n1/v20-n1-(pp42-82).pdf.

[3] Neculai Curteanu, Alex Moruz. *A Procedural DTD Project for Dictionary Entry Parsing Described with Parameterized Grammars*. CogALex-3

Proceedings, The Third Workshop on Cognitive Aspects of the Lexicon, COLING-2012, Bombay, India, 2012, pp. 127-136, http://aclweb.org/anthology-new/W/W12/W12-5110.pdf.

[4] N. Curteanu, A. Moruz, Svetlana Cojocaru. *Steps to a New DTD and SCD-based Dictionary Entry Parser. Optimizing Recursiveness in Sense Dependency Hypergraphs*. Workshop on Linguistic Resources and Instruments for Romanian Language Processing, ConsILR-2013, Miclăuşeni-Iaşi, The Editorial House of "Al. I. Cuza" University, Iaşi, 2013, pp. 119-130.

[5] XCES TEI Standard, Variant P5 (2007): http://www.tei-c.org/Guidelines/P5/.

Neculai Curteanu[1], Alex Moruz[1,2], Svetlana Cojocaru[3]

[1]Institute of Computer Science, Romanian Academy, Iaşi Branch; Iaşi, România.
E-mail: ncurteanu@yahoo.com

[2]Faculty of Computer Science, "Al.I. Cuza" University, Iaşi, România.
E-mail: alex.moruz@gmail.com

[3]Institute of Mathematics and Computer Science, Chişinău, Republic of Moldova.
E-mail: svetlana.cojocaru@math.md

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Concept of computer-aided tools for diagnostics and classification of early stages of non-alcoholic fatty liver disease

Constantin Gaindric, Iulian Secrieru,
Olga Popcova, Svetlana Ţurcan

Abstract

Diagnostics, differential diagnostics and monitoring of non-alcoholic fatty liver disease (NAFLD) remain a difficult clinical problem. The concept's general goal is to design Computer-Aided Tools for Diagnostics and Classification of early stages of Non-Alcoholic Fatty Liver Disease (CATDC-NAFLD). Utilization of CATDC-NAFLD in clinical practice will allow deeper understanding of the pathogenesis of early stages of NAFLD, and will encourage new research and trials dedicated to the liver.

Keywords: early stages of NAFLD, medical diagnostics, knowledge acquisition, knowledge representation, logic inference, NAFLD classification.

1 Introduction

NAFLD is one of the major public health problems both for developed countries and for developing ones. The formalized knowledge, investigating trigger and etiological agents, will allow a better understanding of the onset and progress of fatty liver diseases.

We will concentrate on clinical features, laboratory tests and nosological signs detected by sonography. Sonography is reliable and accurate method for diagnostics of the diseases of hepato-pancreato-biliary

region with sensitivity and specificity more than 70-80%. Due to low cost, safety and accessibility, it is admitted by experts as the imaging technique of choice for screening in clinical practice.

The outcomes will be presented as computer-aided tools – CATDC-NAFLD.

The goal of CATDC-NAFLD development is to discover, formalize and reinforce medical knowledge about risk factors, diagnostics process, onset and progress of early stages of NAFLD. To achieve this goal we join corresponding research of scientists from computer science, medical and medical informatics domains. The existing experience in formalization of medical diagnostics, grounded on sonographic features and experience in creation of predictive models in medicine, will serve as a basis for successful concept development.

2 Early NAFLD diagnostics based on sonographic examination technique

NAFLD encompasses a broad spectrum of liver diseases. NAFLD is a disease that can progress, and in its natural evolution passes through several stages. It begins with free fatty acids synthesis in the liver, resulted in steatosis. Simple steatosis refers to diseases with a favorable course and possibility of a complete regression. However, in 10-20% of cases steatosis is associated with inflammation, which leads to non-alcoholic steatohepatitis [1]. In present, hepatic steatosis and steatohepatitis are considered early stages of NAFLD. The evolution of fibrosis, in turn, causes transformation into cirrhosis and cancer – advanced stages of NAFLD [2].

Despite the experimental and clinical research advances in the discovery of mechanisms of NAFLD onset and progress, at present there is neither generally accepted theory on NAFLD pathogenesis, nor complete understanding of the transition mechanisms from steatosis to steatohepatitis [3].

Filling of the liver with fat can be detected by sonography – an inexpensive, safe and informative method for NAFLD diagnostics. A

benefit is the dynamic registration of steatosis signs and stage differentiation. However, this process is not standardized enough: there are different classifications and NAFLD sonographic criteria, there is no a general algorithm for examination of patients with suspicion of NAFLD.

The recent trend in ICT applications is to obtain extended classifications, when identification and diagnostics of liver disease stages take place basing on sonographic signs, laboratory and clinical data [4]. The majority of ICT applications for NAFLD diagnostics and prediction are based on development and processing of the identification scores. Besides, the risk scores are developed to serve as predictors of NAFLD onset and progression [5]. All known scores systems, whatever exhaustive are, include the following characteristics: anthropometric, biochemical, histological and sonographic.

Existing predictive models have different levels of formalization from spreadsheet tables to computer-based toolkits. Now researchers are concentrating their efforts on finding of less evident characteristics, which can improve the scores.

3 Design of the concept of CATDC-NAFLD

The computer aided tools CATDC-NAFLD to be developed under this concept, in fact, is an information system based on expert knowledge, which will allow i) specialized physicians, ii) general practitioners, iii) patients:

- to obtain classification of early stages of NAFLD and support in complex diagnostics (a second opinion);
- to determine risks of health state aggravation and diseases progress scenarios for a particular patient;
- to guide patient in his self care activities, minimizing aggravation risks and/or improving health status.

The structure of the proposed concept is based on generally accepted technology of design and development of knowledge-based systems, SonaRes methodology (comprehensive and integrated approach

for design and development of clinical decision support systems) [6], as well as on modular programming.

SonaRes methodology combines new advanced methods for acquisition and management of medical professional knowledge with effective algorithms of sonographic images processing. SonaRes technology, in turn, provides effective algorithms for storage and documentation of specific cases, corresponding to normal/pathological states and anomalies of organs from the hepato-pancreato-biliary region, detected by sonographic diagnostics.

Under the presented concept, SonaRes methodology will be used: to select the strategy for expert knowledge acquisition, properly at the stage of clinical knowledge acquisition, and to choose the method for representation of the acquired knowledge.

In the frame of this concept SonaRes technology will be used to incorporate the kernel of SonaRes knowledge base of the liver into knowledge base of CATDC-NAFLD. This kernel includes the following data and expert knowledge: 167 facts, 31 decision rules, 87 model images, 111 images with ROIs marked.

CATDC-NAFLD consists of six main modules:

- Database, containing documentary registered cases about early stages of NAFLD;
- Knowledge base, containing declarative knowledge; formalized diagnostics process; formalized information about risk factors, onset mechanisms, and progress scenarios of early stages of NAFLD;
- Knowledge and data acquisition module;
- Module of inference, including search for similar cases and explanation for drawing conclusions;
- Predictive models of NAFLD onset and risk assessment of complications of early stages of NAFLD;
- User interface, adaptive to different categories of users (specialized physicians, general practitioners, patients).

The development of each module will be done separately. The relationships between them will be implemented at the level of exchange of data streams.

In terms of quality implementation and use CATDC-NAFLD will allow:

- clinicians and researchers in domain of sonographic diagnostics to collect and share medical data about NAFLD cases, and to obtain a consultation, basing on documentary registered cases, annotated by experts;
- operators to familiarize with well-tried recommendations of effective scanning;
- NAFLD pacients to have a guide in their self care activities to minimize aggravation risks and/or improve health status;
- health care policy makers to disseminate information about good practice to follow and malpractice to avoid.

4 Conclusion

The main scientific aim of the concept is to develop methodology of formalization and analysis of vast volume of corresponding data to find and explore NAFLD progress in order to accelerate knowledge discovering in the field of health care.

The methodology developed in the framework of the concept will include: effective methods for the acquisition of NAFLD specific professional knowledge, original form of representation of the acquired knowledge, innovative inference algorithms, feasible statistical models for risk assessment of complications. It will provide the basis for NAFLD early stages diagnostics and classification, development of scenarios of NAFLD progress and assessment of complications risk for specific individuals.

Acknowledgments. This research is supported by AŞM-BMBF project Nr. 13.820.18.02/GA "Computer-Aided Tools for Diagnostics and Classification of Early Stages of Non-Alcoholic Fatty Liver Disease with Predictive Models for Risk Assessment of Complications (CATDC-NAFLD)".

References

[1] Loria P., et al. *Practice guidelines for the diagnosis and management of nonalcoholic fatty liver disease.* A decalogue from the Italian Association for the Study of the Liver (AISF) Expert Committee. Dig Liver Dis. 2010 42(4): 272–82.

[2] Lewis J. R., et al. *Non-alcoholic Fatty Liver Disease: A Review and Update // Digestive Diseases and Sciences.* 2010, 55: 560–578.

[3] Petta S., Muratore C., Craxi A. *Non-alcoholic fatty liver disease pathogenesis: The present and the future //* Dig Liver Dis. 2009, 41 (9): 615–625.

[4] Ribeiro R., et al. *Chronic Liver Disease Staging Classification Based On Ultrasound, Clinical And Laboratorial Data,* Eighth IEEE International Symposium on Biomedical Imaging (ISBI 2011), Chicago Illinois, USA, March 30 – April 2, 2011.

[5] Brunt E.M., et al. *Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings,* Hepatology. 2011 Mar; 53(3): 810-20. doi: 10.1002/hep.24127.

[6] Burtseva L., Cojocaru S., Gaindric C., Popcova O., Secrieru Iu. *SonaRes – Diagnostic Decision Support System for Ultrasound Examination.* In: Proceedings of the 3rd International Conference on e-Health and Bioengineering (EHB 2011). Iaşi, România, November 24-26, 2011, pp. 239–242.

Constantin Gaindric[1], Iulian Secrieru[2], Olga Popcova[3], Svetlana Ţurcan[4]

[1], [2], [3] Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova
Email: gaindric@math.md

[4] State University of Medicine and Pharmacy "Nicolae Testemitanu" of the Republic of Moldova
Email: veisa@mail.ru

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Pragmatics of journalistic discourse

Daniela Gîfu, Rodolfo Delmonte

Abstract: This paper presents a multilingual system for Italian, derived from GETARUNS, which integrates a range of natural language processing tools with the intent to characterize the journalistic discourse. The method could help journalists by evidencing pragmatic aspects of the discursive abilities of speaker.

Keywords: pragmatics, journalistic discourse, Italian tool, natural language processing.

1 Introduction

This tool, GETARUNS (*General Text And Reference UNderstanding System*), implements a theoretical approach to non-compositional interactions of semantic and pragmatic interpretations in determination of noun phrase reference. Content analysis requires an extremely laborious methodology for objective interpretations (Gîfu and Cristea: 2013). We focus on two aspects critical to a successful evaluation: creation of large quantities of reasonably good training data, pragmatic analysis (sentiment) (Wiebe et. al., 2005: 165-210). This distinction is obtained by searching for factivity markers at propositional level (Sauri et. al., 2012: 261-299).

The paper is structured as follows. Section 2 shortly describes state of the art. Section 3 presents briefly the GETARUNS system, and Section 4 discusses a use case from Italian press. Finally, Section 5 presents conclusions.

2 State of the art

One aspect of the platform that we present touches a semiotic functionality. There is a large number of well-documented systems in the literature which compare with GETARUNS, in particular TACITUS and KENEL. The design of Tacitus is that the system, to the maximum possible extent, should not discard any information that might be semantically or pragmatically relevant to a full, correct interpretation. Pragmatics problems are solved by abductive inference in a pragmatics, or

interpretation, component (Hobbs et. al, 1990). Kernel's architecture is closer to ours in that semiotic tasks are segregated into separate processing modules but they are allowed to communicate. Kernel performs its analysis in two stages: first syntactic parsing "which has limited access to shallow semantic constraints for parse disambiguation" (Palmer, 1993: 17-68) and second integrated semantic and pragmatic processing which has constrained access to external knowledge sources. This rigidity of the system could be overcome in case the system could choose to delay reference resolution of all nouns as for instance in CANDIDE (Pollack, 1991: 37-82).

3 The system GETARUNS

The current version (see 3.0) of the multilingual system can be used for deep text understanding with limited domain dependent vocabulary and semantics that works for English, German and Italian. The versions can work in sequence in order to prevent failures of the deep version or they work separately to produce less constrained interpretations of the text at hand. The tool has been used for the RTE challenges and for TAC summarization tasks (Delmonte, 2011: 81-96).

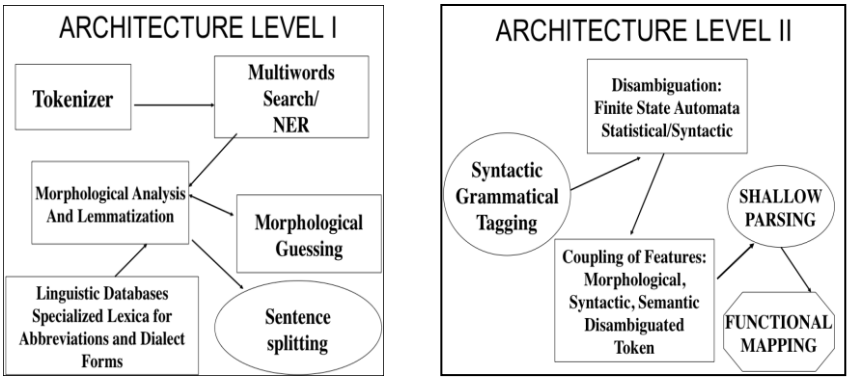


Figure 1. Italian GETARUNS – fist level one and second level.

The system works in a usual NLP pipeline that we show in Figure 1. It tokenizes the raw text and then searches for Multiwords. This computation is then extended to NER (*Named Entity Recognition*), which is performed on the basis of a big database of entities, lately released by JRC (*Joint Research Centre*). The words that are not recognized by simple

matching procedures in the big wordform dictionary (500K entries), are then passed to the morphological analyser. If this may fail, the guesser is activated, which will at first strip the word of its affixes. It will start by stripping possible prefixes and then analysing the remaining portion; then it will continue by stripping possible suffixes. If none of these succeeds, the word will be labelled as foreign word if the final character is not a vowel; a noun otherwise. We then perform tagging and chunking. In order to proceed to the semantic level, each nominal expression is classified at first on the basis of the assigned tag: proper nouns are used in the NER task. The remaining nominal expressions are classified using the classes derived from ItalWordNet (*Italian WordNet*).

4 A use case

For the elaboration of preliminary conclusions we analyzed editorials published by three Italian newspapers having similar profiles. We considered the type of article (*Corriere della Sera*, impartial, *Libero*, pro Berlusconi, and *La Republica*, against Berlusconi), and the period of time (a month before the resignation of Berlusconi, abbreviated to OMBB, the period between the presentation of Berlusconi's resignation and the appointment of Mario Monti as premier of the Italian Government, PTMB, and a month after the resignation of Berlusconi, OMAB).

Table 1. Sentiment analysis of three Italian newspapers

Newspapers / period of time	Corriere della Sera		Libero		La Republica	
	positive words	negative words	positive words	negative words	positive words	negative words
OMBB	59,85%	40,15%	62,62%	37.38%	62,93%	37,07%
PTMB	49,76%	50,24%	51,00%	49,00%	51,31%	48,69%
OMAB	49,60%	50,40%	42,16%	57,84%	42,91%	57,09%

This work consisted from two phases: we checked the automatically verbal mapping with GETARUNS system, and we mapped manually each verb, which wasn't mapped automatically. So, the values in Table 1 should be interpreted as: in OMBB, both Libero and La Repubblica has more positive contents than *Corriere della Sera*, which can be interpreted that Berlusconi's Government is considered a good one; on the contrary,

Corriere della Sera, has the highest percentage of negative opinions (the same evaluation applies for the intermediate period, PTMB). In OMAB, we assist to a change of opinions. *Corriere della Sera* becomes more positive than other newspapers so the new prime minister seems a good chance for the Italian situation; however, *Libero* – the newspaper owned by Berlusconi – becomes a lot more negative than the others.

5 Conclusion

The analysis we proposed in this paper aims at testing if a pragmatic perspective anchored in natural language processing techniques (in this case, GETARUNS system) could be of some use in evaluating journalistic discourse. In terms of pragmatics, the information has a strong impact if the event described is more unpredictable for the readers. The system helps to outline distinctive features which bring a new and, sometimes, unexpected vision upon the discursive feature of journalists' writing.

References

[1] J.R. Hobbs, M. Stickel, D. Appelt, P. Martin. *Interpretation as Abduction*, SRI International Artificial Intelligence Center Technical Note 499, 1990.

[2] J. Wiebe, T. Wilson, C. Cardie. *Annotating expressions of opinions and emotions in language in Language Resources and Evaluation*, 39(2), 2005.

[3] R. Sauri, J. Pustejovsky. *Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text*//Computational Linguistics, 38(2), 2012.

[4] M. Palmer, C. Weir, R. Passonneau, T. Finin. *The Kernel Text Understanding System in Artificial Intelligence*, 63, 1993.

[5] M. Pollack, F. Pereira. *Incremental interpretation in Artificial Intelligence*, 50, 1991.

[6] R. Delmonte, P. Vincenzo. *Opinion Mining and Sentiment Analysis Need Text Understanding in Advances in Distributed Agent-based Retrieval Tools*, Springer, 2011.

[7] D. Gîfu, D. Cristea. *Towards an Automated Semiotic Analysis of the Romanian Political Discourse,* In Computer Science Journal of Moldova, vol. 21, no. 1(61), 2013.

Daniela Gîfu[1], Rodolfo Delmonte[2]

[1] "Alexandru Ioan Cuza"University of Iaşi, E-mail: daniela.gifu@info.uaic.ro
[2] Ca'Foscari University of Venice, E-mail: delmont@unive.it

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

The Quoting-Based Algorithm for Cooperative Decision Making in Production Systems

Gennady Ginkul, Sergey Soloviev

Abstract

This paper discusses the task of cooperative decision making in distributed production systems. The quoting-based algorithm is presented which allows a production system to increase its capabilities basing on the rules borrowed from another production system. Originality of the obtained results is also discussed.

Keywords: artificial intelligence, production systems, inference engine, conflict resolution.

1 Introduction

Rapid advances in Internet technologies have opened new opportunities for enhancing traditional expert systems to distributed expert systems. The earlier success of rule-based expert systems employing more efficient inference engines pushed forward investigations of distributed production systems where multiple rule-based systems solve a common problem together. The basic idea is that a collection of production systems could be a model of group decision making, since single production system can represent logics of an individual human expert. Although researchers have offered a number of particular algorithms for cooperation between production systems, this research direction is still at the stage of formalization. Mostly the offered algorithms exploit such global ideas as data sharing [1], knowledge sharing, learning and quoting from external sources.

2 Production systems

Further we use the following restrictive assumptions about monotonic production systems [2]:
F_0 is an enumerated set of facts presented in working memory;
D is a set of potential goals;
P is a set of production rules like IF f_1 & ... & f_n THEN f;
E is an algorithm the task of which is to return just one proper rule on any step of inference.

Assumption. Rule interpreter is goal-driven; for a given potential goal d it either generates well-grounded inference (in the form of a chain of rules) supporting d or returns reject message informing it is impossible to draw the conclusion d. To search for an appropriate rule, the rule interpreter every time turns to the algorithm E.

Definition. The *h-rule* is any production rule like IF ... THEN h.

Definition. Given a set of production rules P and the goal d, we say that P_d is a *well-grounded inference for* d if P_d is minimal set of production rules which satisfy the following conditions:
(1) P_d contains just one d-rule;
(2) if P_d contains the rule IF f_1 & ... & f_n THEN f, then for every fact f_i either $f_i \in F_0$ or P_d contains just one f_i-rule.

Notice that the last definition is exhaustive: a given set of productions Q can be matching with the above conditions.

Now let us consider the algorithm E. If the fact h is selected as a hypothesis, then rule interpreter turns to the algorithm E which returns either h-rule or reject message. If $E(h)$ is reject message, then rule interpreter starts with another hypothesis. If $E(h)$ is not reject message, then rule interpreter uses $E(h)$ for construction of inference.

In its operation, the algorithm E provides for that the same rules are not selected once again. At the very beginning all the rules are marked as executable; once selected rules are marked as non-executable.

In general, the algorithm E consists of two steps:
Step 1. To form the subset $P(h)$ which consists of all h-rules from the given set of production rules P.

Step 2. If $P(h) = \emptyset$, then to return reject message.
If $P(h) \neq \emptyset$, then to return the only rule from $P(h)$ on the basis of the predefined conflict resolution mechanism [3].

3 The quoting-based algorithm

Not to concentrate on minor aspects, hereinafter we will assume a certain level of similarity of the cooperating production systems, namely (1) rules of the production systems should be generated on the basis of the same syntax, (2) despite facts of the production systems may differ, the facts identical in meaning should have the same name, (3) the production systems should have at least one common goal.

Let us consider the two production systems: $S =< F_0, P, D, E >$ and $S_1 =< F_0, Q, D_1, E_1 >$. Suppose that for the same problem situation F_0 and for the same hypothesis d the system S returned reject message and the system S_1 generated a well-grounded inference Q_d. Backing to analogy with human collective decision making, imagine that the system S nevertheless tried to generate well-grounded inference P'_d referring to the respected opinion of the system S_1. In this case we say that the production system S "quoted" the system S_1.

Formally, the task of quoting for the production system S is
$\triangleright$ to provide the well-grounded inference P_d for the goal d
$\triangleright$ taking advantage of some of the rules of the other system S_1,
$\triangleright$ with the post factum estimation of originality of the received results.

To solve the task of quoting we offer to upgrade the algorithm E of the production system S to the algorithm E^+, which explicitly uses the borrowed well-grounded inference:

`if` $E(h)$ is reject message `and` $Q_d \setminus P$ contains the h-rule q
`then` return the rule q `else` return $E(h)$.

Note, in the case $Q_d = \emptyset$ the algorithm E^+ turns into E. If $Q_d \neq \emptyset$ the production system $< F_0, P, D, E^+ >$ guarantees the well-grounded inference P' for the goal d, using the rules $P \cup Q_d$, where P'_d contains at least one rule from $Q_d \setminus P$.

The simplest formula for estimation of originality of the received result: $k = \| P'_d \setminus Q_d \| \ / \ \| P'_d \|$. In the worst case (full borrowing of knowledge): $P'_d = Q_d$ and $k = 0$. At best, but unattainable, case (no borrowing of knowledge): $k = 1$. The more sophisticated estimations can be calculated by using the measure of similarity of set Q_d and set P'_d [4].

4 Conclusion

We proposed a quoting-based algorithm which allows a production system to enhance its capabilities with the help of borrowing from another production system. In general case, the proposed algorithm does not guarantee the maximum originality for the generated inference P'_d. At the same time, realization of the algorithm requires minor changes of the traditional backward chaining engine.

References

[1] G.P. Ginkul. *The Coordinated Decision Making in Distributed Expert Systems.* Proceedings of the 54th MIREA Scientific-Technical Conference, part 1 (2004), pp. 90–94 (in Russian).

[2] Ela Kumar. *Artificial Intelligence.* I.K. International Publishing House (2008).

[3] J.C. Giarratano, G.D. Riley. *Expert Systems: Principles and Programming.* Thomson Course Technology (2005).

[4] N.G. Zagoruiko *Cognitive Data Analysis.* Acadimic Publishing House “Geo” (2012). (in Russian)

Gennady Ginkul[1], Sergey Soloviev[2]

[1]ZAO Sberbank-Technologii, Moscow, Russia
Email: ghincul@hotmail.com

[2]Lomonosov Moscow State University, Moscow, Russia
Email: soloviev@glossary.ru

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Compactified HVG for the Language Network

D.V.Lande, A.A.Snarskii

Construction of networks with text elements, words, phrases or fragments of natural language as nodes in some cases allows one to detect the structural elements of the text critical for its connected structure and find informationally significant elements, as well as words that are secondary for understanding of the text. Such networks may also be used to identify unconventional text components, such as collocations, supra-phrasal units [1], as well as for finding similar fragments in different texts [2].

There is a multitude of approaches to constructing networks from the texts (so-called language networks) and different ways of interpreting nodes and links, which causes, accordingly, different representation of such networks. Nodes are connected if corresponding words are either adjacent in the text [3, 4], or are in a single sentence [5], or are syntactically [6, 7] or semantically [8, 9] connected.

At the intersection of digital signal processing (DSP) theory and complex network theory there are several ways of constructing networks from the time series, among those are visibility graph construction methods (see survey [10]), namely the horizontal visibility graph (HVG) [11,12]. Based on these approaches, networks can also be constructed from texts in which numeric values are assigned in some manner to each word or phrase. The examples of functions assigning a number to a word are: ordinal number of a unique word in a text, length of the word, "weight" of the word in a text, e.g., generally accepted TFIDF metric (canonically, a product of the term frequency in a text fragment and a binary logarithm of the inverse number of text fragments containing this word– inverse document frequency) or its modifications [13, 14] and other word weight estimates.

In this paper, the standard deviation estimate of word weight is used for constructing word networks [15]. If all the words in the text of N words are numbered in succession (let $n = 1, \ldots N$ be the ordinal number

of the word in a text, the word position), layout of a certain word A can be designated as $A_k(n)$, where $k = 1,2,\dots,K$ denotes the number of occurrence of this word in a text, and n is a position of this word in a text. For example, $A_3(50)$ means that the third occurrence of the word A has position 50 in the text.

The distance between successive occurrences of the word in these terms would be $\Delta A_k = A_{k+1}(m) - A_k(n) = m - n$, where m and n are the positions of the $k+1$-th and k-th occurrences of the word A in the text, respectively.

Standard deviation estimate proposed in [15] is calculated as follows:

$$\sigma_A = \frac{\sqrt{\langle \Delta A^2 \rangle - \langle \Delta A \rangle^2}}{\langle \Delta A \rangle}, \quad (1)$$

where $\langle \Delta A \rangle$ is a mean value of the sequence $\Delta A_1, \Delta A_2, \dots, \Delta A_K$, $\langle \Delta A^2 \rangle$ is a mean value of $\Delta A_1^2, \Delta A_2^2, \dots, \Delta A_K^2$, and K is a number of occurrences of the word A in the text.

As opposed to other series examined in DSP theory, the series of numerical values assigned to words are transformed into horizontal visibility graphs (HVG), where each node not only has a corresponding numerical value, but also the corresponding word itself.

The process of constructing the language network using HVG consists of two stages. At the first stage, the traditional HVG is constructed [16]. To do that a series of nodes is put on the horizontal axis, where each node corresponds to a word in order of occurrence in the text, and standard deviation estimates are put on the vertical axis (visually a histogram, see Fig. 1). There is a connection between nodes, if they are in "line of sight" with each other, i.e., if they can be connected by a horizontal line that does not cross any other histogram bar. This (geometric) criterion can be written down as follows, according to [10,11]: the two nodes (words), e.g., $B_3(n)$ and C_7 $(m = n + 5)$, are connected if (see Fig. 1)

$$\sigma_n, \sigma_m > \sigma_p, \text{ for all } n < p < m. \quad (2)$$

The process of constructing can be algorithmized. For example, in Figure 1 the word node $A_1(n+2)$ is considered incident (and is connected with edges) to the words $B_3(n)$ and $C_1(n+5)$, $B_3(n)$ being the closest word to the left of $A_1(n+2)$ with a standard deviation estimate $\sigma_n = \sigma_B$ greater than that of the word A: $\sigma_{n+2} = \sigma_A$, and C_7

$(m = n + 5)$ being the closest word to the right of $A_1(n + 2)$, for which $\sigma_m > \sigma_A$.

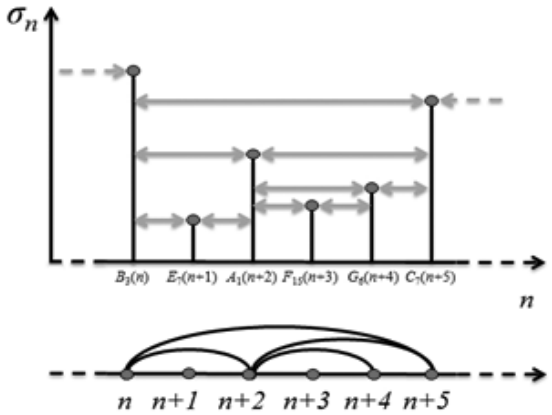


Figure 1. An example of HVG construction

At the second stage, the derived network is compactified. All the nodes corresponding to a single word, e.g., the word A, are combined into a single node (naturally, occurrence numbers and positions of the words are lost). The connections of these nodes are also combined. Note that there is no more than one edge left between any pair of nodes, multiple connections are removed (see Fig. 2).

This means, in particular, that the degree (number of connections) of the node A does not exceed the sum of degrees $\sum_k A_k(n)$. As a result, the new network of words – *compactified horizontal visibility graph* (CHVG) – is constructed (Fig. 2).

Texts used for CHVG construction were the novels "The Master and Margarita" by Mikhail Bulgakov and "Moby-Dick; or, The Whale" by Herman Melville, as well as arrays of news information from the Web.

For all CHVG networks of words described here, the degree distribution is close to power law, i.e., these networks are scale free.

For comparison, there were studied the properties of the simplest language networks, for which during the first stage of the network construction the adjacent words were connected, and, at the second stage, the network was compactified. It is obvious that the weight of a node in such network corresponds to the word frequency, and the distribution of

these weights follows the Zipf law [18]. The most connected are the nodes corresponding to the most frequently occurring words – conjunctions, prepositions, etc., which are very important for the text coherence, but are of little interest for the aspect of informational structure.

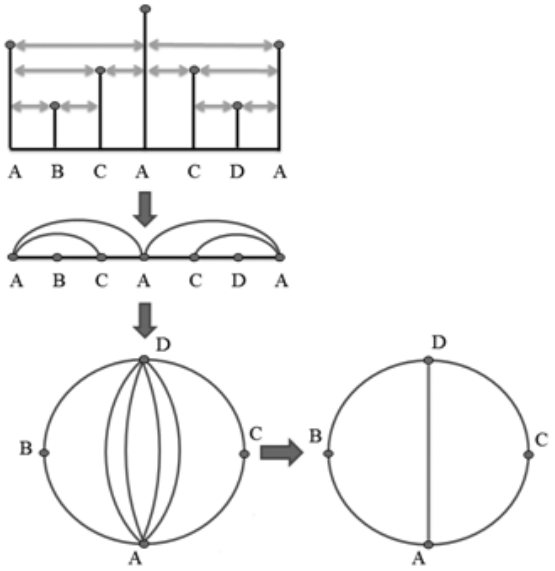


Figure 2. Two stages in construction of CHVG

Among the nodes with largest degrees, alongside with personal pronouns and other function words (particles, prepositions, conjunctions, etc.), there are the words, which determine the informational structure of the text [16, 17].

Let Ψ be a set of N different words (in our case N=100) corresponding to the largest-weight nodes of the aforementioned simple language network, and let Λ be a set of words corresponding to the largest-weight nodes of the CHVG. Then the set $\Omega = \Lambda \setminus \Psi$ will contain informational words, which are also important for the text coherence. Appendix gives juxtaposition of the top 100 largest-weight nodes for the two types of language networks constructed from the novels “The Master and Margarita” by Michael Bulgakov and “Moby-Dick; or, The Whale” by Herman Melville.

In particular, the Ω set of the CHVG built from "The Master and Margarita" contains such words as Иван, Мастер, Варенуха, Берлиоз, Бегемот, Римский, профессор, Левий, Иешуа.

The following results were obtained from studying the language networks:

1. An algorithm for constructing compactified horizontal visibility graph (CHVG) was proposed.
2. Language networks were built from different texts based on series of standard deviation estimates and CHVG.
3. In CHVG obtained from literary works, among the largest-degree nodes there are words responsible not only for the coherence of the text, but also for its informational structure. They reflect the meaning of the mentioned texts.

References

[1] Dijk van T.A. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse.*–London:Longman. – 357p. (1977).

[2] Broder A. *Identifying and Filtering Near-Duplicate Documents,* COM'00 // Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching. – P. 1-10 (2000).

[3] Ferrer-i-Cancho R., Sole R. V. *The small world of human language* // Proc. R. Soc. Lond. – B 268, 2261 (2001).

[4] Dorogovtsev S.N., Mendes J. F. F. *Language as an evolving word web* // Proc. R. Soc. Lond. – B 268, 2603 (2001).

[5] Caldeira S. M. G., Petit Lobao T. C., Andrade R. F. S., Neme A., Miranda J. G. V. *The network of concepts in written texts* // Preprint physics/0508066 (2005).

[6] Ferrer-i-Cancho R., Sole R.V., Kohler R. *Patterns in syntactic dependency networks* // Phys. Rev. E 69, 051915 (2004).

[7] Ferrer-i-Cancho R. *The variation of Zipf's law in human language.* // Phys. Rev. E 70, 056135 (2005).

[8] Motter A. E., de Moura A. P. S., Lai Y.-C., Dasgupta P. *Topology of the conceptual network of language* // Phys. Rev. E 65, 065102(R) (2002).

[9] Sigman M., Cecchi G. A. *Global Properties of the Wordnet Lexicon* // Proc. Nat. Acad. Sci. USA, 99, 1742 (2002).

[10] Nunez A. M., Lacasa L., Gomez J. P., Luque B. *Visibility algorithms: A short review* // New Frontiers in Graph Theory, Y. G. Zhang, Ed. Intech Press, ch. 6. – P. 119 – 152 (2012).

[11] Luque B., Lacasa L., Ballesteros F., Luque J. *Horizontal visibility graphs: Exact results for random time series* // Physical Review E,– P. 046103-1–046103-11 (2009).

[12] Gutin G., Mansour T., Severini S. *A characterization of horizontal visibility graphs and combinatoris on words* // Physica A, – 390 – P. 2421-2428 (2011).

[13] Jones K.S. *A statistical interpretation of term specificity and its application in retrieval* // Journal of Documentation. – 28 (1). – P. 11–21 (1972).

[14] Salton G., McGill M. J. *Introduction to Modern Information Retrieval.* – New York: McGraw-Hill. – 448 p. (1983).

[15] Ortuño M., Carpena P., Bernaola P., Muñoz E., Somoza A.M. *Keyword detection in natural languages and DNA* // Europhys. Lett, – 57(5). – P. 759-764 (2002).

[16] DijkvanT.A. *Issues in Functional Discourse Analysis* / In H. Pinkster (Ed.), Liber Amicorum for Simon Dik. – Dordrecht: Foris. – P. 27) 46. (1990).

[17] Giora R. *Segmentation and Segment Cohesion: On the Thematic Organization of the Text* // Text. An Interdisciplinary Journal for the Study of Discourse Amsterdam. – 3. – № 2. – P. 155-181 (1983).

[18] Zipf G.K. *Human Behavior and the Principle of Least Effort*. – Cambridge, MA: Addison-Wesley Press. – 573 p. (1949).

D.V.Lande[1,2], A.A.Snarskii[1,2]

[1]Institute for information recording, NAS Ukraine, Kiev, Ukraine
[2]National Technical University “Kiev Polytechnic Institute”, Kiev, Ukraine
E-mails: dwlande@gmail.com, asnarskii@gmail.com

Stackelberg equilibria set in multi-matrix mixed strategy games

Victoria Lozan, Valeriu Ungureanu

Abstract

The Stackelberg equilibrium set (SES) is described as the set of optimal solutions of a sequence of optimization problems that reduces the graph of best response mapping of the last player to the SES. The problem of SES computing in multi-matrix finite mixed-strategy games (Stackelberg games) is considered. A method for SES computing is studied.

Keywords: Noncooperative game, finite mixed-strategy game, hierarchical game, graph of best response mapping, maximum, Stackelberg equilibrium, Stackelberg equilibrium set..

1 Introduction

The Stackelberg equilibrium set (SES) may be determined by reducing the graph of best response mapping of the last player, via a set of optimization problems, to the SES. This idea serves as a basis for the method of SES computing in finite mixed-strategy n-player hierarchical games which is proposed in this paper. The results from [1, 4, 2, 3] are generalized.

Consider the game $\Gamma = \langle \mathbf{N}, \{\mathbf{X_p}\}_{p\in\mathbf{N}}, \{f_p(\mathbf{x})\}_{p\in\mathbf{N}}\rangle$, where

- $\mathbf{X_p} = \{\mathbf{x^p} \in \mathbf{R}^{\mathbf{m_p}}_{\geq} : x^p_1 + x^p_2 + \cdots + x^p_{m_p} = 1\}$ is a set of mixed strategies of player $p \in \mathbf{N}$;
- $f_p(\mathbf{x})$ is the utility function of the player $p \in \mathbf{N}$ defined on the $\mathbf{X} = \underset{\mathbf{p}\in\mathbf{N}}{\times} \mathbf{X_p}$, where $f_p(\mathbf{x}) = \sum\limits_{s_1=1}^{m_1}\sum\limits_{s_2=1}^{m_2}\ldots\sum\limits_{s_n=1}^{m_n} a^p_{s_1 s_2 \ldots s_n}\prod\limits_{p=1}^{n} x^p_{s_p}$.

The players make their moves hierarchically. **When player $p \in \mathbf{N}$ moves, players $1, 2, ..., p-1$ are leaders or predecessors of player p and players $p+1, ..., n$ are followers or successors of the player p. Players have all the information about the predecessors choices and doesn't have information about the choices of the successors, but the p^{th} player ($p < n$) has all the information about the all strategy sets and the cost functions of the players $p, p+1, ..., n$.** Without loss of generality **it is supposed that all the players maximize the values of their cost functions.**

By backward induction, every player $n, n-1, ..., 2$ computes his best move mapping and the first player computes the set of his best moves:

$$\mathbf{Br_n}(\mathbf{x^1}, ..., \mathbf{x^{n-1}}) = \underset{\mathbf{y^n} \in \mathbf{X_n}}{\text{Argmax}} \; f_n\left(\mathbf{x^1}, ..., \mathbf{x^{n-1}}, \mathbf{y_n}\right),$$

$$\mathbf{Br_i}(\mathbf{x^1}, ..., \mathbf{x^{i-1}}) = \underset{\substack{\mathbf{y^i}, ..., \mathbf{y^n}: \\ (\mathbf{x^1}, ..., \mathbf{x^{i-1}}, \mathbf{y^i}, ..., \mathbf{y^n}) \in \mathbf{Gr_{i+1}}}}{\text{Argmax}} f_i\left(\mathbf{x^1}, ..., \mathbf{x^{i-1}}, \mathbf{y^i}, ..., \mathbf{y^n}\right),$$

$$\hat{\mathbf{S}} = \underset{(\mathbf{y^1}, ..., \mathbf{y^n}) \in \mathbf{Gr_2}}{\text{Arg max}} \; f_1\left(\mathbf{y^1}, ..., \mathbf{y^n}\right),$$

where

$$\mathbf{Gr_n} = \left\{ \mathbf{x} \in \mathbf{X} : \begin{array}{l} \mathbf{x^1} \in \mathbf{X_1}, ..., \mathbf{x^{n-1}} \in \mathbf{X_{n-1}}, \\ \mathbf{x^n} \in \mathbf{Br_n}(\mathbf{x^1}, ..., \mathbf{x^{n-1}}) \end{array} \right\},$$

$$\mathbf{Gr_i} = \left\{ \mathbf{x} \in \mathbf{Gr_{i+1}} : \begin{array}{l} \mathbf{x^1} \in \mathbf{X_1}, ..., \mathbf{x^{i-1}} \in \mathbf{X_{i-1}}, \\ (\mathbf{x^i}, ..., \mathbf{x^n}) \in \mathbf{Br_i}(\mathbf{x^1}, ..., \mathbf{x^{i-1}}) \end{array} \right\}.$$

Evidently, $\mathbf{Gr_2} \subseteq \mathbf{Gr_3} \subseteq \cdots \subseteq \mathbf{Gr_n}$.

Definition 1. *Any profile $\hat{\mathbf{x}} \in \hat{\mathbf{S}}$ of the game is called Stackelberg equilibrium.*

Theorem 1. *For every finite hierarchical game the set $\hat{\mathbf{S}}$ of the Stackelberg equilibrium is non empty.*

Theorem 2. *If every strategy set* $\mathbf{X_p} \subset \mathbf{R^{m_p}}, p = \overline{1,n}$ *is compact and every cost function* $f_p(x^1, ..., x^p, ..., x^n)$, $p = \overline{1,n}$ *is continuous by* $(x^p, ..., x^n)$ *on* $\mathbf{X_p} \times \cdots \times \mathbf{X_n}$ *for every fixed* $\mathbf{x^1} \in \mathbf{X_1}, ..., \mathbf{x^{p-1}} \in \mathbf{X_{p-1}}$ *and the corresponding best respond set is compact, then the Stackelberg equilibria set* $\hat{\mathbf{S}}$ *is non empty.*

2 SES in n-player mixed-strategy games

Consider a n-player game formulated in section 1. The utility functions of the player p are linear if the strategies of the remaining players are fixed: $f_p(\mathbf{x}) = \sum_{k=1}^{m_p} \left(\sum_{s_{-p} \in S_{-p}} a^p_{1\|s_{-p}} \prod_{q=\overline{1,n}, q\neq p} x^q_{s_q} \right) x^p_k$.

Thus, at first stage of backward induction the player n solves a linear parametric problem with parameter vectors $\mathbf{x^{-n}} \in \mathbf{X_{-n}}$:

$$f_n(\mathbf{x^n}, \mathbf{x^{-n}}) \to max,\ \mathbf{x^n} \in \mathbf{X_n}. \tag{1}$$

According to the method described in [2], the solution (1) is

$$\mathbf{Gr_n} = \bigcup_{i_n \in U_n, I_n \in 2^{U_n \setminus \{i_n\}}} X(i_n I_n),$$

where $X(i_n I_n)$ is a set of solutions of the system:

$$\begin{cases} \sum\limits_{s_{-n} \in S_{-n}} (a^n_{k\|s_{-n}} - a^n_{i_n\|s_{-n}}) \prod\limits_{q=\overline{1,n-1}} x^q_{s_q} = 0, k \in I_n, \\ \sum\limits_{s_{-n} \in S_{-n}} (a^n_{k\|s_{-n}} - a^n_{i_n\|s_{-n}}) \prod\limits_{q=\overline{1,n-1}} x^q_{s_q} \leq 0, k \notin I_n \cup \{i_n\}, \\ \mathbf{er}^T \mathbf{x^r} = 1,\ \mathbf{x^r} \geq \mathbf{0}, r = \overline{1, n-1}, \\ \mathbf{ep}^T \mathbf{x^n} = 1, \mathbf{x^n} \geq \mathbf{0}, \\ x^n_k = 0, k \notin I_n \cup \{i_n\}. \end{cases}$$

and $U_n = \{i_n \in \{1, 2, \ldots, m_n\} : X_{-n}(i_n) \neq \emptyset\}$, $I_n \in 2^{U_n \setminus \{i_n\}}$, $\mathbf{ep}^T = (1, \ldots, 1) \in \mathbf{R^{m_n}}$. The player $n-1$ solves a parametric optimization problem on $\mathbf{Gr_n}$:

$$f_{n-1}(\mathbf{x^1}, \ldots, \mathbf{x^{n-2}}, \mathbf{y^{n-1}}, \mathbf{y^n}) \to max, (\mathbf{x^1}, \ldots, \mathbf{x^{n-2}}, \mathbf{y^{n-1}}, \mathbf{y^n}) \in \mathbf{Gr_n}.$$

The $n-1$ player computes the set of his best moves on the $\mathbf{Gr_n}$. **He determines the optimal values on the each non-empty component $X(i_nI_n)$, simultaneously comparing them with the preceding value, and the best one is saved**. The procedure is repeated by the $n-2, n-3, ..., 2$ players. Finally, the first player determines the set of equilibria by solving an optimization problem on $\mathbf{Gr_2}$.

3 Conclusion

The proposed approach for the SES computing works effectively in the case of two-matrix games. For the games with more than two players the performance of the method is computationally limited by the complexity of the optimization problems that must be solved. Nevertheless, the proposed method has both theoretical and practical values.

References

[1] G.Leitmann, *On Generalized Stackelberg Strategies.* Journal of Optimization Theory and Applications,Vol. 26, 1978, pp. 637–648.

[2] V. Ungureanu, *Nash equilibrium set computing in finite extended games*, Computer Science Journal of Moldova, Volume 14, no. 3 (42), 2006, pp. 345–365.

[3] V. Ungureanu, *Solution principles for simultaneous and sequential games mixture.* ROMAI Journal, 4 (2008), pp. 225–242.

[4] H. Von Stackelberg, *Marktform und Gleichgewicht.* Springer Verlag, Vienna, 1934.

Victoria Lozan[1], Valeriu Ungureanu[2]

[1] Moldova State University
Email: victoria.lozan@gmail.com

[2]Moldova State University
Email: v.a.ungureanu@gmail.com

Developing a Question Answering System

Victoria Maxim, Victoria Lazu, Victoria Bobicev

Abstract: The paper presents a question-answering system created in the project „Research in the field of Information Retrieval for question answering system creation". The system consists of question analysis and answer retrieval and extraction module. At the first stage the system is working on the basis of documents provided by IDSI (Institutul de Dezvoltare a Societăţii Informaţionale - Information Society Development Institute). The system is available online in order to collect questions from different users and to improve it by using collected information.

Keywords: automatic text processing, question answering system, automatic analysis of questions, information retrieval.

1 Introduction

At present, the most effective method of finding and acquiring information represents search engines, whose goal is to provide a list of links to websites for the user where he can find the necessary information. Often, the links proposed by the search engines do not meet the user's desire to get an adequate answer. In addition, they do not provide a specific response to user's queries, but only a set of links to web pages and the user is forced to spend time extracting the necessary information.
Thus, the next step in the information acquisition is development of systems able to answer user's questions in natural language.
A Question Answering (QA) system requires a more complex natural language processing than a simple search engine. Most question answering systems use a variety of linguistic resources to help in understanding the user's query and matching sections in documents.

Our system is created within the project "Research in information retrieval to create electronic public information system" and operates with documents provided by IDSI (Institutul de Dezvoltare a Sicietăţii Informaţionale - Information Society Development Institute) at the first

stage of the project. It is available online on http://lilu.fcim.utm.md/QASystem/ in order to collect questions from different users and to improve it by using collected information.

2 Question Processing

Almost all QA systems use question type to judge the answer. They classify the question types on the basis of types of answers users are expecting For example, "Who is …?" will be assigned to "PERSON/ORGANIZATION" type; while "When did … happen?" will be classified as "DATE/TIME" question [2].

The classical methodology of question's semantic analysis proposed in [1] requires dividing the interrogative sentence into the following items: interrogative pronouns (eg, who, what, where, when) or interrogative prepositional group (eg, in what, to whom, from where) and the inverted sentence. Inverted sentence begins with verbal group followed by the noun group. At the end of the sentence there is the so-called "gap" - free space provided for the answer. Such methodology is developed for questions in English and it was adapted to analyze the interrogative sentences in Romanian.

Table 1. The pattern of question analysis

Nr.	Asking point	Verb part	Key words
1	care	este	domeniul de reglementare al prezentei legi
2	în ce	constă	scopul prezentei legi
3	ce	reprezintă	un proiect de inovare şi transfer tehnologic
4	la propunerea cui	se formează	parcul ştiinţifico-tehnologic sau incubatorul de inovare
5	cum poate	activa	incubatorul de inovare
6	cine	își desfășoară	activitatea în parcul ştiinţifico-tehnologic

The noun group of the inverted sentence is considered the key phrase for searching response. Verb and noun groups in the reversed sentence change their places and if this question begins with a preposition, it is attached to the end. The free space for the answer is reserved after these elements.

Table 1 presents the types of questions and methodology of their analysis. After processing, the question is transformed into a query phrase for the search system.

3 Evaluation

The answer extraction module selects the sentence or paragraph that is top ranked by the weight calculated on the basis of the question's keywords. Figure 1 presents an answer of the system in the current form with some additional details in order to evaluate their performance.

Analiza întrebării şi vizualizarea răspunsului

Întrebarea pusă: Ce este cercetare

Răspunsul la întrebarea pusă s-a căutat în Codul cu privire la ştiinţă şi inovare al Republicii Moldova

Răspunsul cu scorul maxim este #7 din 22 alese:

Titlu: I Capitol: II Articol: 6

Cercetare fundamentala - activitate orientata spre dobindirea de noi cunostinte stiintifice, spre formularea si verificarea de noi ipoteze si teorii.

alte răspunsuri selectate
la documentul în care s-a căutat răspunsul
la pagina precedenta

***

Figure 1. Web page with the answer and choices offered to the user.

We performed the preliminary testing of the question-answering system developed within the project, and highlighted the strong and weak points of the system

We concentrated on the reasons of the incorrect answers extracted by the system and the most important are:

- First of all, the questions wrong formulated from grammatical point of view. Most of people are posing the questions as they speak and it's difficult to analyze a question that looks like a simple sentence.
- Incorrect parsing of the question by the system.

- The above mentioned reasons lead to wrong keywords extraction, and incorrect answers.

While we cannot influence the first reason, we focused on the question parsing and key words formation. We should improve the analysis of the question and keywords extraction in order to obtain the adequate answer.

After the complex analysis of answers we calculated that the system correctly answers only 35.8% of questions. It's a small range but it gives us new challenge to improve the system.

4 Conclusion

This article describes analysis of question answering system that is created in the project "Research in the field of Information Retrieval for question-answering system creation" The system is available online as a web application. The system was evaluated with a set of questions and the causes of wrong answers are analyzed. The next step of the project is to improve the question analysis on the base of presented evaluation.

References

[1] R. Baeza-Yates, B. Ribeiro-Nieto, *Modern Information Retrieval*, Addison Wesley, 2000.

[2] D. Tufiş, D. Ştefănescu, R. Ion, A. Ceauşu. RACAI's *Question Answering System at QA@CLEF 2007.* In Alessandro Nardi and Carol Peters, editors, Working Notes for the CLEF 2007 Workshop, pages 15–21, 2007.

[3] V.Bobicev *Preprocesarea textelor în sistemele de tip „întrebare-răspuns*. 7th International Conference on Microelectronics and Computer Science, Chişinău, 2011.

[4] V. Lazu, T. Prodan *Procesarea şi analiza întrebărilor într-un sistem de „ÎNTREBARE-RĂSPUNS"*, In International Conference ICTEI 2012, Chişinău. Vol. I, pp.376-380, 2012.

Victoria Maxim[1], Victoria Lazu[2], Victoria Bobicev[3]

[1,2,3]Technical University of Moldova
[1]E-mail: maxivica@yahoo.com
[2] E-mail: lazu.vica@mail.utm.md
[3] E-mail: vika@rol.md

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Increasing the Effectiveness of the Romanian Wordnet in NLP Applications

Verginica Barbu Mititelu

Abstract

The Romanian wordnet is a valuable linguistic resource: it contains 52357 literals distributed in 58725 synsets and covers 86175 senses, which makes it one of the richest such resources worldwide. Besides its intrinsic value, it is also used in various NLP applications developed in our group. In this paper, we present a scenario that proves that enriching the Romanian wordnet with derivational relations increases its effectiveness in Question Answering tasks. The statement is true for whatever wordnet, irrespective of the language it represents.

Keywords: wordnet, derivational relations, lexical chain.

1 Introduction

Wordnets are semantic networks containing nouns, verbs, adjectives and adverbs organized according to psycholinguistic principles, by means of semantic relations. Princeton WordNet (PWN henceforth) ([5, 1]) served as a model for more than 50 such language resources. The Romanian wordnet (RoWN henceforth) has been being developed for 12 years. The most recent activities involving it were semantic validation, alignment to PWN version 3.0 and import of DOMAINS 3.2 and of SUMO/MILO annotations (as described in [6]). The ceaseless development of RoWN is justified by its use in various applications implemented in our institute: Word Sense Disambiguation in a monolingual ([4]) or in a multilingual context ([2]), Question Answering ([3]), Machine translation.

2 Adding Value to the RoWN by Marking Derivational Relations

RoWN has been developed by following the expand method. Thus, the semantic relations in PWN have been transferred and organize the content of the RoWN, too. However, PWN also contains derivational relations. As they are lexical relations valid between literals, so language specific, they could not be transferred into RoWN. In order to mark such relations in our RoWN, we followed the steps below: i) Create a list of Romanian affixes; ii) Find possible pairs of root-derived words among the simple literals in RoWN; validate them; iii) Extract (in a set) all synsets in which each member of the above validated pairs occur; calculate the Carthesian product of the sets for a pair of literals; validate the members of the Carthesian product, thus obtaining a list of pairs of word senses between which a derivational relation was marked (notice that it is not valid at the synset level, but at the literal one); iv) Add a semantic label for each derivational relation in the form of a semantic relation in the network between the synsets to which the literals in derivational relation belong. Marking such relations in our wordnet, we increased the number of cross-part of speech relations to a high extent, as 66% of the suffixed words and 97% of the prefixed words have a different part of speech than their root.

3 Short Demonstration

For proving that by adding derivational relations to the RoWN we increase its effectiveness in NLP applications, let us consider the Question Answering task. Our corpus for searching answers can be RoWikipedia. One possible question of a user is "Cine a inventat motorul cu reacţie?" (Who invented the jet engine?). A sentence such as "Henri Coandă a inventat motorul cu reacţie." ("Henri Coandă invented the jet engine.") does not occur in RoWikipedia. However, one can find the answer in the corpus sentence "Henri Marie Coandă (n. 7 iunie 1886 - d. 25 noiembrie 1972) a fost un academician şi inginer român, pionier

al aviaţiei, fizician, inventator, inventator al motorului cu reacţie şi descoperitor al efectului care îi poartă numele." (Henri Marie Coandă (born 7 June 1886 - died 25 November 1972) was a Romanian academician and engineer, pioneer of aviation, phisicist, inventor, inventor of the jet engine and discoverer of the effect bearing his name.). The only term common to both the question and the answer is "motor cu reacţie". This unique match is not enough for giving a high score to the sentence so that it should be returned to the user. However, expanding the query, the system will also consider in the search words that are semantically related to those introduced by the user. So, one more match will be possible: between "inventat" and "inventator". In fact, the maximum number of matches is now complete, so the sentence is retained by the system.

For calculating the semantic distance or similarity between two word senses lexical chains are created, i.e., the links and nodes in the network that are crossed for getting from one node (containing one of the target word sense) into another (containing the other target word sense). The shortest the chain, the more similar the senses. For the pair "inventa" (occurring in the user's question) - "inventator" (occurring in the corpus), the lexical chain between them crossed 6 nodes and 7 relations previously: inventator(1.1) *instance_hyponym* James_Watt(x); James_Watt(x) *instance_hypernym* inginer(1.1); inginer(1.1) *hyponym* inginer_software(1); inginer_software(1) *domain_member_TOPIC* ştiinţa_calculatoarelor(x); ştiinţa_calculatoarelor(x) *domain_TOPIC* programa(3); programa(3) *hyponym* crea_mental(1); crea_mental(1) *hypernym* inventa(1). However, now that derivational relations are marked, there is a direct link between the two words: inventator(1.1) *agent* inventa(1).

4 Conclusion

Derivational relations are part of our mental lexicon and are in semantic relations to their roots, creating micro-networks. The more relations are marked in the wordnet, the more effective it becomes in

applications. We have proved this in a Question Answering scenario for Romanian.

Acknowledgments. The Sectorial Operational Program Human Resources Development, contract number SOP HRD/89/1.5/S/59758, has supported part of the research for this paper. Another part was supported by METANET4U European project, grant #270893.

References

[1] C. Fellbaum (ed.). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press (1998).

[2] R. Ion, D. Tufiş. *Multilingual versus Monolingual Word Sense Disambiguation.* International Journal of Speech Technology, vol. 12, no 2-3 (2009), pp. 113–124.

[3] R. Ion, D. Ştefănescu, A. Ceauşu, D. Tufiş, E. Irimia, V. Barbu Mititelu. *A Trainable Multi-factored QA System.* Multilingual Information Access Evaluation, Vol. I Text Retrieval Experiments Proc. LNCS, vol. 6241 (2010), pp. 257–264.

[4] R. Ion, D. Ştefănescu. *Unsupervised Word Sense Disambiguation with Lexical Chains and Graph-Based Context Formalization.* LTC 2009 Proc. LNCS, vol. 6562 (2011), pp. 435–443.

[5] G.A Miller et al. *Five Papers on WordNet*, 1993. http://wordnetcode.princeton.edu/5papers.pdf.

[6] D. Tufiş, V. Barbu Mititelu, D. Ştefănescu, R. Ion. *The Romanian Wordnet in a Nutshell.* Language Resources and Evaluation. 2013. DOI 10.1007/s10579-013-9230-7.

Verginica Barbu Mititelu

Romanian Academy Research Institute for Artificial Intelligence
E–mail: vergi@racai.ro

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Consequence Operators Related to Logical Friendliness

Alexei Y. Muravitsky

Abstract

We consider nonmonotonic formula operators related to (logical) friendliness and investigate properties of some of them in comparison with classical consequence. We give an example of nonmonotonic noncompact operator, which is strictly less than classical consequence and strictly greater than friendliness operator. Also, we generalize the friendliness relation to that which is based on any Boolean algebra and show that each operator associated with such a relation is greater than or equal to that of friendliness. We show that in the last inequality relation the equality occurs, in particular, when a Boolean algebra is finite.

Keywords: Classical consequence, nonmonotonic operator, logical friendliness, Boolean algebra, Heyting algebra.

1 Preliminaries

Nonmonotonic operators first were studied in Artificial Intelligence. They are supposed to model a way of obtaining consequences, when an ordinary logical deduction either is not applicable or does not work. In this research we consider nonmonotonic consequence operators which act in computer-related environment with limited resources. Limitation can be caused by restrictions in relation to resources of information used for premises of the operator in question. In particular, we focus on the consequence operator determined by (*logical*) *friendliness*

introduced in [1]. In the sequel, we also consider other consequence operators related to friendliness. (See about friendliness also in [2], [3], [4] and [5].)

We assume that information can be expressed in an ordinary (i.e. assertoric) propositional language based on an enumerable set **Var** of propositional variables $p, q, \ldots$ with the connectives: $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional) and $\neg$ (negation). The set of formulas is denoted by **Fm**. It is convenient to regard **Fm** sometimes as a free algebra with the connectives as signature operations and **Var** as generators, and sometimes as a set.

An operator $\boldsymbol{C}$ over **Fm** (regarding the latter as a set) is a mapping defined on the powerset of **Fm** into itself. In particular, classical consequence operator $\boldsymbol{Cn}$ is defined in terms of valuations as follows.

A *valuation* as usual is a mapping $v : \mathbf{Var} \rightarrow \mathbf{B}_1$, where $\mathbf{B}_1$ is a 2-element Boolean algebra with the carrier $\{\mathbf{0}, \mathbf{1}\}$. It is well known that v can be extended to a homomorphism from **Fm** (regarded as an algebra) into $\mathbf{B}_1$. We say that v *validates* a set Γ of formulas, in symbols $v(\Gamma) = \mathbf{1}$, if for any $\gamma \in \Gamma$, $v(\gamma) = \mathbf{1}$. As usual, a set Γ is *satisfiable* if for some valuation v, $v(\Gamma) = \mathbf{1}$.

We obtain classical consequence relation, $\Gamma \models \alpha$, if any valuation validates α, whenever it validates Γ. Then, $\boldsymbol{Cn}(\Gamma) = \{\alpha \mid \Gamma \models \alpha\}$ is called *classical consequence*.

We denote by $E(\Gamma)$, the set of variables which occur in Γ. Thus in order to check whether $v(\Gamma) = \mathbf{1}$, we can restrict the domain of v, that is before v is extended to a homomorphism, to $E(\Gamma)$. It means that instead of using ordinary valuations we can limit ourselves to *partial valuations*, the domain of which can be a proper subset of **Var**. From now on, we deal with partial valuations only. We do not exclude ordinary valuations but simply regard them as partial. Given a valuation v, we denote its domain by $E(v)$. We say that a valuation w is an extension of v, in symbols $w \geq v$ or $v \leq w$, if $E(v) \subseteq E(w)$ and for any $p \in E(v)$, $w(p) = v(p)$.

Now in terms of partial valuations, classical consequence relation can be redefined as follows: $\Gamma \models \alpha$ if, whenever valuation v with $E(v) =$

$E(\Gamma)$ validates Γ, any extension $w \geq v$ with $E(w) = E(\Gamma, \alpha)$ validates α.

We arrive at the definition of (*logical*) *friendliness*, when change the $\forall\forall$-format of the definition of classical consequence to the $\forall\exists$-format. Namely, following [1], we say that Γ is *friendly* to α, in symbols $\Gamma \mid\!\approx \alpha$, whenever v validates Γ with $E(v) = E(\Gamma)$, there is an extension $w \geq v$ such that $E(w) = E(\Gamma, \alpha)$ and w validates α. We notice that $\emptyset \mid\!\approx \alpha$ for any $\alpha \in \mathbf{Fm}$.

2 Operator $\boldsymbol{C}_F$

Let $\boldsymbol{C}$ below be an operator defined on the powerset of a set. The following properties of set operators have been a long period of time in focus.

(1) $\Gamma \subseteq \boldsymbol{C}(\Gamma)$; (*reflexivity*)
(2) $\Gamma \subseteq \Delta \Rightarrow \boldsymbol{C}(\Gamma) \subseteq \boldsymbol{C}(\Delta)$; (*monotonicity*)
(3) $\boldsymbol{C}(\boldsymbol{C}(\Gamma)) \subseteq \boldsymbol{C}(\Gamma)$; (*closure*)
(4) $\Delta \subseteq \boldsymbol{C}(\Gamma) \Rightarrow \boldsymbol{C}(\Delta) \subseteq \boldsymbol{C}(\Gamma)$; (*cumulative transitivity*)
(5) $\Delta \subseteq \boldsymbol{C}(\Gamma) \Rightarrow \boldsymbol{C}(\Gamma \cap \Delta) \subseteq \boldsymbol{C}(\Gamma)$; (*strong cumulative transitivity*)
(6) $\Gamma \subseteq \Delta \subseteq \boldsymbol{C}(\Gamma) \Rightarrow \boldsymbol{C}(\Delta) \subseteq \boldsymbol{C}(\Gamma)$; (*weak cumulative transitivity*)
(7) $\boldsymbol{C}(\Gamma) \subseteq \cup\{\boldsymbol{C}(\Gamma') \mid \Gamma' \text{ is finite}, \Gamma' \subseteq \Gamma\}$; (*compactness*)
(8) $\cup\{\boldsymbol{C}(\Gamma') \mid \Gamma' \text{ is finite}, \Gamma' \subseteq \Gamma\} \subseteq \boldsymbol{C}(\Gamma)$; (*finitary inclusion*)
(9) $\Gamma \neq \emptyset \Rightarrow \boldsymbol{C}(\Gamma) \subseteq \cup\{\boldsymbol{C}(\Gamma' \mid \Gamma' \text{ is finite}, \Gamma' \neq \emptyset, \Gamma' \subseteq \Gamma\}$; (*strong compactness*)
(10) If $\alpha \notin \boldsymbol{C}(\Gamma)$, then there is a maximal Γ^* such that $\Gamma \subseteq \Gamma^*$ and $\alpha \notin \boldsymbol{C}(\Gamma^*)$. *(maximalizability)*

It is from the area of folklore that classical operator $\boldsymbol{Cn}$ satisfies all of the properties (1) – (10) above. An essential part of this conclusion is that operator $\boldsymbol{Cn}$ is monotonic.

We define $\boldsymbol{C}_F(\Gamma) = \{\alpha \mid \Gamma \mid\!\approx \alpha\}$. Operator $\boldsymbol{C}_F$ is not monotonic, because, for example, $p \mid\!\approx q$ is true but $p, \neg q \mid\!\approx q$ is not. Also, we notice that $\boldsymbol{C}_F(\emptyset) = \mathbf{Fm}$.

Given two operators $\boldsymbol{C}$ and $\boldsymbol{C}'$, we define $\boldsymbol{C} \leq \boldsymbol{C}'$ to mean that for any set Γ, $\boldsymbol{C}(\Gamma) \subseteq \boldsymbol{C}'(\Gamma)$. It is obvious that $\boldsymbol{Cn} \leq \boldsymbol{C}_F$.

Also, it is obvious that operator $\boldsymbol{C}_F$ is reflexive. It is not that obvious at all that $\boldsymbol{C}_F$ possesses the strong compactness property, which has been proved in [1]. (The proof of strong compactness has been considerably shortened in [2], [3], and [4]; the last three papers are largely alike in content. A syntactic treatment of friendliness and a proof-theoretic proof of strong compactness is given in [5].) As to compactness of $\boldsymbol{C}_F$, if $\Gamma \not\approx \alpha$, then obviously $\emptyset \not\approx \alpha$.

Furthermore, we make the following two observations.

Proposition 2.1. *Property* (10) *holds for* $\boldsymbol{C}_F$.

In addition that $\boldsymbol{C}_F$ is not monotonic (see [1]), we have the following.

Proposition 2.2. *Neither of the properties* (3) – (6) *and* (8) *holds for* $\boldsymbol{C}_F$.

Then, summing up our comparison of the two operators $\boldsymbol{Cn}$ and $\boldsymbol{C}_F$, we obtain the following.

Proposition 2.3. *The following properties hold:*

a) *If a set* Γ *is unsatisfiable, then* $\boldsymbol{C}_F(\Gamma) = \boldsymbol{Cn}(\Gamma) = \mathbf{Fm}$.
b) *If* Γ *is satisfiable and* $E(\Gamma) = \mathbf{Var}$, *then* $\boldsymbol{C}_F(\Gamma) = \boldsymbol{Cn}(\Gamma)$ *and the output set is satisfiable.*
c) *If* Γ *is satisfiable and* $E(\Gamma) \subset \mathbf{Var}$, *then* $\boldsymbol{Cn}(\Gamma) \subset \boldsymbol{C}_F(\Gamma)$ *and the former set is satisfiable but the latter is not.*

3 Some operators related to friendliness

Let $\mathbf{B}$ be any nontrivial Boolean algebra. We define $\Gamma|\!\approx_{\mathbf{B}} \alpha$ if for any valuation v on $\mathbf{B}$ with $E(v) = E(\Gamma)$ and $v(\Gamma) = \mathbf{1}$, there is an extension $w \geq v$ with $E(w) = E(\Gamma, \alpha)$, which validates α on $\mathbf{B}$. Thus, when $\mathbf{B} = \mathbf{B}_1$, $\not\approx_{\mathbf{B}} = \not\approx$. Then, we define $\boldsymbol{C}_{\mathbf{B}}(\Gamma) = \{\alpha \mid \Gamma \not\approx_{\mathbf{B}} \alpha\}$.

Proposition 3.1. *For any nontrivial Boolean algebra* $\mathbf{B}$, $\boldsymbol{C}_F \leq \boldsymbol{C}_{\mathbf{B}}$. *If* $\mathbf{B}$ *is isomorphic to* $\mathbf{B}_1^I$, *then* $\boldsymbol{C}_F = \boldsymbol{C}_{\mathbf{B}}$.

Problem 1. *Is the equality* $\boldsymbol{C}_{\mathbf{B}} = \boldsymbol{C}_F$ *true for any Boolean algebra* $\mathbf{B}$*?*

Remark. Since any Boolean algebra is embedded into $\mathbf{B}_1^I$, for some I, so that each projection from $\mathbf{B}_1^I$ into its component is an epimorphism, we do not exclude that $\boldsymbol{C}_F = \boldsymbol{C}_{\mathbf{B}}$ for any Boolean algebra $\mathbf{B}$.

4 A nonmonotonic operator between $\boldsymbol{Cn}$ and $\boldsymbol{C}_F$

According to Proposition 2.3, operators $\boldsymbol{Cn}$ and $\boldsymbol{C}_F$ may differ only on satisfiable sets Γ with $E(\Gamma) \subset \mathbf{Var}$. We use this observation to define a nonmonotonic operator $\boldsymbol{C}_1$ such that $\boldsymbol{Cn} < \boldsymbol{C}_1 < \boldsymbol{C}_F$.

First we notice that all finite unsatisfiable sets are effectively enumerable. If we fix one such a numeration, we can talk about the first (with respect to this numeration) finite unsatisfiable set, which is contained in a given unsatisfiable set. After this remark, whenever $E(\Gamma) \subset \mathbf{Var}$ and $\boldsymbol{C}_F(\Gamma)$ is unsatisfiable, we define $\boldsymbol{C}_1(\Gamma)$ be equal to the union of $\boldsymbol{Cn}(\Gamma)$ and the first finite unsatisfiable subset of $\boldsymbol{C}_F(\Gamma)$. Otherwise, $\boldsymbol{C}_1(\Gamma) = \boldsymbol{Cn}(\Gamma)$.

Proposition 4.1. *The operator* $\boldsymbol{C}_1$ *is not monotonic and* $\boldsymbol{Cn} < \boldsymbol{C}_1 < \boldsymbol{C}_F$.

Sketch of proof. It is obvious that $\boldsymbol{Cn} < \boldsymbol{C}_1 < \boldsymbol{C}_F$. Now we show that $\boldsymbol{C}_1$ is nonmonotonic. Suppose Γ is satisfiable and $E(\Gamma) \subset \mathbf{Var}$. Then, by definition, $\boldsymbol{Cn}(\Gamma) \subset \boldsymbol{C}_1(\Gamma) \subset \boldsymbol{C}_F(\Gamma)$, therewith the two last sets are unsatisfiable. We know that Γ is contained in some maximal satisfiable set Γ'. It is obvious that $\mathbf{T} \subseteq \Gamma'$ and hence $E(\Gamma') = \mathbf{Var}$. Therefore, $\boldsymbol{C}_F(\Gamma')$ is satisfiable, for $\boldsymbol{C}_F(\Gamma') = \boldsymbol{Cn}(\Gamma')$ and, by virtue of Proposition 2.3, the latter is satisfiable. It in turn implies that $\boldsymbol{C}_1(\Gamma) \not\subseteq \boldsymbol{C}_1(\Gamma')$. Thus $\boldsymbol{C}_1$ is not monotonic.

Proposition 4.2. *The operator $\boldsymbol{C}_1$ is not compact; that is it does not satisfy* (7)*.*

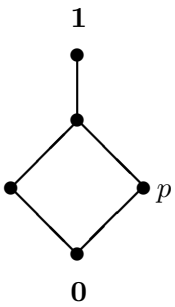
Sketch of proof. Indeed, if $\alpha \in \boldsymbol{C}_1(\Gamma)$, where Γ is satisfiable, and $E(\Gamma) \subset \mathbf{Var}$, then, by definition, $\boldsymbol{C}_1(\Gamma) = \boldsymbol{Cn}(\Gamma) \cup \Delta$, where Δ is some finite unsatisfiable subset of $\boldsymbol{C}_F(\Gamma)$. Since $\Delta \setminus \boldsymbol{Cn}(\Gamma) \neq \emptyset$, it may be the case that $\alpha \in \Delta \setminus \boldsymbol{Cn}(\Gamma)$. Then, by compactness of $\boldsymbol{C}_F$, there is a finite set $\Gamma_0 \subseteq \Gamma$ such that $\alpha \in \boldsymbol{C}_F(\Gamma_0)$. Certainly, Γ_0 is satisfiable and $E(\Gamma_0) \subset \mathbf{Var}$. Since set $\boldsymbol{C}_F(\Gamma_0)$ is unsatisfiable, by definition, $\boldsymbol{C}_1(\Gamma_0) = \boldsymbol{Cn}(\Gamma_0) \cup \Delta_1$, for some finite unsatisfiable Δ_1. Formula α does not belong to $\boldsymbol{Cn}(\Gamma_0)$, but it may not lie in Δ_1, either.

It is clear that operator $\boldsymbol{C}_1$ is not very "efficient," though it is recursive for recursive sets Γ.

Problem 2. *Find a "good" nonmonotonic operator $\boldsymbol{C}$ such that $\boldsymbol{Cn} < \boldsymbol{C} < \boldsymbol{C}_F$.*

From Proposition 3.1, we see that there is no help in defining such $\boldsymbol{C}$ via Boolean algebras.

Also, we should not expect a big fortune in using Heyting algebras, either. Indeed, let $\mathbf{H}_5$ be a 5-element cyclic Heyting algebra.



There is a valuation on $\mathbf{H}_5$, which validates $\neg\neg p \to p$ but not $\neg p \vee p$. Indeed, assigning p any atom of $\mathbf{H}_5$, the first formula takes $\mathbf{1}$, but the latter takes the pre-top element of $\mathbf{H}_5$. However, $\neg\neg p \to p \models \neg p \vee p$.

Problem 3. *What kind of structure is the family of nonmonotonic operators between* $\boldsymbol{Cn}$ *and* $\boldsymbol{C}_F$*? Can nonmonotonic operators strictly between* $\boldsymbol{Cn}$ *and* $\boldsymbol{C}_F$ *approximate* $\boldsymbol{Cn}$*?*

References

[1] D. Makinson. *Logical Friendliness and Sympathy.* In: J.-Y. Beziau (ed.), Logica Universalis, Birkhäuser Verlag, 2005, 195–224.

[2] D. Makinson. *Friendliness for Logicians.* In: We Will Show Them! Essays in Honor of Dov Gabbay, vol. 2, College Publications, 2005, 259–292.

[3] D. Makinson. *Friendliness and Sympathy in Logic.* In: http://eprints.ac.uk/3412.

[4] D. Makinson. *Logical Friendliness and Sympathy.* In: J.-Y. Beziau (ed.), Logica Universalis, Birkhäuser Verlag, 2nd edition, 2007, 191–205.

[5] A.Y. Muravitsky. *Satisfaction and Friendliness within Classical Logic: Proof-Theoretic Approach.* In: P. Bosch, D. Gabelaia, and J. Lang (editors), Proceedings of the 7th Tbilisi Symposium on Languages, Logic and Computation, Lecture Notes in Artificial Intelligence, no. 5422, Springer-Verlag, 2009, 180–192.

Alexei Y. Muravitsky

Louisiana Scholars' College
Northwestern State University
U.S.A.
E-mail: alexeim@nsula.edu

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

e-Learning System "Mental and Behavioral disorders in epilepsy"

Alexandru Popov, Maria Butnaru,
Ana Capatana, Gheorghe Capatana

Abstract

The paper describes the structure of National Clinical Concept "Mental and Behavioral Disorders in epilepsy" and e-Learning system developed as support for the first one. The system is intended for: facilitating the diagnostic process of epilepsy; increasing the quality of treatment administration and quality of patients life with epilepsy; early detection of patients with insidious onset of epilepsy; avoiding invalidation and stigmatization based on *hospitalism* effect. E-Learning system will be useful in education and within improvement of health care professionals, also to popularize among the population knowledge about prevention and patients care with epilepsy.

Keywords: epilepsy, support system for decisions, artificial intelligence, e-Learning.

1 Introduction

The studies of mental and behavioral disorders in epilepsy (abbreviated Epi) are of interdisciplinary character. The team of authors includes specialists in medicine, genetics and bioinformatics. The national clinical concept "Mental and Behavioral Disorder in Epilepsy" (further used as Concept) represent an outcome of this collaboration.

The Epilepsy (from Greek "epilamvano", which means "to grab, to gather") is a mental illness from the category of endogenous psychoses.

It is a chronic brain disease, polyetiologic, which, depending on the location of the pathological brain outbreak is manifested by repeated paroxysms (convulsive and non-convulsive) and/or by paroxysmal psychopathological manifestation as result of excessive neuronal discharges with gradual development of mental and emotional disorders [1].

2 The Rules of Users' Actions

The concept includes the description of users actions rules (hereinafter, rules) in diagnosis process, assistance and treatment of patients with psychiatric and behavioral disorders in epilepsy.

The syntax of rules is:

$$< name >:\ IF\ \ < motives >\ \ THEN\ \ < steps >,$$

where:

(i) $< name >$ is the generic name of rule. Each rule describes a stage of diagnosis, assistance and treatment;

(ii) $< motives >$ are the premises based on which the user takes the decision to apply the rule to this $< name >$. The component may also contain references, which can be consulted additionally by the user in the context of the rule.

(iii) $< steps > ::= < step > | < steps >< step >$

(iv) $< step > ::= < box > | < algorithm > | < annex >$

(v) in computer, the objects: *box*, *algorithm*, and *annex* are represented by *frames*.

The classification of mental and behavioral disorders in epilepsy includes the following hierarchy of rules (abbreviated as R):

A. Primary care institution (P)

A.1. Diagnosis

$R_{A,1,1}$: Suspected diagnosis of Epi;

$R_{A,1,2}$: Decision-making: specialist consultation and / or hospitalization.

A.2. Supervision

$R_{A,2,1}$: Supervision.

B. Specialized consultation (by psychiatrist)

B.1. Diagnosis

$R_{B,1,1}$: Confirmation of the diagnosis of Epi;

$R_{B,1,2}$: Selection of therapeutic strategy: inpatient versus outpatient.

B.2. Treatment at home

$R_{B,2,1}$: Blocking non-psychotic and behavioral symptoms;

$R_{B,2,2}$: Prophylactic therapy.

B.3. Long time supervision

$R_{B,3,1}$: Long time supervision.

B.4. Rehabilitation

$R_{B,4,1}$: Rehabilitation.

C. In-patient Care

C.1. Hospitalization

$R_{C,1,1}$: Hospitalization.

C.2. Diagnosis

$R_{C,2,1}$: Confirmation of the diagnosis of Epi.

C.3. Treatment

$R_{C,3,1}$: Blocking therapy;

$R_{C,3,2}$: Ending therapy

C.4. Discharge

$R_{C,4,1}$: Discharge;

$R_{C,4,2}$: Discharge, by district psychiatrist.

3 Antipsychotics

The e-Learning system can be regarded as a decision support system (DSS) for synthesis of programs of treatments after patients were diagnosed with mental and behavioral disorders in epilepsy. Structure and principle of operation of these systems can be found in literature [2–4]. DSS employs for this scope the knowledge regarding the antipsychotics. Antipsychotics drugs are medicines which show the following clinical effects: sedative effect, antipsychotic effect and anti-delirium effect. The terms "*sedative effect*" (E_s), "*antipsihotic effect*" (E_{ap}) and "*antidelirium effect*" (E_{ad}) can be examined as *linguistic variables* with following *universe of definition* (abbreviated as U): $U = \{0, 0.25, 0.5, 0.75, 1\}$. The elements of universe have the following meanings: (a) 1 - *full action*; (b) 0.75 - *expressed action*; (c) 0.5 - *moderate action*; (d) 0.25 - *weak action*; (e) 0 - *absent action*.

The term "Clinical effect of antipsychotics (CEA)" can be represented by a predicate:

$$CEA(M, E_s, E_{ap}, E_{ad}) = \textbf{true},$$

where: M is the name of the group of antipsychotic medicines.

The predicate executes on the computer application:

$$cea: \ M \rightarrow E_s \times E_{ap} \times E_{ad}.$$

4 An Example

Section B "Specialized consultation" of the Concept contains six rules. The first rule has the following syntax:

$$\mathrm{R}_{B,1,1}\text{: } Name_{B,1,1} \ \textbf{\textit{IF}} \ M_{B,1,1,1} \wedge M_{B,1,1,2} \ \textbf{\textit{THEN}} \ P_{B,1,1,1}, \ P_{B,1,1,2}, \ P_{B,1,1,3}, \ P_{B,1,1,4}, \ P_{B,1,1,5};$$

where:

- $Name_{B,1,1}$ – <*confirmation of Epi diagnosis*>;

- $M_{B,1,1,1}$ – <*in Epi are not present pathognomonic symptoms*>, <*references*>;
- $M_{B,1,1,2}$ – <*appropriate treatment administration requires specialized diagnostic*>, <*references*>;
- $P_{B,1,1,1}$ – Mandatory: Box 4. Anamnesis;
- $P_{B,1,1,2}$ – Mandatory: Evaluation of symptomatic (boxes 6, 7, 8, 9; table 2);
- $P_{B,1,1,3}$ – Mandatory: Laboratory examination (box 10);
- $P_{B,1,1,4}$ – Mandatory: Performing differential diagnosis (table 4);
- $P_{B,1,1,5}$ – Mandatory: Evaluation of disease prognosis (box 5).

5 Conclusions

In the article, we described National Clinical Concept "Mental and Behavior Disorder in Epilepsy" and some parts of the e-Learning system designed for:

- facilitating the diagnosis process of epilepsy;
- improving the management, treatment and life quality of patients with epilepsy;
- early detection of patients with insidious onset of epilepsy;
- avoiding invalidation and stigmatization based on hospitalism effect;
- education and improvement of qualification of health care professionals;
- popularization among population about knowledge regarding prevention and care of patients with epilepsy.

References

[1] P.M. Sarajishvili, T.S. Geladze. *Epilepsy.* - Moscow: Medicine, 1977 (in Russian).

[2] M. Butnaru. *Expert system for medical diagnosing.* Scientific Conference of master and PhD students "Research and innovation - European perspectives of evolution and integration", 23 September 2009. Conference proceedings. Real science. – Chisinau: CEP USM, 2009, pp. 136–137. (in Romanian).

[3] M. Butnaru. *Intelligent system for assistance in psychiatry,* Studia Universitatis: "Economic and hard science" no. 7 (27). – Chisinau: CEP USM, 2009, pp. 76–80. (in Romanian)

[4] M. Butnaru, A. Capatana, Gh. Capatana, O. Cobileanschi, A. Popov. *The Intelligent Support System for Remission in Patients with Psychiatric Disorders in Epilepsy.* In: International Conference on Nanotechnologies and Biomedical Engineering German-Moldovan Workshop on Novel Nanomaterials for Electronic (ICNBME-2011). Photonic and Biomedical Applications, Chisinau, July 7-8, 2011. – Chişinău: "Elan Inc" SRL, 2011, pp. 452–454.

Alexandru Popov[1,2], Maria Butnaru[3], Ana Capatana[3,4], Gheorghe Capatana[3]

[1]State Medical and Pharmaceutical University "Nicolae Testemitanu" of the Republic of Moldova
Email: san.popov90@gmail.com

[2]Psychiatric Hospital of the Republic of Moldova

[3]State University of the Republic of Moldova
Email: butnaru_m05@yahoo.com, ana_c27@yahoo.com, gh_capatana@yahoo.com

[4]Yale University, New Haven, Conecticut, USA

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Examiner – Software Product for Students' Knowledge Assessment

Larisa Pşenicinîi, Lidia Popov

Abstract: The article touches upon the problem of software usage for both current and final student's knowledge assessment. It presents as well the main principles of elaboration of this program, its requirements and possible results.

Keywords: assessment, test, software, educational program, soft.

1 Introduction

Lately, in higher education radical changes have occurred that have influenced greatly the economic, social and cultural fields. Especially noticeable is the democratization of relations between the actors of the educational process, the focus on skills acquirement and, as a result, the teacher and the student's role changing in the learning process.

One of the main issues that may arise in the relationship between the student and the lecturer is the technology used to assess students' knowledge. A basic requirement that should be specified by the assessment guideline in higher education institutions is the evaluation transparency.

Nowadays the innovative methods of assessment assisted by computer are increasingly used in the field of education. The usage of computer-based version of testing has a considerable number of advantages in comparison with the standard testing, traditional one and namely the following ones: maximum objectivity, maximum standardization, automatic recording of results, technological flexibility, systematic topics evaluation with the detection of the topics that should be reviewed, tools savings in case of a large number of students and et alia [1], [2].

The main purpose of elaborating such a project is to cover the insufficiency of course practical assignments and increase as well the effectiveness of student - teacher interaction.

This study presents the description of a software program that allows creating and updating tests, and testing students' knowledge at different subjects.

2 The Program of Students' Knowledge Evaluation Used in Higher Education

The stages of the evaluation process consist in fixing the following steps: asking key-questions, formulation of evaluation criteria, determination of the aspects that should be analyzed to answer the test items, identification of required sources of information, selection of applying methods, preparation of evaluation program etc.

Nowadays, there are many software products designed to operate the computer-based tests and organize the students' knowledge assessment process in higher education. The testing program that was described in this article is called Examiner, being easy to use, with a handy and clear interface, consisting of two parts:

- *user* (allows knowledge testing);
- *administrator* (allows test managing).

Creating an educational program for knowledge evaluation implies knowledge of programming languages such as Pascal (Borland Delphi), C (Borland C ++ Builder) and the use of HTML interfaces or some development environments – Microprogram Visual C ++, Visual Basic, etc.

To create a program, these steps are to be followed: test visualization, items completion, response selection, pending of correct answer, display of feedback message regarding the acquired score (correct or incorrect), calculation of the acquired mark.

The program was elaborated in an object-oriented programming environment, using consequently, its specific possibilities. The program interface is accessible to all users and it is well-structured. The additional objects that would distract users' attention or mislead them were omitted. Each object has a hint (hidden message that appears on the proximity of

mouse pointer) with short and clear cues in regards to what should be done for topic selection.

At the launching of the program, a window appears from which only one topic to test the knowledge is chosen.

The user part allows choosing one from two offered levels: low and average ones; by default the program suggests the low one. The knowledge testing program has a non-adaptive algorithm, the level of complexity playing an important role during the evaluation process: the difficulty of items of different levels is gradually changing. The test results together with the author's recommendations are displayed on the screen.

The complexity of items is predefined by the administrator at the introduction of items. So, the items of lower level are of a minimum difficulty, especially those with dual response, and those of average complexity of higher level – have as basis some program sequences.

For the elaboration of items, two important aspects were taken into account and namely the starting point that is the delimitation of the test area covering and the identification, within this area, of the main knowledge, abilities, and skills to be tested, they being the base of the evaluation objectives [3].

The test contains items of different types, both theoretical and practical, having an evaluation scale, each item being rated with one point.

The testing system Examiner is a program designed for the students' knowledge checking which allows using an unlimited number of tests, items and answers. The program allows managing the items of the following types: short-response items, dual-choice items, multiple choice items, keyboard response items.

At the elaboration of items two main aspects were taken into account:

1. The starting point is the area demarcation of what the test will cover. Based on Curriculum, the modules that will constitute the object evaluation will be indicated [3], [4].
2. The identification of the main knowledge, abilities, skills to be tested within this area, as they are the assessment targets (didactic tasks on cognitive levels) and, based on them, creating the items [3], [4].

To confirm the answer it's enough to click OK button. As a result, the answer is processed and the content of the next item is displayed on the

screen. In the window, below, the Correct or Incorrect message will be exhibited, depending on the chosen answer; next to it – the number of correct answers received till that moment.

The items are stored in a database, each test having 15 questions which are at random displayed on the screen (with the help of the random number generator). The large number of items in the database reduces the probability of their repetition from one test to another one.

3 Conclusion

The existence of information technologies as well as the increase of students' number has led to a considerable number of computer-based program testing. The information technologies are increasingly used to evaluate students' knowledge in higher education. This article presents the description of a testing program named Examiner that allows checking the knowledge at different subjects within the educational institutions.

The modern technologies are increasingly used to assess students' knowledge in higher education. This article describes Examiner testing system as being a universal program for students' knowledge checking. It was made a comparative analysis of these two assessment methods that have already been applied to several groups of students in our higher education and we can say it makes a difference in the acquired results.

If we analyze the results of the same group assessed through the traditional method, the grade point average is lower than that of those assessed through the non-traditional one whose grade point average is higher.

The advantage of the non-traditional method is that it's easy to check students' knowledge and it's also more objective.

From all the above said, we can affirm that the modern technology described in this paper can be used in any educational institution and at any discipline. The non-traditional method is considered to be one of the most beneficial ones from all the methods used to assess students' knowledge in higher education.

The experiment presented in this paper opens new perspectives for studying the students' development process, especially at the stages of adaptation and enrichment with upgrading procedures of academic subjects.

References

[1] V. Cabac. *Assessment by tests in education* State University „A. Russo" Balti, 1999. (romanian)

[2] V. Cabac, N. Deinego. *Testing strategies: adaptive testing*. In: Compendium of Communication International Symposium „Philosophy of education - imperative searches guidelines".: „Ion Creanga", 1997, p. 279-283. (romanian)

[3] N. Deinego. *Adaptive technology training focused on skills training*. In: Materials of International Scientific Conference „Changing Paradigm in Theory and Educational Practice", Volume II. Chisinau: CEP USM, 2008, p. 65-70. (romanian)

[4] N. Deinego. *Adaptive testing - a means of improving the quality of education*. In: Materials of the International Scientific Conference „Actual problems of the theory and practice of evaluation in education". Chisinau: Institute of Education Sciences, 2007, p. 159-162. (romanian)

Larisa Pşenicinîi[1], Lidia Popov[2]

Alecu Russo Balti State University
E-mail: [1]larisa.pshenicinii@gmail.com, [2]lidiapopov64@mail.ru

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Infinitely many precomplete relative to parametric expressibility classes of formulas in a provability logic

Andrei Rusu

Abstract

Artificial Intelligence (AI) systems simulating human behavior are often called intelligent agents. By definition, these intelligent agents exhibit some form of human-like intelligence. Intelligent agents typically represent human cognitive states using underlying beliefs and knowledge modeled in a knowledge representation language, specifically in the context of decision making. In the present paper we investigate some functional properties of the underlying knowledge representation language based on the provability logic.

Keywords: intelligent agents, modal logic, provability logic, parametric expressibility of formulas, precomplete classes of formulas).

1 Introduction

Artificial Intelligence (AI) systems simulating human behavior are often called intelligent agents. These intelligent agents exhibit somehow human-like intelligence. Intelligent agents typically represent human cognitive states using underlying beliefs and knowledge modeled in a knowledge representation language, specifically in the context of decision making [1]. In the present paper we investigate some functional properties of the underlying knowledge representation language of intelligent agents which are based on the provability logic G [2].

The notion of parametric expressibility of formulas via a system of formulas of a given logical system, which is strongly connected with the notion of precomplete class of formulas relative to parametric expressibility, was proposed in [3]. In the present paper we state that there are infinitely many precomplete with respect to parametric expressibility classes of formulas in the propositional provability logic of Gödel-Löb.

2 Definitions and notations

Provability logic. We consider the propositional provability logic G, which formulas are based on propositinal variables $p, q, r, \ldots$ and logical connectives $\&, \vee, \supset, \neg, \Delta$, its axiomes are the classical ones together with the following Δ-formulas:

$$\Delta(p \supset q) \supset (\Delta p \supset \Delta q), \quad \Delta(\Delta p \supset p) \supset \Delta p, \quad \Delta p \supset \Delta\Delta p,$$

and the rules of inference are the rules of: 1) substitution; 2) the modus ponens, and 3) the necessitation, which allows to get formula ΔA if we already have got formula A. The normal extentions of the propositional provability logic are defined as usual [2].

Diagonalizable algebras. A diagonalizable algebra [4] is a universal algebra of the form $\mathfrak{A} =< M; \&, \vee, \supset, \neg, \Delta >$, where $< M; \&, \vee, \supset, \neg >$ is a boolean algebra, and the unary operation Δ satisfies the relations

$$\Delta(x \supset x) = (x \supset x), \qquad \Delta(\Delta x \supset x) = \Delta x,$$
$$\Delta(x\&y) = (\Delta x\&\Delta y), \qquad \Delta 1_{\mathfrak{A}} = 1_{\mathfrak{A}},$$

where $1_{\mathfrak{A}}$ is the unit of $\mathfrak{A}$, which is denoted also by 1 in case the confusion is avoided.

Diagonalizable algebras are known to be algebraic models for provability logic and its extensions [5]. Obviously we can interpret any formula of the calculus of G on any diagonalizable algebra $\mathfrak{A}$. As usual a formula F is said to be valid on $\mathfrak{A}$ if for any evaluation of variables of F with elements of $\mathfrak{A}$ the value of the formula on $\mathfrak{A}$ is $1_{\mathfrak{A}}$. The set

of all valid formulas on $\mathfrak{A}$, denoted by $L\mathfrak{A}$ and referred to as the logic of the algebra $\mathfrak{A}$, forms an extension $\mathfrak{A}$ of the provability logic G [5].

We consider the diagonalizable algebra $\mathfrak{M} = (M; \&, \vee, \supset, \neg, \Delta)$ of all infinite binary sequences of the type $\alpha = (\mu_1, \mu_2, \dots), \mu_i \in \{0, 1\}, i = 1, 2, \dots$. The boolean operations $\&, \vee, \supset, \neg$ over elements of M are defined component-wise, and the operation Δ over element α we define by the equality $\Delta\alpha = (1, \nu_1, \nu_2, \dots)$, where $\nu_i = \mu_1 \& \dots \& \mu_i$. Let $\mathfrak{M}^*$ the subalgebra of $\mathfrak{M}$ generated by its zero $0_{\mathfrak{M}^*}$ element $(0, 0, \dots)$. Remark, the unit $1_{\mathfrak{M}^*}$ of the algebra $\mathfrak{M}^*$ is the element $(1, 1, \dots)$.

Parametical expressibillity [3]. They say the formula F is expressible in the logic L via a system of formulas Σ if F can be obtained from variables and Σ applying finitely many times 2 kinds of rules: a) the rule of weak substitution, b) the rule of passing to equivalent formula in L. Formula F is said to be parametrically expressible via Σ if there exist formulas $B_1, \dots, B_k, C_1, \dots, C_k, D_1, \dots, D_n$ not containing variables $\pi, \pi_1, \dots, \pi_n$ such that $B_1, \dots, B_k, C_1, \dots, C_k$ are expressible via Σ and the following first-order formulas with equalities are valid ($\wedge, \to, \sim$ are first-order connectives):

$$\begin{aligned} (F \sim \pi) &\to \wedge_{i=1}^{k}(B_i \sim C_i)[\pi_1/D_1, \dots, \pi_n/D_n] \\ \wedge_{i=1}^{k}(B_i \sim C_i) &\to (F \sim \pi) \end{aligned}$$

A system of formulas Σ is said to be complete with respect to parametric expressibility in the logic L if any formula of the calculus of L is parametrically expressible via Σ. The system Σ is precomplete with respect to parametric expressibility in L if it is not complete, but for any F, which is not parametrically expressible via Σ, the system $\Sigma \cup \{F\}$ is already parametrically complete.

3 Main result

Now we are able to formulate the main result of the present work.

Theorem 1. *The are infinitely many classes of formulas in the propositional provability logic $L\mathfrak{M}^*$ which are precompete relative to parametric expressibility in $L\mathfrak{M}^*$.*

4 Conclusion

In view of the Theorem 1 it is clear that a traditional algorithm for detecting the functional completeness with respect to parametrical expressibility in the logic $L\mathfrak{M}^*$ formulated in terms of a finite collection of precomplete classes of formulas is impossible to achieve.

Acknowledgments. Information Society Development Institute (http://www.idsi.asm.md/) has supported part of the research for this paper.

References

[1] Subrata Das. *Foundations Of Decision-Making Agents: Logic, Probability and Modality.* World Scientific Publishing Company, 2008.

[2] G. S. Boolos. *The Logic of Provability.* Cambridge University Press, 1995.

[3] A. V. Kuznetsov. *On detecting non-deducibility and non-expressibility.* In: Logical Deduction, Nauka, Moscow, 1979, pp. 5–33 (in russian).

[4] R. Magari. *The diagonalizable algebras.* Boll. Unione Matem. Italiana (1975) **12**, 117–125.

[5] R. M. Solovay. *Provability interpretations of modal logic* // Israel J. Math., 1975, **25**, p. 287–304.

Andrei Rusu

Information Society Development Institute
http://www.idsi.md/
E-mail: andrei.rusu@idsi.md

Numerical analysis of reaction of buried charge to explosive or seismic loading

Boris Rybakin, Grigore Secrieru, Elena Gutuleac

Abstract

The impulsive loading on the buried charge with exploder was investigated in this work. The loading is simulated by seismic wave or by detonation of explosive substance with trinitrotoluene parameters. The numerical investigations show: the geometry and location of booster charge strongly influences the nature and time of detonation of main charge; the characteristics of the seismic wavelet is also affect the behavior of charge.

Keywords: elastoplastic model, booster charge, explosive substance, saturated soil, seismic loading.

1 Introduction

The impact of the shock wave from the detonation charge with exploder was studied in this paper. To describe materials, composing the shell and the explosive substance (ES) we use elastoplastic model [1, 2, 3]. We propose here a second order accurate finite-difference numerical scheme that is an extension of Wilkins scheme [4]. This scheme is based on the Lagrange coordinates description. The physical processes of impulsive loading of such structures are of significantly unstable character.

2 The mathematical model

The mathematical model includes main equations, expressing the laws of preservation of mass, the quantity of motion and energy in

case of two spatial variables with cylindrical symmetry. The system of equations is complemented by the equation of state of matter and the Prandtl-Reuss equations for the elastoplastic condition with von Mises yield criterion. The behavior of various materials is described within the framework of state equation in the form of Mie-Grneisen, taking into account the complex stress-strain behavior of the matter [1, 2, 3]. The soil is represented as a three-component substance consisting of gaseous, liquid and solid components [5]. ES state equation before the detonation is written in the form of Tets law for the solid phase. The polytropic state equation was used for the detonation products. The seismic loading was modeled by specifying pressure impulse at the border of the calculation area [6].

3 Numerical results

The booster charge in the calculations may need to take in consideration the initial and the boundary conditions. The charge is being modeled as shell (R is radius, $2R$ is the length, $\frac{R}{10}$ is the thickness of the shell). By shell we denote a construction, composed of cylindrical (or plain) plates of a finite thickness. The explosive material of the main charge has parameters of trinitrotoluene (TNT). The exploder $\frac{R}{10} \times \frac{R}{4}$ is in the top of the charge. The exploder is more sensitive to detonation than filling explosive. The impact loading is simulated by explosion of booster charge or specifying pressure at the border of the calculation area. The whole structure is surrounded by water saturated ground, consisting of 10% air, 32% water and 58% quartz.

We consider following variants of calculations. Variant 1: The initiating ES is the square with size $\frac{R}{4} \times \frac{R}{4}$, situated at a distance $\frac{R}{4}$ from the axis of symmetry OY and a distance R from the booster charge shell, is detonated at a first time moment. Variant 2: The initiating ES is the rectangle with size $\frac{3R}{4} \times \frac{R}{4}$, it is at the same distance from the charge shell. Variant 3: The initiating ES with size $\frac{3R}{4} \times \frac{R}{4}$ is at the distance R from the charge shell and from the axis of symmetry OY.

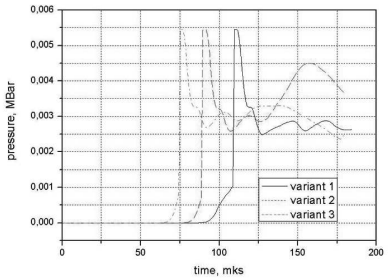


Figure 1.The dependence of the pressure/time inside the exploder

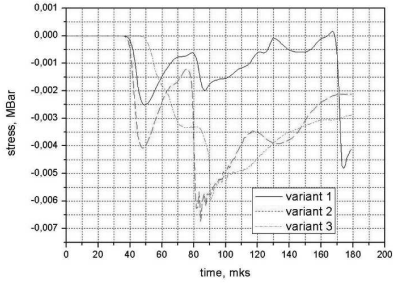


Figure 2.The normal stress component on the border of ES/shell

Figure 1 shows the pressure inside the exploder during the time of calculation. Detonation occurred inside the exploder in all variants of the calculation, but at different times: variant 1 – $t = 109$ mks., variant 2 – $t = 90$ mks., variant 3 – $t = 75$ mks. The maximum pressure inside the exploder in all variants is $P = 0.00545$ Mbar. The detonation appeared inside the shell earlier than in the exploder in variant 2.

Figure 2 shows the normal stress component (the sensor is located on the border of shell/TNT and is nearest to the axis OY). The curve (for variant 1) shows the stress change before initiation of detonation. For variant 2, 3 the detonation occurred inside the charge at the same time or before the initialization of the charge exploder (graphs on Figure 2 are different for this reason).

There is a detonation of the charge from the detonator at the time t=50 mks, when seismic loading is simulated.

4 Conclusion

Therefore, we can conclude that the geometry and location of booster charge strongly influences the nature and time of detonation of the shell. Also we can conclude the detonation of buried charge occurs

depending on characteristics of the seismic wavelet.

Acknowledgments. This work is supported by STCU project 5807.

References

[1] A.B. Kiselev, M.V.Yumashev *Deforming and fracture under impact loading. The model of thermoelastoplastic medium.*J. Appl. Mech. Tech. Phys., vol. 31 no. 5, 1990, pp. 116–123.

[2] B. Rybakin. *Numerical Modeling of Hyper-Velocity Collision of Debris Particles with Space Vehicles.* with Space Vehicles Proceedings of the 15-th IMACS World Congress, v.III, Computational Physics, Biology and Chemistry,Berlin, 1997, pp. 233–238.

[3] B. Rybakin. *Computer Modeling of Dynamic Processes.* CSJM, v.8, N 2(23), 2000, pp. 150–180.

[4] M.L. Wilkins. *Modelling the behavior of materials.* Proceeding of International Conference. V.2, London, New York, 1984, pp. 243–277.

[5] G.M. Lyakhov, G.I. Pokrovskii. *Blast Waves in Soils.*[in Russian]. Gosgortekhizdat (1962), p. 105.

[6] I.K. Naval, B.P. Rybakin, V.G. Cheban. *Mathematical modeling of ecological processes.* [in Russian]. Chishinau: Evrika, 1998, pp. 125–134.

Boris Rybakin[1], Grigore Secrieru[2], Elena Gutuleac[3]

[1] MSU, Lenin Mountains, Building 1; SRISA RAS, Moscow, Russia
Email: rybakin@vip.niisi.ru

[2] Institute of Mathematics and Computer Science, Academy of Sciences of Moldova
Email: secrieru@renam.md

[3] Institute of Mathematics and Computer Science, Academy of Sciences of Moldova
Email: elena.gutuleac@math.md

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

About new modification of Petri nets with dynamic weights of tokens

Olga Alexandrovna Savina, Tatiana Yurievna Savva

Abstract

This article introduces new modification of the apparatus of Petri nets, what allows simulating of behavior of systems where execution of operations causes a change of weights of used resources.

Keywords: formal models of computing; mathematical modeling; modification of the apparatus of Petri nets; modeling systems; dynamic weights of used resources.

Numerous researches in the field of simulation of the execution of parallel-sequential processes use the theory of Petri nets and its extensions. The wide using of the specified apparatus, on the one hand, is caused by existence of modifications of Petri nets justified within the solution of a certain circle of tasks, and on the other hand, is explained by comparative simplicity of using the apparatus as a whole. At present time the most known and often used modifications include Well Formed (Colored), Functional, Timed, Stochastic, Inhibitor, E-nets, Combi-nets, Workflow (WF) nets and others [1-4].

During the research of approaches [5] to the solution of the task of simulation of technological processes in the food industry authors have justified an approach using the apparatus of Petri nets. Thus the sequence of operations is defined by network structure. Operations correspond to transitions with a non zero time delay by analogy to Timed Petri nets. Movement of a token in some place of net points out to obtaining an ingredient with a certain quality as result of execution of previous operations. But at the same time there was defined a range

of features [6] which can not be taken into account simultaneously within a known modifications of this apparatus. For example:

– possibility of definition of the current weight of each ingredient in certain quality state in each moment of simulation time;

– definition of the needed weights of the input ingredients for every operation and the weight for the half-finished products of these operations;

– possibility of firing the transitions for given integral positive number of input sets of tokens (where input set of tokens includes certain number of tokens in all input places needed for firing the transition);

– necessity of setting source and sink places for movement of tokens in certain quality state;

– setting an enabled intervals of holding of tokens in places (movement of tokens out of these intervals is impossible).

An analysis of the approaches presented in scientific literature has allowed formulating a new modification of apparatus of Petri nets which differs from the previously proposed. The modification of Petri nets proposed by the authors is a six-tuple:

$$C = \langle S, T, M, D(S), D(T), \mu^0 \rangle.$$

Here S and T are finite sets of places and transitions: $S \cup T = \emptyset$, $S \cap T = \emptyset$.

According to the described requirements, it is reasonable to pick out 4 not being crossed subsets of places:

$$S = \langle S^{Sr}, S^{Sk}, S^{In}, S^{Bf} \rangle,$$

where

S^{Sr} – is subset of source places,

S^{Sk} – is subset of sink places,

S^{In} – is subset of places which have input and output transitions at the same time,

S^{Bf} – is subset of places which are connected with 1 or more transitions with self-loops. So places contain only a special type of tokens –

buffer tokens. Number of these tokens in buffer place defines a number of tokens from preset places allowed to be taken for every transition connected with buffer place.

If transition allows taking only one preset of tokens, it is named a simple transition. Others had to be connected with one of existing buffer places with self-loops. Algorithm of movement of buffer tokens is similar to the approach proposed by Dijkstra [1], where ones transitions have P- and V-operations which change number of tokens in semaphores.

Input and output function for transitions of proposed modified Petri net can be described with an incidence matrix:

$$M[i,j] = \begin{cases} v(i,j) & \text{if } S_i \text{ has input transition } T_j, \\ 0 & \text{if } S_i \text{ has not input and output transition } T_j, \\ -v(i,j) & \text{if } S_i \text{ has output transition } T_j, \\ \infty & \text{if } S_i \text{ has input and output transition } T_j; \end{cases}$$

where $v(i,j)$ – weight of token placed in S_i (in T_j) and needed for firing T_j (for movement in S_i as result of firing transition), $i = \overline{1,2,\dots s}$, $j = \overline{1,2,\dots t}$, $s,t \in N$.

μ^0 is an initial marking. Thus in the proposed modification there are defined two types of tokens: with dynamic weights and service-token. Initial marking include maximum number of service-tokens in buffer places.

In Figure 1 there is presented an example of the fragment of modified Petri net, where place S_2 is an input place for both transitions T_1 and T_2 and contains token with weight v^*. According to the relation between the expression $(v(2,1)+v(2,2))$ and v^*, execution of net can be extended with one of following scripts:

– if $(v(2,1)+v(2,2)) \geq v^*$, transitions T_1 and T_2 are enabled simultaneously. Thus it is necessary to divide of token in input place S_2 into 2 or more parts with weights $v(2,1)$, $v(2,2)$ and $v^* - (v(2,1)+v(2,2))$;

– if, for example, $\begin{cases} (v(2,1)+v(2,2)) < v^* \\ v(2,1) \geq v^* \end{cases}$, it is possible to fire

transition T_1 with a token having weight $v(2,1)$ in S_2. If $v^* - v(2,1) > 0$, a token having this non zero weight is held in place S_2;

– if $\begin{cases} v(2,1) < v^* \\ v(2,2) < v^* \end{cases}$, transitions T_1, and T_2 are disabled for firing.

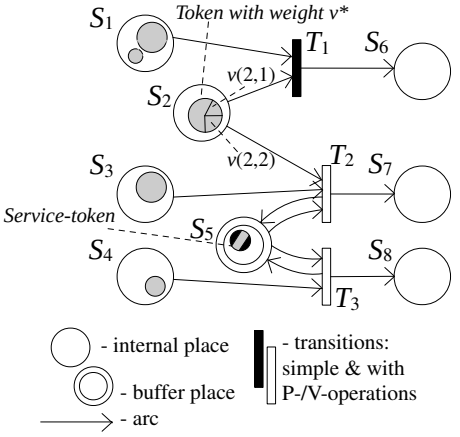


Figure 1. Example of definition of input tokens with needed weights for firing different types of transitions (situation: $v(2,1) + v(2,2) \geq v^*$)

Tokens moved in buffer places have not any weight. But the number of them in current moment of simulation explains how many transitions connected with given buffer place is enabled for one set of tokens from preset places. Also the given transitions have time interval for firing. Due to firing service-tokens are held in transition and because of it, this token cannot be used for some other transition while time interval lasts.

So, in situation from Figure 1, transition T_1 is enabled if weight of token in place S_1 is equal or more then $v(1,1)$. Transitions T_2 and T_3 are not enabled at the same time even if conditions $v(2,2)$, $v(3,2)$, $v(4,3)$ are executed, because buffer place S_5 has only one service-token. Weights of tokens, moved in output places S_6, S_7, S_8 after firing transitions, can be also defined by the incidence matrix. Sets $D(S)$ and $D(T)$ have data about timed interval for places (except buffer places)

and for transitions.

A graphical representation of the proposed modified Petri net is presented in Figure 2.

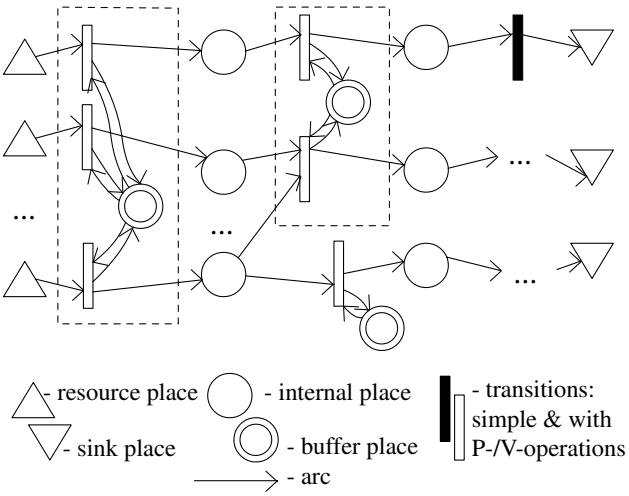


Figure 2. An example of a graphical representation of the proposed modified Petri net

Subsets of transitions connected with the same buffer place are circled in line with dotted lines. It must be told that buffer place can be set for one or more transitions.

So, the proposed modification of apparatus of Petri nets is based on the theory of Petri nets and also it borrows some ideas from Colored (two types of tokens), Timed (setting time interval for transitions and for places), *WF*-nets (resource and sink places) and the approach with setting *P-*/*V*-operations. Thus, the proposed by the authors modification of Petri nets includes the new possibility of determining weights of internal tokens in every moment of simulation. Information about needed weights is kept in the incidence matrix. Therefore, presented modified apparatus has sufficiently functional possibilities for modeling systems having the described features.

References

[1] J.L. Peterson *Petri Net Theory And The Modeling Of Systems*. Prentice-Hall, Inc., Englewood Cliffs, 1981, ISBN 0136619835.

[2] *Petri Nets World*. URL: http://www.informatik.uni-hamburg.de/TGI/PetriNets/

[3] D. René, A. Alla. *Discrete, Continuous, and Hybrid Petri Nets*. Springer-Verlag Berlin Heidelberg, 2010, ISBN 978-3-642-10668-2.

[4] I.B. Virbitskaite. *Petri nets: modifications and extensions*. Novosibirsk: NSU publishing house, 2005, 123 p. – (in Russian)

[5] T.Yu. Savva. *Mathematical modeling of the perishable raw material processing*. Applied mathematics, computer science and management. Belgorod: Belgorod publishing house, 2012, vol. 1, pp. 256–259. ISBN 978-5-9571-0556-5. – (in Russian)

[6] T.Yu. Savva. *Development of mathematical model for download equipment at the plant for the processing of perishable materials*. Informational systems and technologies, vol. 6 (74), (2012), pp. 47–56. – (in Russian)

Olga Alexandrovna Savina[1], Tatiana Yurievna Savva[2]

[1]Affiliation/Institution State University – ESPC, Russia, Oryol,
Email: o.a.savina@gmail.com

[2]Affiliation/Institution State University – ESPC, Russia, Oryol,
Email: t.savva@mail.ru

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

A simple reliable web application for Public Discourse Analysis

Ana Maria Timofciuc, Daniela Gîfu,
Corina Forăscu

Abstract: This paper presents a new tool for public speech analysis, especially its lexicon and its semantic interpretations. The concept behind this application is that every user, no matter his or hers computer science education, can use this tool and see results of the analysis of the input text. This tool will open a new road in the political discourse analysis, not only by its accuracy and richness of information, but also by its ease-of-use.

Keywords: political discourse, content analysis, domain of interest, natural language processing.

1 Introduction

The motivation for our study and for building our application lies in the importance given nowadays to the need of discourse analysis, especially in the case of political discourses, since our world is now dominated by politics. A discourse has many interesting attributes. The software we developed, PDA (*Political Discourse Analyzer*), offers the possibility to upload any kind of text and to analyze it automatically just by one click. The result of the analysis is as close as possible to a "manual" one made by a human expert.

The paper has the following structure: Section 2 shortly presents state of the art, and then Section 3 describes briefly the functionality of the web application and the associated resources for the Romanian language. Section 4 presents a use case of how the application can be used and Section 5 shows the conclusions.

2 State of the art

Our study combines tools that enable classification of the texts [Gifu, 2012] and automatic recovery methods of text information monitored by

applicable pragmalinguistics studies, that are used in natural language in order to identify features of journalistic profiles [Schiaffino et. al., 2009], [Zukermann et. al., 2001]. Of major importance are some researches based on the collection of new media texts, used in identifying characteristics underlying implementation of text classifiers [Stark et. al., 2011], [Pennebaker et. al., 2001].

3 PDA-2013 software. A short description

The PDA-2013 web application wants to do more but letting the user to do less in order to see the results the user wants. First of all, the PDA tool wants to gather many features, so that the user can make all kinds of analysis to its text, from a single place. The UI of the application is simple and user-friendly and can be basically used by anyone, linguists and computer science specialists and, why not, by all unspecialized users.

Firstly, the user can upload a text from its computer (.txt, .doc, .docx, .pdf) or can choose a text from our database. This database is made using a crawler that gets political discourses and articles from public web pages. The database can always be extended. Despite the way the user chooses to upload it, the input text is loaded and the user can see it all in a text area. The user can also edit the text.

Secondly, the user has 6 options of analysis that can be used. Five of them are web services developed by the Faculty of Computer Science from Iasi, Romania. We incorporated them in our tool because they offer a variety of features for discourse analysis. These 5 web services are: Discourse Parser, PoS Tagger, NP Chunker, Dependency Parser, Clause Splitter, all for Romanian. The 6th option - See discourse statistics - is the one that represents most of our interest and is the basis of the PDA application. We want to create a lexicon from some given texts with known domain. Firstly, we want to find manually some keywords that represent the category of the type of the text and insert it in a database. Then, we want to increase our lexicon by automatically adding keywords linked to categories. In this way, we want to be able to categorize and analyze texts in an automatical way.

Using this lexicon stored in our database, we give to the user the results of the input text analysis. When the user presses the *See results*

button, a Bar Chart appears that shows the categories related to the text and the number of words from the texts that appear in each category. The user can see the words in the input text that appear in each category by clicking the bar in the chart corresponding to that category and can also see the frequency of the word and of the class. In addition to this, the user can choose from a variety of Charts to see the results: Line, Column, Area, Pie Charts and even download the results in .cvs format.

Figure 1. The PDA interface

The Romanian lexicon contains now approximately 5700 lemmas and 33 semantic classes.

4 Use cases

For example, a user wants to see an article about a Romanian politician. He/she goes to the *Get text from Database* section, writes the name of the politician in the Search field. Now, the articles related to the searched name appear. By clicking an article, the text appears now in the *Plain Text*

Area, where the user can do further editing, if he/she wants. By clicking the *See results* button[1], a new window appears where the user can see the input text and the results. The user can click on the charts to see more information and can also change the type of the chart.

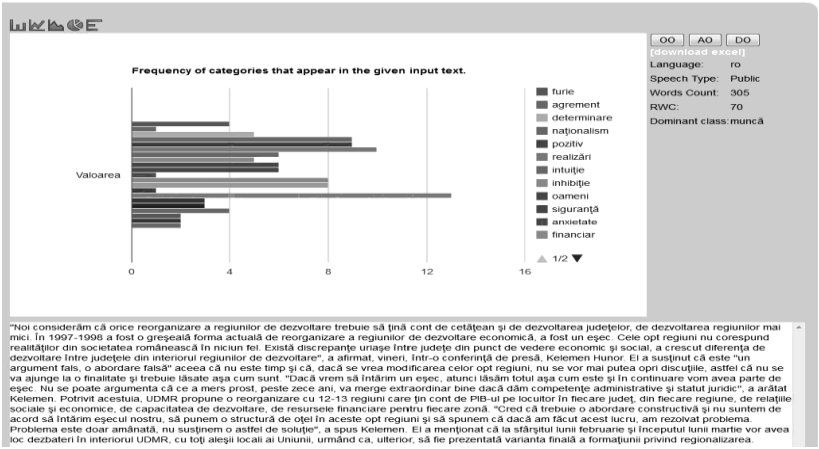


Figure 2. Text analysis

5 Conclusion

The PDA web application has a range of features that make it reliable and easy to use as a tool for discourse analysis. Its interface is user-friendly, elegant and the response of the application to the user's requests is fast. It can also be adapted to new domains and new languages and even create a word sense disambiguation module in order to determine the correct senses, in context, of those words which are ambiguous between different semantic classes belonging to the lexicon, or between classes in the lexicon and outside the lexicon.

References

[1] D. Gîfu, D. Cristea, D. *Multi-dimensional analysis of political language* in J.J. (J. Hyuk) Park, V. Leung, T. Shon, C. Wang (Eds.), Future Information

[1] The frequency of a word = (how many times it appears in the text) / (total number of words from the text) and the frequency of a class = (number of words from that category in the text) / (total number of words from the text).

Technology, Application, and Service: FutureTech2012 (volume 1) Springer, Netherlands, pp. 213-221, 2012.

[2] J.W. Pennebaker, M.E. Francis, R.J. Booth. *Linguistic Inquiry and Word Count* – LIWC2001, Mahwah, NJ, Erlbaum Publishers, 2001.

[3] S. Schiaffino, A. Amandi, A.. *Intelligent user profiling in Artificial intelligence*, Max Bramer (Ed.). Lecture Notes In Computer Science, Vol. 5640. Springer-Verlag, Berlin, Heidelberg, pp. 193-216, 2009.

[4] A. Stark, C. Dürscheid. *SMS4science: An international corpus-based texting project and the specific challenges for multilingual Switzerland* in Crispin Thurlow/Kristine Mroczek (Hrsg.): Digital Discourse. Language in the New Media. Oxford: Oxford University Press, pp. 299-320, 2011.

[5] I. Zukerman, D. Albrecht. Predictive Statistical Models for User Modeling in User Modeling and User-Adapted Interaction, 11(1-2) pp. 5-18, 2001.

Ana Maria Timofciuc, Daniela Gîfu, Corina Forăscu

Universitatea "Alexandru Ioan Cuza" din Iaşi
E-mail: {ana.timofciuc, daniela.gifu, corinfor}@info.uaic.ro

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Workflow Petri nets used in modeling of Parallel architectures

Inga Titchiev, Nicolae Iliuha, Mircea Petic

Abstract

In this article we intend to illustrate the relationship between distributed systems and workflow Petri nets in order to model parallel routing through them. Using of workflow Petri nets is another way to organize interactions in parallel architectures based on natural description of parallelism.

Keywords: workflow Petri nets, parallel computing.

1 Introduction

Adoption of cloud environment is quickly transformed from a competitive advantage to an operational necessity, facilitating innovation and thus allowing definition of new computational models and occurrence of new working opportunities. Also not less significant costs reduction by using techniques of parallel processing and storing data in the cloud, which being combined with increasing performance requirements of applications led to their successful use in various fields (bioinformatics, physics, mathematical modeling, web servers and database, the optimization of business decisions, medicine).

The development of computer is characterized not only by increasing the number of elements involved in data processing but by presenting the relationship between them and the management of interactions with a more complex structure. The quality of such new interactions contributed to the occurrence of new problems related to the analysis, modeling and representation of causal relations by such objects of complex systems that can act in parallel.

In this context there appeared the idea to study some of the aspects related to systems and technologies for distributed processing of data and which contribute to the formation of skills of the use the advanced computing technologies for distributed processing of data in solving various problems (in particular, for example, related to parallel processing of images).

2 Distributed systems by means of workflow Petri nets

Debugging of the parallel interactions [1] is a difficult and complicated process. For a better understanding of these processes Petri nets formalism is used, which is an environment of interaction of the processes (including the parallel one) and in which they can control the execution of successions and allow one or another operation to be held.

Modeling of distributed systems using workflow Petri nets [2] is performed at component level (each component can be a single system, but its behavior can be described independently of other components, except for interactions which are well defined with other components). Components have a specific state (state represents abstraction of information which is necessary to describe further actions) which determines what actions can occur in system, what states precede these actions and the state of the system after the occurrence of actions. Actions of the components of the system are characterized by combining or parallelization of them. Actions of the one component of the system can occur at the same time with actions of the other components of the system.

Through the simulation of model of the states the description of the dynamic behavior of the system is obtained. Also after simulation we obtain system properties that satisfy or not the established rigors and which can lead to a better system. This process continues until the desired outcome is obtained. Workflow Petri nets [3], as it was defined, serve for modeling the distributed systems with multiple processes that are executed concurrently in time. Next section will illustrate this.

3 Parallelism

Parallelism (competition) can be introduced in several ways. We consider the case of two concurrent processes. Each process can be represented by a workflow Petri net. Therefore, the composed workflow Petri net, which is simply the union of workflow Petri nets for each of the two processes, may represent concurrent execution of the two processes. Initial marking of compound workflow Petri net has two tokens, one in each source place representing the beginning of each process. This fact introduces parallelism that can not be represented by a logical scheme and for which the representation as a workflow Petri net is a very useful solution.

Another approach is to consider how parallelism can be introduced in a normal process into computer system. It is considered branching operations (FORK) and union (JOIN) initially proposed by Dennis and Van Horn [1966]. A branching operation performed with the precondition p_i determines current continuation of postcondition p_{i+1} and starts a new process execution at location p_j. A union operation will recombine two processes in one (or, equivalently, destroy one of the two processes and will allow the other to be performed). These operations can be modeled by a Petri net, as it is shown in Figure 1 a). If we consider parallel execution of k processes, then we obtain the combination of the two operations defined above – branching and union, as in Figure 1 b). Parallelism is useful in solving a problem if concurrent processes can cooperate in solving the problem. Such cooperation involves common processes information and resources [4]. Shared access to information and resources must be controlled in order to ensure correct functioning of the system. To run applications with parallel technologies at the Institute of Mathematics and Computer Science the 48 core IMI-RENAM cluster is used. On virtual machines there are deployed two systems: Windows Compute Cluster 2003, allows running Sequential, Parallel (OpenMP, MPI and Hybrid) applications and Grid-cluster.

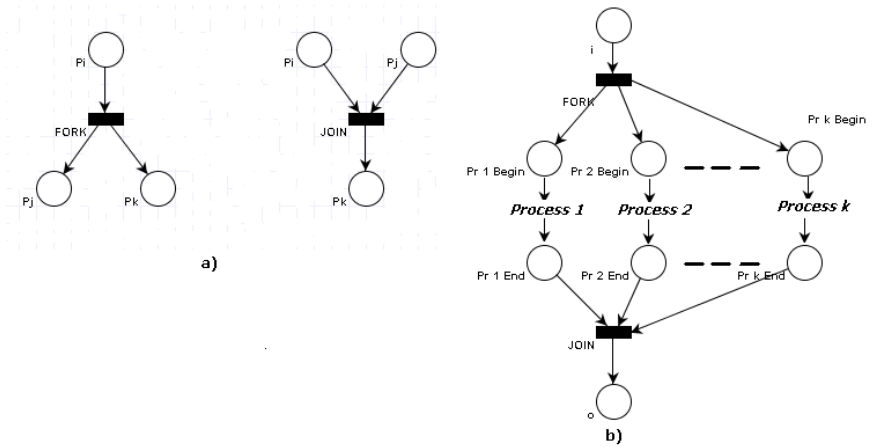


Figure 1. Parallel processes in workflow nets

4 Conclusion

In this article it was shown that workflow Petri nets is a convenient formal method for modeling information flow. Thus, it was shown how workflow Petri nets have been used for representation of parallel processes in order to better understand these processes. On the local cluster the applications can be tested and prepared for execution on more productive resources.

Acknowledgments. This paper was elaborated within the projects: 13.819.18.06A Systems and technologies of distributed processing of information and evaluating the effectiveness of their use; 12.839.08.01F Continuous structures and their applications.

References

[1] V. Ghergheli. *Theory and Practice of Parallel Computing*, -M, Internet–University, BINOM, Laboratory of knowledge, 2007.

[2] W. Aalst. *Interorganizational workflows: An approach based on message sequence charts and petri nets*, Systems Analysis – Modelling – Simulation, vol. 34, no. 3 (1999), pp. 335–367.

[3] W. Aalst, K. Hee. *Workflow Management – Models, Methods, and Systems* MIT Press, 2002.

[4] I. Titchiev. *Modeling of the Communication Protocol by Means of Colored Petri Nets. Case Study*, In Proceedings of "International Workshop on Intelligent Information Systems", September 13–14, Chisinau (2011), 203–206.

Inga Titchiev [1], Nicolae Iliuha[2], Mircea Petic[3]

Institute of Mathematics and Computer Science

[1]E-mail: inga.titchiev@gmail.com
[2]E-mail: wolf911@gmail.com
[3]E-mail: mirsha@math.md

Modeling and calculating of electrotechnical systems in Scilab

Topolsky Dmitry, Topolskaya Irina, Topolsky Nikita

Abstract

Modeling and calculation of a number of electrotechnical systems, including renewable energy, have been done in the free software of computer mathematics Scilab. The results confirm the validity of using the package Scilab for this purpose and demonstrate some of its advantages for use in electrotechnical education and scientific research.

Keywords: modeling in Scilab, free software, electrotechnical calculations, renewable energy.

1 Introduction

It is almost impossible to imagine modern science without the extensive use of mathematical modeling. This method of learning, design and engineering combines many of advantages of both theory and experiment. Work not with a real object (phenomenon, process), but with its model makes it possible relatively quickly and without significant costs to investigate its properties and behavior in all conceivable situations (the benefits of theory).

At the same time, computer experiments with models of objects can help (relying on advanced computational methods and technical capabilities of computers) to study objects in sufficient detail and in depth that is not possible in purely theoretical approaches (the benefits of the experiment). That is why the methodology of mathematical modeling rapidly develops and embraces new fields, including electrical engineering calculations of renewable energy.

2 The relevance of choice of modeling tool

The development of science depends largely on the enhancement of quality of education, which in turn is possible on the basis of new information technologies.

In recent years, Russia has undertaken a number of measures for the development and implementation of the free software in educational institutions. The main advantage of free software for the purpose of education is to have a legal basis to freely explore the documented source code of free software for computer and modify it, including those created on the basis of its own design. For the purpose of education it is necessary to provide teachers and students the opportunity to choose between free and proprietary software. In this context, the purpose of the study was testing the ability to offer the free software in relation to the mathematical modeling and calculation of electrotechnical systems.

Modeling is one of the most common methods of research that combines the capabilities of theory and practice. Simulation as well is particularly relevant for solving electrical problems [1, 2]. Development of methodology for modeling and analysis of electrical systems using free software is an important problem, which is of great practical importance and relevance.

The simulation results are largely dependent on the software. Mathematical packages are widely used in the modeling. In the market there is a number of computer mathematics systems. They all have great opportunities, advantages and disadvantages. However, during the involvement of students in research, the major drawback of these packages is to buy a license for their use, that often makes scientific research difficult and expensive [3].

One of the most popular open source mathematical package is Scilab. Scilab is designed for engineering and scientific calculations. In terms of features Scilab package is comparable to known mathematical package Mathcad, and at its interface is similar to the package Mathlab. There are versions of Scilab for different operating systems: Linux, Windows, MacOS.

The system itself Scilab, as Mathlab, is primarily intended for numerical computations and works with matrices. Main features of Scilab: works with elementary and special functions used in mathematics (approximation of elliptic integrals, Bessel functions), a set of tools for working with conventional and matrix polynomials; tools for creating and manipulating arrays (vectors, matrices, etc.) and hundreds of mathematical functions with the possibility to add the new ones, written in different languages (C, C ++, Fortran); 2D and 3D graphics, animation, differential and not differential optimization, stats and much more [4]. The package also includes Scicos – a tool for editing block diagrams and simulations (similar to the Simulink package Mathlab). There is the opportunity to work with Scilab software LabVIEW. In Scilab powerful programming language with support of objects is built in. Scilab allows you to create not just an ordinary program to automate the calculations, but also visual applications that run in the environment of Scilab. The main object in the environment of Scilab is a graphical window. The correct version of Scilab can always be downloaded from the official website of the program.

3 Results

During the testing of the package of Scilab in order to perform modeling and analysis of electrical systems a number of problems had been solved. Design solutions were obtained in Scilab for direct current (DC) electrical circuits with a DC power supply, complex DC electrical circuits, single-phase circuits with sinusoidal current, and also there were designed and built the mechanical characteristics of DC machines (see Figure 1). All analysis solutions obtained for Scilab were analyzed in comparison with similar solutions obtained for Mathcad. The results were identical.

Also, a number of models of renewable energy sources was developed: the model output with the new Chelyabinsk landfill of municipal solid waste (MSW) from the equation of the reaction kinetics of gas formation of the first order and the model of Tabasaran-Rettenberger;

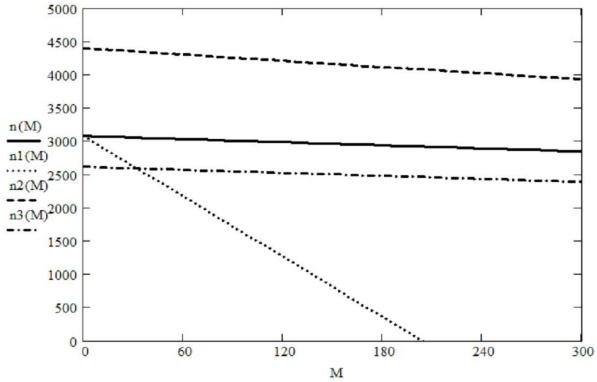


Figure 1. The results of calculation of the mechanical characteristics of the DC machine.

active and reactive hydraulic turbines; model of biogas plant for farming and others. These models are a set of sequential calculations. Figure 2 shows the modeling results of parameters of the new Chelyabinsk landfill of MSW.

One of the types of problems solved in Scilab, is the optimization problem. In this case, we consider the problem of linear programming, where it is required to find the values of variables $x1$, $x2$, $x3$, $x4$, when the objective function reaches its maximum value and the number of constraints is satisfied. There is function *linpro* to solve these problems in the Scilab. The results of solving optimization problems are presented in Figure 3.

In the modeling of the asynchronous electric drive the range of possible variations of the time constants was investigated in order to create an adaptive controller of the drive. One of modeling results is the surface, which characterizes the variation of electromagnetic time constants of electric drive (Figure 4).

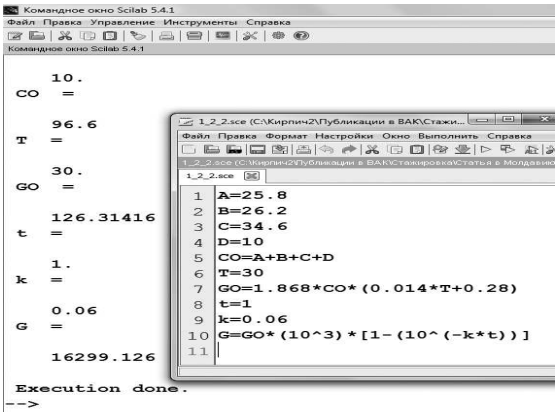


Figure 2. The modeling results of parameters of the new Chelyabinsk landfill of MSW.

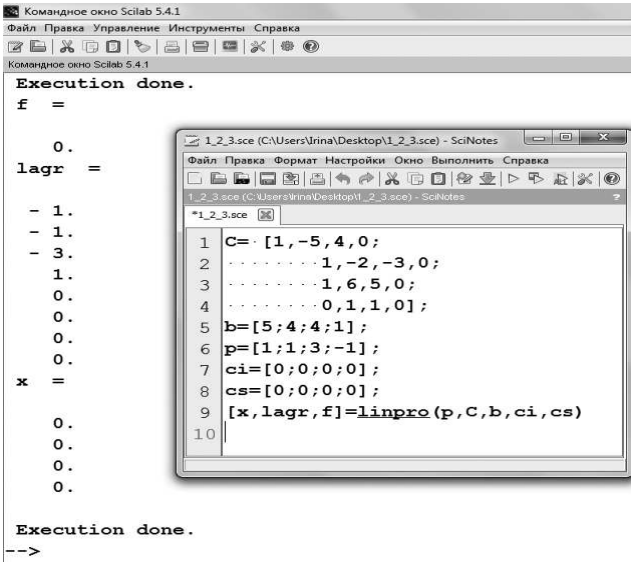


Figure 3. The results of solving optimization problems.

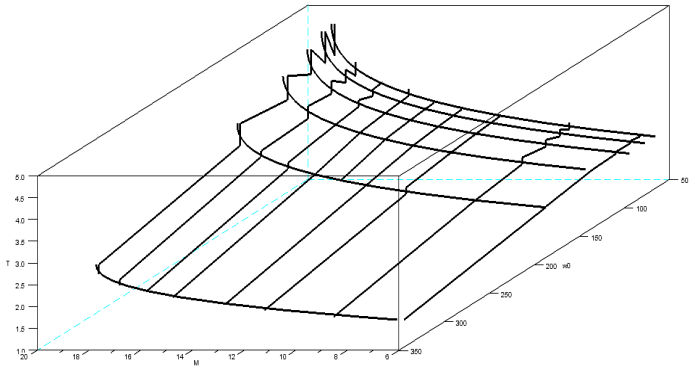


Figure 4. The variation of electromagnetic time constants of the asynchronous electric drive.

4 Conclusion

The study of the possibility of using free software package Scilab for mathematical modeling and calculation of electrical systems shows its suitability for this purpose. There is also a positive experience of its use in the study of the physical basis of obtaining information, methods and means of measurement, as well as solving optimization problems. In their capacity to solve most problems of this type it is not inferior to the existing paid mathematical packages. Accordingly, it is correct to speak about the benefits of Scilab, a freely distributed program for education and science. All this suggests the relevance of extending the scope of use of the package Scilab.

References

[1] D.V. Savinova. *Use of simulation modeling during study course of TOJe.* Bulletin of South Ural State University. Serija "Power Engineering", no 12(2007), pp. 58–60.

[2] N.V. Klinachev, N.V. Klinacheva. *Presentation of mathematical models of electrical machines using graphical language of Jigrein program.* Bulletin of South Ural State University. Serija "Power Engineering", vol. 10, no. 26 (2008), pp. 46–49.

[3] I.M. Kirpichnikova, D.V. Topolsky, I.G. Topolskaja. *Importance of use of free software during modeling of electrotechnical systems containing renewable energy.* Science actual problems: Materials for VII International theoretical and practical conference (25.01.2013), 2013, pp. 65–69.

[4] E.R. Alekseev, O.V. Chesnokova, E.A. Rudchenko. *Scilab: Solution of engineering and mathematical problems.* Moscow, 2008. http://www.altlinux.org/Books:Scilab.

Topolsky Dmitry[1], Topolskaya Irina[2], Topolsky Nikita[3]

[1]Electrical Engineering and Renewable Energy Department, South Ural State University, Chelyabinsk, Russian Federation
E-mail: topol69@mail.ru

[2]Electrical Engineering and Renewable Energy Department, South Ural State University, Chelyabinsk, Russian Federation
E-mail: tig@susu.ac.ru

[3]Student, South Ural State University, Chelyabinsk, Russian Federation
E-mail: ntopolsky@mail.ru

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Mathematical Theory of Pareto-Nash-Stackelberg Processes

Valeriu Ungureanu

Abstract

Control Processes of Pareto-Nash-Stackelberg type are defined. Mathematical models are presented and solving principles are identified. A direct-straightforward method for solving linear discrete-time optimal control problem is applied to solve control problem of a linear discrete-time system as a mixture of multi-criteria Stackelberg and Nash games. For simplicity, the exposure starts with the simplest case of linear discrete-time optimal control problem and, by sequential considering of more general cases, investigation finalizes with the highlighted Pareto-Nash-Stackelberg and set valued control problems. Different principles of solving are compared and their equivalence is proved.

Keywords: Linear discrete-time control problem, non-cooperative game, multi-criteria strategic game, Pareto-Nash-Stackelberg control.

1 Introduction

Optimal control theory which appeared due to Lev Pontryagin and Richard Bellman, as natural extension of calculus of variations, often doesn't satisfy all requirements and needs for modelling and solving problems of real dynamic systems and processes. A situation of this type occurs for problem of linear discrete-time system control by a decision process that evolves as Pareto-Nash-Stackelberg game with constraints – a mixture of hierarchical and simultaneous games. For such system, the notion of optimal control evolves naturally to the notion

of Pareto-Nash-Stackelberg type control and to the natural principle for solving the highlighted problem by applying a concept of Pareto-Nash-Stackelberg equilibrium with a direct-straightforward principle for solving [1].

The direct method and principle for solving linear discrete-time optimal control problem is extended to control problem of a linear system in discrete time as a mixture of multi-criteria Stackelberg and Nash games [1]. The exposure starts with the simplest case of linear discrete-time optimal control problem and, by sequential considering of more general cases, finalizes with the Pareto-Nash-Stackelberg and set valued control problems. The maximum principle of Pontryagin is formulated and proved for all the considered problems. Its equivalence with the direct-straightforward principle for solving is established.

2 Linear discrete-time Pareto-Nash-Stackelberg control problem

The conference presentation begins with the problem of optimal control and its optimal principles, and by considering different types of control and principles; it finalizes with the control of Pareto-Nash-Stackelberg type with T stages and $\nu_1 + \cdots + \nu_T$ players, where $\nu_1, \ldots, \nu_T$ are the correspondent numbers of players on stages $1, \ldots, T$. Every player is identified by two numbers: τ – stage on which player selects his strategy and π – player number at stage τ. In such game, at each stage τ the players $1, 2, \ldots, \nu_\tau$ play a Pareto-Nash game by selecting simultaneously their strategies according to their criteria ($k_{\tau 1}, k_{\tau 2}, \ldots, k_{\tau \nu_\tau}$ are the numbers of criteria of respective players) and by communicating his and all precedent selected strategies to the following $\tau + 1$ stage players. After all stage strategy selections, all the players compute their gains on the resulting profile. Such type of control is named Pareto-Nash-Stackelberg control, and the corresponding problem – linear discrete-time Pareto-Nash-Stackelberg control problem.

The decision control process may be modelled as:

$$
\begin{aligned}
&f_{\tau\pi}(x, u^{\tau\pi}||u^{-\tau\pi}) = \sum_{t=1}^{T}\left(c^{\tau\pi t}x^t + \sum_{\mu=1}^{\nu_t} b^{\tau\pi t\mu}u^{t\mu}\right) \underset{u^{\tau\pi}}{\longrightarrow} \text{ef max},\\
&\tau = 1,\dots,T, \pi = 1,\dots,\nu_\tau,\\
&x^t = A^{t-1}x^{t-1} + \sum_{\pi=1}^{\nu_t} B^{t\pi}u^{t\pi}, t = 1,...,T,\\
&D^{t\pi}u^{t\pi} \le d^{t\pi}, t = 1,...,T, \pi = 1,\dots,\nu_t,
\end{aligned}
\tag{1}
$$

where $x^0, x^t \in R^n$, $c^{\tau\pi t\mu} \in R^{k_{tp}\times n}$, $u^{\tau\pi} \in R^m$, $b^{\tau\pi t\mu} \in R^{k_{tp}\times n}$, $A^{t-1} \in R^{n\times n}$, $B^{\tau\pi} \in R^{n\times m}$, $d^{\tau\pi} \in R^k$, $D^{\tau\pi} \in R^{k\times n}$, $t, \tau = 1,\dots,T, \pi = 1,\dots,\nu_\tau, \mu = 1,\dots,\nu_t$.

By performing direct transformation, (1) is reduced to a sequence of multi-criteria linear programming problems

$$
\begin{aligned}
&f(u^{\tau\pi}||u^{-\tau\pi}) =\\
&= \left(c^{\tau\pi\tau}B^{\tau\pi} + c^{\tau\pi\tau+1}A^{\tau}B^{\tau\pi} + c^{\tau\pi\tau+2}A^{\tau+1}A^{\tau}B^{\tau\pi} + \dots +\right.\\
&\left. + c^{\tau\pi T}A^{T-1}A^{T-2}\dots A^{\tau}B^{\tau\pi} + b^{\tau\pi\tau\pi}\right)u^{\tau\pi} \underset{u^{\tau\pi}}{\to} \text{ef max},\\
&D^{\tau\pi}u^{\tau\pi} \le d^{\tau\pi},
\end{aligned}
\tag{2}
$$

$\tau = 1,\dots,T, \pi = 1,\dots,\nu_\tau$.

Equivalence of (1) and (2) proves the following Theorem 1.

Theorem 1. *Let (1) be solvable. The sequence* $\bar{u}^{11}, \bar{u}^{12}, \dots, \bar{u}^{T\nu_T}$ *forms a Pareto-Nash-Stackelberg equilibrium control in (1) if and only if* $\bar{u}^{\tau\pi}$ *is an efficient solution of multi-criteria linear programming problem (2), for* $\tau = 1,\dots,T, \pi = 1,\dots,\nu_\tau$.

Pontryagin maximum principle may be generalized for (1). By considering recurrent relations

$$
\begin{aligned}
p^{\tau\pi T} &= c^{\tau\pi T},\\
p^{\tau\pi t} &= p^{\tau\pi t+1}A^t + c^{\tau\pi t}, \quad t = T-1,...,1,
\end{aligned}
\tag{3}
$$

where $\tau = 1,\dots,T, \pi = 1,\dots,\nu_\tau$, Hamiltonian vector-functions are defined as

$$
H_{\tau\pi t}\left(u^{\tau\pi}\right) = \left\langle p^{\tau\pi t}B^{\tau\pi} + b^{\tau\pi\tau\pi}, u^{\tau\pi}\right\rangle, t = T,\dots,1,
$$

where $\tau = 1, \ldots, T, \pi = 1, \ldots, \nu_\tau$ and $p^{\tau\pi t}, t = T, \ldots, 1, \tau = 1, \ldots, T, \pi = 1, \ldots, \nu_\tau$.

Theorem 2. *Let (1) be solvable. The sequence* $\bar{u}^{11}, \bar{u}^{12}, \ldots, \bar{u}^{T\nu_T}$ *forms a Pareto-Nash-Stackelberg equilibrium control if and only if*

$$\bar{u}^{\tau\pi} \in \underset{u^{\tau\pi}: D^{\tau\pi} u^{\tau\pi} \leq d^{\tau\pi}}{\operatorname{Arg\,ef\,max}} H_{\tau\pi t}\left(u^{\tau\pi}\right),$$

for $t = T, \ldots, 1, \tau = 1, \ldots, T, \pi = 1, \ldots, \nu_\tau$.

Theorems are equivalent.

3 Concluding remarks

There are different types of processes control: optimal control, Stackelberg control, Pareto-Stackelberg control, Nash-Stackelberg control, Pareto-Nash-Stackelberg control, etc.

The direct-straightforward, dual and classical principles (Pontryagin and Bellman) may be applied for determining the desired control of dynamic processes. These principles are the bases for pseudo-polynomial methods, which are exposed as a consequence of theorems for linear discrete-time Pareto-Nash-Stackelberg control problems.

The direct-straightforward principle is applied for solving the problem of determining the optimal control of set-valued linear discrete-time processes. Pseudo-polynomial method of solving is constructed.

The results obtained for different types of set-valued control will be exposed in a future paper.

References

[1] V. Ungureanu. *Linear discrete-time Pareto-Nash-Stackelberg control problem and principles for its solving,* CSJM, v.21, n.1 (61), 2013, pp. 65–85.

Valeriu Ungureanu

State University of Moldova
Email: v.a.ungureanu@gmail.com

System Questions of Automation Control of Processes of Rendering Electronic Services to The Population on The Basis of Administrative Monitoring Data

Vadim Nikolaevich Volkov, Aleksey Ivanovich Frolov

Abstract

System decisions in the field of automation control of rendering electronic services to the population on the basis of data of administrative monitoring are considered in this article. The purposes and types of control are formulated, the contours of control in the organizational and technical environment of existence of electronic services are described, the role of system of administrative monitoring in the allocated contours is shown.

Keywords: electronic service; monitoring; control; automated system; organizational and technical environment.

One of the most actual directions of development of many countries now is creation of the information society, one of basic principles of which is wide circulation and accessibility of electronic services to the population, both state and non-state (banking, insurance, communications, etc.).

At the present stage of a services sector (both state, and non-state) development there are some problems [1] the organization-and-technical reasons of which are considered in [2]. One of them is a gap in a circuit of automated control of process of rendering services at a stage (section) of collection and processing of information on control object status. Also in [2] the system and technological tasks of administrative monitoring organization of processes of rendering the services

are formulated. It will allow creating necessary tools for unification of services control. The role of the monitoring system in the circuit of control of rendering the electronic services to the population (ESP) is considered in this article.

The purpose of monitoring is information and analytical support of rendering electronic services control. On the basis of the analysis of the organizational and technical environment (OTE) models and life cycle of rendering ETS [3] it is possible to select two levels (or two circuits) of control:

1. Operational management (or control of functioning) – the control exercised during rendering specific service and maintenance having the purpose to support the rendering this service on each of its stages. Controlling influence in this case is directed on the executive mechanisms of the organizational and technical environment.

2. Project management (or control of changes) – the control exercised during the organization of rendering services and having the purpose to correct the rules of rendering service to the demands of OTE.

Let us select basic elements in a circuit of control in automated systems: the control object (CO), the measuring mechanism (MM), the system making decisions (SMD) and the executive mechanism (EM). The description of these elements and communications is provided in Table 1 for both considered circuits. The external relations at this level of abstraction aren't considered here.

Based on the analysis of elements of circuits and communications it is possible to conclude that they will form difficult downlink structure. On the one hand, the hierarchy traditional for difficult control systems takes place: the top level – the circuit of project management, the lower level is provided by circuits of operational management of specific implementations of service. On the other hand circuits of the upper and lower levels have common structural elements and information flows.

Graphic representation of the considered circuits is given in Figure 1. The set of circuits of operational management is shown in it in a simple way.

Table 1.

Circuits element	Circuit of operative management	Circuit of project management
CO	Processes of rendering services by changing the implementations of rules of their rendering	Control is exercised by changing rules directly
MM	System of administrative monitoring	
SMD	The person making the decision	
EM	Executive bodies which are responsible for organization the processes of execution of operations of the rules	Executive bodies which are responsible for regulation and technical supply of processes of rendering services

It is especially necessary to mark an information flow (a dash-dotted arrow), providing adaptation of system of administrative monitoring to changeable rules of rendering services. Setup of model of data storage for each service happens automatically based on the rules of its rendering.

Implementation of control of rendering ESP leads to formation of hierarchical management system with circuits of the upper and lower levels. The feature of this decision caused by specifics of a lifecycle model of ESP and use of the uniform measuring mechanism is the common elements and information flows in these circuits.

On this basis further formalization of processes of information exchange, data storage and control in considered OTE of EUN is possible.

Researches were conducted within a grant of the President of the Russian Federation MK-3750.2012.9.

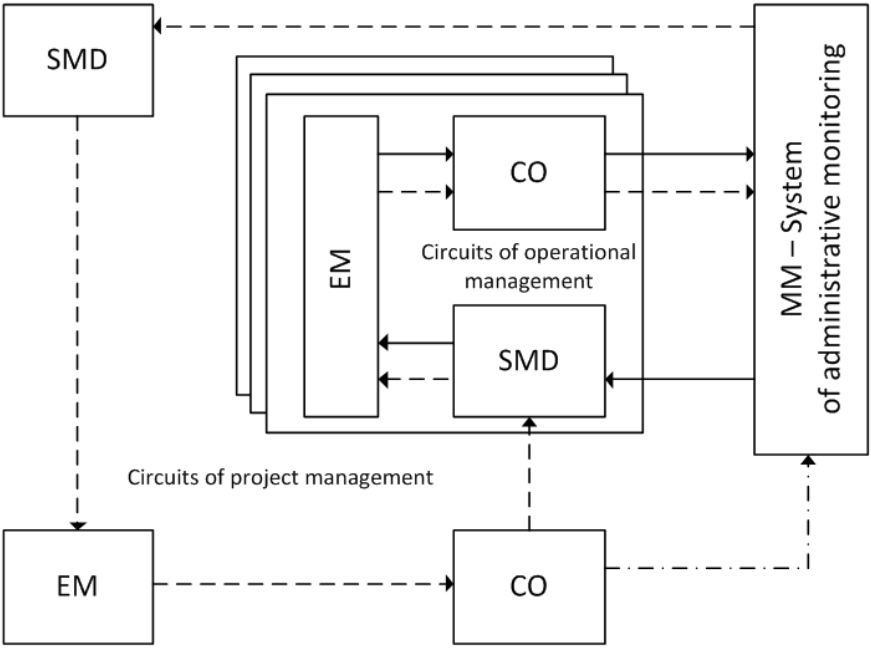


Figure 1. The structure of the circuits of operative (solid line) and project (dash line) management.

References

[1] O.A. Ivashhuk, O.P. Arhipov, O.A. Savina et al. *The market of electronic services to the population in Russia: problems and prospects.* Information resources of Russia., vol. 4 (122) (2011), pp. 2–6.

[2] A.I. Frolov. *The system and technological problems of the organization of the electronic services delivery processes monitoring.* Messenger of computer and information technologies., vol. 2 (2013), pp. 32–36.

[3] A.I. Frolov. *The organization of the electronic services delivery*

processes control on the basis of the administrative monitoring data. Belgorod State University Scientific Bulletin, series "History, Political science, Economics, Information technologies.", vol. 13 (132) (2012), pp. 199–204.

Vadim Nikolaevich Volkov[1], Aleksey Ivanovich Frolov[2]

[1]Affiliation/Institution State University – ESPC, Russia, Oryol
Email: vadimvolkov@list.ru

[2]Affiliation/Institution State University – ESPC, Russia, Oryol
Email: aifrolov@ostu.ru

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Multicriteria Allocation of Limited Resources

Albert Voronin

Abstract: The problem of distribution of the given global resource of the system under the constraints, imposed on individual resources is considered. It is shown, that the problem lies in constructing an adequate objective function for optimization of the resources distribution under their limitations. For solving the considered problem, the multicriteria optimization approach is undertaken with the nonlinear trade-off scheme.

Keywords: distribution of global resource, multicriteria optimization, nonlinear trade-off scheme, constructing an adequate objective function for optimization, decision making method.

1 Problem Description

The problem of allocating of limited resources is a main problem of economics. It is believed that the proper distribution and redistribution of resources – this is just the economics. Similar problems arise in other subject areas. The art is to allocate properly limited resources, depending on the circumstances.

One of such circumstances is usually resources limitation. The most prevalent is the case of upper limitation of a total (global) system resource to be distributed among the individual objects. In practical cases, constraints are imposed not only on the global resource, but also on the individual resources, given to individual objects. The constraints may be imposed both from below and from above. It is easy to see, that for all the individual resources the sum of the constraints from below is a lower bound for the global resource, and the sum of the constraints from above restricts the global resource from above.

The problem lies in constructing an adequate objective function to optimize resource distribution under the condition of their limitation.

2 Formalization of the Problem

Since the considered problem is urgent for different domains, we shall present the problem formulation in a general form.

The global resource R is given, which is to be allocated, and $n \geq 2$ elements (objects) of the system, every of which is provided with the individual resource r_i, their set forming the vector $r = \{r_i\}_{i=1}^{n}$.

The formula for determining the domain of this vector has the form

$$r \in X_r^{\circ} = \{r \mid 0 \leq r_i \leq R, i \in [1,n]\}. \quad (1)$$

At the same time, the condition

$$\sum_{i=1}^{n} r_i = R \cdot \quad (2)$$

holds true.

For every object it is known (or determined by the peer review method), the maximum permissible value of the resource $B_{i\min}$, below which the given object can not function. Thus, the given system is constrained from below

$$r_i \geq B_{i\min}, \sum_{i=1}^{n} B_{i\min} \leq R, i \in [1,n] \cdot \quad (3)$$

At the same time, for every object the quantity $B_{i\max}$ is known, which can not or should not be exceeded by the object resource. The system of constraints from above has the form

$$r_i \leq B_{i\max}, \sum_{i=1}^{n} B_{i\max} \geq R, i \in [1,n] \cdot \quad (4)$$

From (3) and (4) it follows that

$$B_{i\max} \geq r_i \geq B_{i\min}, i \in [1,n] \quad (5)$$

$$\sum_{i=1}^{n} B_{i\max} \geq R \geq \sum_{i=1}^{n} B_{i\min} \cdot \quad (6)$$

In view of (5), expression (1) is transformed to the form

$$r \in X_r = \{r \mid B_{i\max} \geq r_i \geq B_{i\min}, i \in [1,n]\}. \quad (7)$$

Let us consider the polar (degenerate) cases of inequality (6). If $R = \sum_{i=1}^{n} B_{i\min}$, then the considered problem is reduced to such distribution of the global resource, for which every object obtains its minimum allowable individual resource: $r_i^* = B_{i\min}, i \in [1,n]$.

If the global resource can fully satisfy the needs of the objects, i.e., $R = \sum_{i=1}^{n} B_{i\max}$, then the problem is solved as $r_i^* = B_{i\max}, i \in [1,n]$.

Thus, in the polar cases of inequality (6) the considered problem has trivial solutions. And only if the expression (6) becomes a strict inequality

$$\sum_{i=1}^{n} B_{i\max} > R > \sum_{i=1}^{n} B_{i\min}, \tag{8}$$

the problem of optimizing distribution of limited resources gets the sense.

The problem is formulated: under the condition (8) to define such individual resources $r^* \in X_r$, for which requirement (2) is fulfilled and some objective function $Y(r)$, the type of which should be selected and justified, takes the extreme value.

3 Method of Solution

In the problem of optimizing the distribution of limited resources the limit $r_i \le B_{i\max}, i \in [1,n]$ from above, is considered as a simple optimization constraint, the approaching to which does not threaten the system very much. Quite a different meaning has the limit $r_i \ge B_{i\min}, i \in [1,n]$ from below. The resource approaching to this limitation threatens the very possibility of the appropriate object functioning. One can say, that the limitation from below is "criteria-forming" in the sense that the objective function must increase the difference between the individual resource and its limit from below.

Therefore, the expression of the desired objective function should: 1) include constraints from below in the explicit form, 2) penalize the system for the partial resources approaching to these constraints, 3) be

differentiable by its arguments. The simplest objective function, satisfying these requirements, is

$$Y(r) = \sum_{i=1}^{n} B_{i\min}(r_i - B_{i\min})^{-1}. \quad (9)$$

The analysis of formula (9) shows that this is nothing but an expression of the scalar convolution of the maximized individual criteria $r_i, i \in [1,n]$, by the nonlinear trade-off scheme (NTS) in the problem of multicriteria optimization]. Indeed, in the considered problem the resources $r_i, i \in [1,n]$, have a dual nature. On the one hand, they can be considered as independent variables, the arguments of optimization of the objective function $Y(r)$. On the other hand, for each of the objects it is logic the desire to maximize their individual resource, to go away as far as possible from the dangerous limit $B_{i\min}$ to improve the efficiency of its operation. From this point of view, the resources $r_i \geq B_{i\min}, i \in [1,n]$ can be regarded as individual quality criteria of operation of the corresponding objects. These criteria are subject to maximization, they are limited from below, nonnegative and contradictory (the increase of one resource is possible only at the expense of reducing the other).

On the ground of the told above, the problem of vector optimization of limited resources, taking into account the isoperimetric constraint for arguments, becomes

$$r^* = \arg\min_{r \in X_r} Y(r) = \arg\min_{r \in X_r} \sum_{i=1}^{n} B_{i\min}(r_i - B_{i\min})^{-1}, \sum_{i=1}^{n} r_i = R. \quad (10)$$

Problem (10) can be solved both analytically, using the Lagrange method of multipliers, and by numerical methods, if analytical solution is difficult.

Albert Voronin

National Aviation University
E-mail: alnv@voliacable.com

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Mental Disorder Diagnostic System Based on Logical-Combinatorial Methods of Pattern Recognition

Anna Yankovskaya, Sergei Kitler

Abstract

The authors describe mental disorder diagnostic system based on logical-combinatorial methods of pattern recognition. The system is designed to diagnostic and prevention of depression. The mathematical apparatus for creation of the proposed system is suggested. The description of the system is given.

Keywords: intelligent system, logical-combinatorial methods of pattern recognition, diagnostic tests, intelligent instrumental software IMSLOG, depression.

1 Introduction

Creation of intelligent systems (ISs) for various semistructured areas, such as medicine, psychology, geology, etc. and development of algorithms underlying this ISs is very relevant [1, 2]. Mathematical apparatus of a number of ISs for above-mentioned problem areas is based on logical-combinatorial methods of test pattern recognition [2-4]. Currently, investigation in practical public health, viz. revealing mental and behavioral disorders is very important. However, the problem of ISs creation for revealing regularities of various kinds, high quality and timely diagnostic and prevention of these disorders is still open. Unlike created intelligent systems [5] for revealing mental and behavioral disorders in an inspected person which are based on a small number

of scales and / or questionnaires, in our proposed mental disorders diagnostic system diagnostic criteria of the international classification of diseases, tenth revision (ICD-10) [6] and 8 clinical-psychological scales and questionnaires are used.

In Laboratory of Intelligent Systems at Tomsk State University of Architecture and Building by chief A.E. Yankovskaya have been developed ISs for revealing regularities and diagnostic and organizational-management decision-making, for example IS for revealing social-psychological factors in communicative stress conditions in learning process [7]; IS DIOS [8] designed to express-diagnostic and intervention (correction) of organizational stress and IS DIAPROD [9] designed to express-diagnostic and prevention of depression.

Unlike IS express-diagnostic DIAPROD [9] using threshold and fuzzy logic for decision-making, in proposed IS DIAPROD-LOG decision-making is fulfilled based on logical-combinatorial methods of test pattern recognition.

2 Basis of mathematical apparatus of creation of intelligent system DIAPROD-LOG

The mathematical apparatus of IS DIAPROD-LOG is based on logical-combinatorial methods of test pattern recognition [3, 4]. For the data and knowledge representation in the IS DIAPROD-LOG a matrix model [4] is used.

The model includes an integer description matrix (**Q**) that describes objects in the space of characteristic features and an integer distinction matrix (**R**) that partitions objects into equivalence classes for each classification mechanism.

We mean under the pattern a subset of objects knowledge base with matching values classification features.

A diagnostic test (DT) [4] is a set of features that distinguishes any pair of objects that belongs to different patterns.

A diagnostic test is called "irredundant" (dead-end [3]) if it includes an irredundant amount of features.

An irredundant unconditional diagnostic test (IUDT) is characterized by simultaneously presentation of all features of the object under investigation included in test, while decision-making.

Definition of regularities in data and knowledge is given in [4]. The regularities can include constant (taking the same value for all patterns), stable (constant inside a pattern, but non-constant), non-informative (not distinguishing any pair of objects), alternative (in the sense of their inclusion in DT), dependent (in the sense of the inclusion of subsets of distinguishable pairs of objects), unessential (not included in any irredundant DT), obligatory (included in all irredundant DT), pseudo-obligatory (which are not obligatory, but included in all IUDT involved in decision-making) features and signal features, as well as all minimal and all (or part, for a large feature space) irredundant distinguishing subsets of features that are essentially minimal and irredundant DTs, respectively and tolerant to measurement (entry) errors IUDT [10]. The weight coefficients of characteristic features calculated by different algorithms are also included in regularities [4].

There is no doubt that wider range of regularities considered provide a higher degree of accuracy while diagnostic decision-making.

We use a procedure for constructing the irredundant implication matrix ($\mathbf{U}'$) [4, 10] for revealing various kinds of regularities at construction of IUDTs.

The matrix $\mathbf{U}'$ is an integer. The matrix $\mathbf{U}'$ defines distinguishability objects from different patterns (classes for each mechanism classification).

The regularities of various kinds are revealed on the matrix $\mathbf{U}'$ in order to reduce the feature space, determine the most important features. Also, all irredundant column coverings of the matrix $\mathbf{U}'$ [4, 10], defining essentially all BBDT are determined with the use of logical-combinatorial algorithms. Then the choice of optimal subset of features is fulfilled. On base of this subset ultimately the final decision-making is fulfilled.

3 Description of intelligent system DIAPROD-LOG

The developed intelligent system DIAPROD-LOG is distributed. The first part of the system is designed to data collection and data and knowledge storage. This part is implemented as a web-application with the use of C#. To data storage in the IS DIAPROD-LOG was chosen relational database management system MySQL, since it has great flexibility, rich functionality and is free of charge.

An inspected person offered to be tested to 8 questionnaires and scales: questionnaire A. Beck including 21 features; Edinburgh postpartum depression scale including 10 features; multivariate Freiburg personality inventory (FPI-B) including 114 features; Tomsk questionnaire rigidity of G.V. Zalewski (TQRZ) including 150 features; questionnaire of relationship of pregnant by I.V. Dobryakov including 9 features; questionnaire about ways of coping R. Lazarus including 50 features; questionnaire symptom levels including 90 features; questionnaire of determining the stress and social adaptation of Holmes and Rage including 43 features.

Unfortunately, the scope of the paper does not permit to include references to the questionnaires and scales. Since the inspected person confirms the correctness of entered answers, the test results are stored in the data and knowledge base.

The second part of a system designed to create matrices $\mathbf{Q}$ and $\mathbf{R}$, to construct matrix $\mathbf{U}'$, revealing different kinds of regularities, to construct diagnostic tests, decision-making and justification of decisions, is implemented as a template of intelligent instrumental software (IIS) IMSLOG [11] including dynamically plug-ins.

The IIS IMSLOG has module designed to data and knowledge base operation. Input data of module is structure of data and knowledge base and objects of knowledge base. Future selection of the necessary features for including in matrices $\mathbf{Q}$ and $\mathbf{R}$ is produced. The IS DIAPROD-LOG has characteristic features space including 28 features of the questionnaire A. Beck; the test result on Edinburgh post-

partum depression scale; the test result on FPI-B; the test result on TQRZ; the test result on questionnaire of relationship of pregnant by I.V. Dobryakov; the test result on questionnaire about ways of coping R. Lazarus; the test result on questionnaire symptom levels; the test result on questionnaire of determining the stress and social adaptation of Holmes and Rage. The test results (characteristic features values) are stored in the data and knowledge base as well as classifications features values filled based on ICD-10 and highly qualified experts knowledge in considered a problem area.

Also the IIS IMSLOG has a module designed to realizing construction matrix $\mathbf{U}'$ by matrices $\mathbf{Q}$ and $\mathbf{R}$ with simultaneously calculating weight coefficients of characteristic features, similar to the algorithm described in [10]. In this case, this module is not implemented condition of tolerant to a preassigned number of measurement (entry) errors of characteristic feature values of the objects under investigation described in [10]. Then in the next module the above mentioned regularities are revealed on the basis of the matrix $\mathbf{U}'$. Next the module is the construction of all irredundant column covering of matrix $\mathbf{U}'$, defining in fact all IUDTs.

The last module fulfils a final decision-making on the diagnostic and prevention of depression in an inspected person based on voting procedure [4] on the set of tests and approaches.

4 Conclusion

The basis of the mathematical apparatus of creating intelligent system DIAPROD-LOG based on the logical-combinatorial methods of test pattern recognition, revealing various kinds of regularities, decision-making and justification decisions are suggested. The description of this system is given.

Application of the developed IS DIAPROD-LOG will allow time to diagnose depression, prevention decision-making, as well as forming the diagnostic and prevention results.

Further investigations are devoted to the creation of complex intel-

ligent systems for diagnosis and prevention of depression based on IS DIAPROD-LOG and IS DIAPROD.

Acknowledgments. This research is funded by grant from the Russian Foundation for Basic Research (projects no. 13–07–00373a and no. 12–07–31109–mol_a) and by grant from the Russian Humanitarian Scientific Foundation (project no. 13–06–00709a).

Professor MD Kornetov A.N. is greatly appreciated for providing questionnaires and scales for the diagnosis of depression.

References

[1] B.A. Kobrinskiy. *A retrospective analysis of the medical expert systems* Novosti iskustvennogo intellekta, no. 2 (2005), pp. 6–17 [in Russian].

[2] A.E. Yankovskaya. *Test recognition medical expert systems with cognitive graphics elements* Komp'yuternaya khronika, no 8/9 (1994), pp. 61–83 [in Russian].

[3] Yu.I. Juravlev, I.B. Gurevich. *Pattern recognition and image analysis* Artificial intelligence in 3 books, book no. 2: Models and methods: Handbook/ Edit by D.A. Pospelov (1990), Moscow: Radio i Svyaz, pp. 149–191 [in Russian].

[4] A.E. Yankovskaya. *Logical tests and cognitive graphic tools* LAP LAMBERT Academic Publishing (2011), 92 p. [in Russian].

[5] V.E. Ekong, U.G. Inyang and E.A. Onibere. *Intelligent Decision Support System for Depression Diagnosis Based on Neuro-fuzzy-CBR Hybrid* Modern Applied Science, vol. 6, no. 7 (2012), pp. 79–88.

[6] *World Health Organization, 1992. The International Classification of Diseases, Tenth Revision (ICD-10). Clinical descriptions and diagnostic guidelines, Geneva, World Health Organization.*

[7] A.E. Yankovskaya, E.A.Rogdestvenskaya. *Revealing of social-psychological factors in conditions of communicative stress in the learning process with use intelligent system* The psychological universe of the formation of human noetic. Proceedings of International Symposium, Tomsk (1998), pp. 184–186 [in Russian].

[8] A.E. Yankovskaya, S.V. Kitler, A.V. Silaeva. *Intelligent system for diagnostics and intervention of organizational stress: its development and approbation* Otkritoe obrazovanie, vol. 91, no. 2 (2012), pp. 61–69 [in Russian].

[9] A.E. Yankovskaya, S.V. Kitler, R.V. Ametov. *Basis for creation of intelligent system for express diagnostics and prevention of depression* Proc. of congress on intelligent systems and information technology, vol. 2, M.: Fizmatlit (2011), pp. 265–272 [in Russian].

[10] A.E. Yankovskaya, S.V. Kitler. Parallel Algorithm for Constructing k-Valued Fault-Tolerant Diagnostic Tests in Intelligent Systems Pattern Recognition and Image Analysis, vol. 22, no 3 (2012), pp. 473–482.

[11] A.E. Yankovskaya, A.I.Gedike, R.V. Ametov, A.M. Bleikher. *IMSLOG-2002 Software Tool for Supporting Information Technologies of Test Pattern Recognition* Pattern recognition and image analysis, vol. 13, no. 4 (2003), pp. 650–657.

Anna Yankovskaya[1], Sergei Kitler[2]

[1]National Research Tomsk State University; Tomsk State University of Architecture and Building
Email: ayyankov@gmail.com

[2]Tomsk State University of Control Systems and Radioelectronics
Email: svkitler@gmail.com

Part 3

STUDENTS' PAPERS

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Prediction of medical conditions of the person on the basis of data on urbanosferas

Valeriy Abashin, Aleksander Pilipenko

Abstract

Authors discuss, the use of concepts urbanosfera and personal urbanosfera in the development of the theory of prediction of medical conditions of the person. Analysis of the methods and systems for medical prediction from the point of view of mathematical support is used. Methods of obtaining information about the urbanosferas and methods of urbanosferas comparison are considered.

Keywords: urbanosfera and personal urbanosfera, prediction of medical conditions of the person, methods, data from urbanosferas, automation.

1 Introduction

The current state of information technologies allow us to develop telemedicine technologies, and to go to the question of automating the collection of data about medical conditions of a human that exists in the urban area; forecasting these medical conditions. The use of concepts urbanosfera and personal urbanosfera in the development of the theory of prediction of medical conditions of the person enables us to make the process iterative and continuous. In this case, the further development of forecasting medical conditions technologies will enhance the level of medical care, and thus the quality of life in general.

Define two main medical conditions of a person: condition of norm (functional optimum) and pathology. Pathology is considered a result

of exhaustion and fracture of mechanisms of adaptation. All possible deviations from the norm are pathologies or pathological conditions and they form a set of pathologies (Figure 1).

In fact forecasting medical conditions is usually reduced to the problem of forecasting one of the most important at the moment state – functional ability of the person. An example is the definition of mental capacity. In this case, there have to be forecasted the number of "failures", i.e., those conditions that require control functions, but a person cannot carry them out because of fatigue, temporary loss of working capacity, etc [1].

To predict medical conditions there are often used technical and economic models [2], which are based on information theory, theory of pattern recognition, discrete differential methods, stochastic processes, separation in the feature space, mathematical logic, logic and discrete method. Among existing methods and systems for medical prediction

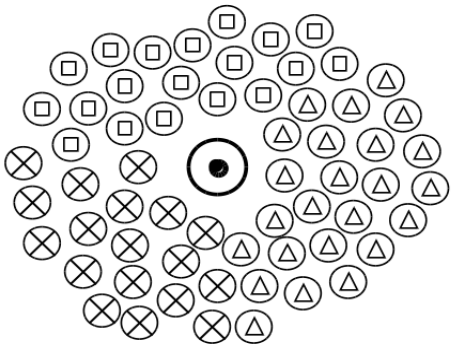


Figure 1. Pathologies of relatively optimal condition of a person. Point - condition of norm, cross - congenital defects (chromosomal, genic), square - specified by the influence of environmental hazards (injuries, burns), triangle - specified by genetic predisposition and impact of initiating agents of environment (diabetes, peptic ulcer, neoplastic disease).

from the point of view of used mathematical support the following can be distinguished:

1. probabilistic methods that are on the basis of signs define one of

the possible events with a certain probability;

2. methods of multivariate statistical analysis that can detect character and structure of relationships between components of the multivariate attribute;
3. sequential statistical analysis that represents the consistent procedure surveys, which achieves the selected level of probability of the diagnosis or prognosis;
4. logical-probabilistic, which uses the value of conditional probability of occurrence of signs typical of a disease and its priori probability.

2 Concept

The increase of the automation level of medical devices is closely connected with the general increase in the automation of the urban environment. The presence of successful automation projects and urban management and its individual elements, as well as the mass distribution of computing devices allows to predict a significant acceleration of the implementation of automation and control of all the processes of municipal services that will become the source to enhance the quality of prediction of medical conditions of the person living in urban territories. When one automates significant portions of the city, or combines multiple systems in one city, there is a problem of determining the boundaries of the future automated system, which makes it difficult to use of a system approach. To solve this problem, the use of a new term *urbanosfera* is proposed (from Lat. Urbanus – urban, Greek. $\Sigma\varphi\alpha$ $\iota\rho\alpha$ – closed surface). Initially under the term of urbanosfera a certain part or parts of urbanized environment was meant; or the environment of environment with the living conditions similar to the living conditions in the urbanized environment; or the environment of the environment resource-dependent on urbanized environments. The distinctive feature of urbanosfera is the presence of a common purpose of its use, operation. For example, the urban water system has a certain objective of functioning, it is controllable and therefore it is an urbanosfera.

However, the autonomous regional water system, as a part of a city-wide water system, also has a single purpose and controllable, then it is an urbanosfera.

An example of combining multiple systems of urban economy into one is the conception of the "smart home". It allows you to build up systems that integrate all the resources in one building, and to manage them. In this case, the resulting system does not cover all the processes taking place in the building and allows the "surface" to interact with elements of the building dependent on external factors and operating by complex laws.

Note that several urbanosferas may be disposed on the same territory, and may use the same resources, control elements and perhaps even be on the same physical properties. Different urbanosferas can have the same goals and the territory of the location, but use different tools to achieve goals or different ways of urbanosfera's elements interacting. Urbanosferas retain such property in urban areas, as extra-openness, which means that they are dependent on their entire environment, and do not contradict the principles of "incomplete intelligibility of the object", and "transfer of part of the elements of a complex object in the category of autonomous subjects". All urbanosferas have a set of common properties and can be classified. Each class of urbanosfera has a set of properties unique to it.

To represent the notion of urbanosferas in mathematical form, it is proposed to use the apparatus of the theory of fuzzy sets. Let's denote the fuzzy set of urbanosfera as $Y = \{(x, \mu_y(x)|x \in X)\}$, where X – the set of all possible elements of urbanosfera in the most general sense, $\mu_y(x)$ the membership function of x on the fuzzy set y. The function $\mu_y(x)$ can take on values in a linear set of accessories M. Usually the observed condition is $= [0, 1]$, which corresponds to a normal fuzzy set, but at this stage of formalization it is more important than the fact that the membership function shows the relationship element of X with probability of achieving the goal Y. The less the target is achievable without an element of x, the more it belongs to Y, that is, the function $\mu_y(x)$ is closer to 1. In addition, on the basis of X it is theoretically

possible to construct an infinite set of urbanosferas.

Initially executed formalization allows us to solve the problem of defining the boundaries of urbanosfera, because it does not require the inclusion of all elements of urbanosfera in the model, but only those that significantly increase the likelihood of achieving the goal Y. Applying fuzzy sets it is possible to perform comparisons of urbanosferas and to determine their properties in a mathematical sense, the implementation of the operations of intersection, product, union, addition and negation. It should also be taken into account that in practical modeling the theory of fuzzy sets is used alone, without the apparatus of the theory of probability and mathematical statistics, which allows you to change the volume of statistical tests on iterative mathematical model. The use of the concept of urbanosfera allows us to scale the problem of automation and management of urban areas, to highlight the most important elements and to investigate the reciprocal influence of individual parts of the urban economy.

Going back to the resources of urban planning, it should be noted that the first texts on planning of the city can be related with the works of Greek philosophers. One of the first models of the ideal city is described by Plato and based on philosophical arguments about the interaction between people. So, initially the center of the city unit was defined as an interpersonal interaction, which means that from the very beginning understanding of Urban Affairs development a man was central core of the city. This approach is used in our days. In accordance with the UN programs on human settlements, the modern international community is a central place to the effort of the international community towards the sustainable development and it gives the person and approves of and adopts the universal purposes of providing shelter for all men and making human settlements safer, healthier and more liveable, equitable, sustainable and productive habitat. It also indicated that the quality of housing and living conditions in human settlements continues to deteriorate for a large number of people and the matter is urgent on the agenda of the international community [3].

Applying the concept of urbanosfera to anthropocentric approach of

constructing the urban areas, habitat of a particular person is defined as a personal urbanosfera. By this concept, everything that touches people, living in the urban area and having in their life predominance of urban lifestyle, may be included. However, for solving a particular matter and building up a model of its solution, any sets of elements which interact with a particular person can be used. First of all, the authors consider it appropriate to study personal urbanosfera to determine the effect of specific elements of the urban area on the psychophysical and psycho-physiological states of a particular person. Thus, the tool should appear, allowing a person to show what he gets from the city and what he pays for it.

Considering the personal urbanosfera of a person on a concrete example of the car driver, you can select the personal urbanosfera, which includes his car. In this case, to the personal urbanosfera can be also referred objects encountered on the road, which may have a direct impact on a person, for example, the road surface and passing nearby cars, trees, houses and other elements do not have significant emotional or physical effects on him. It should be noted that if a driver is considered to pass by the tree planted by him in childhood, as an element associated with its past, this tree might have a significant emotional impact on him.

The difference of the concept of personal urbanosfera from the concept of life style of a person is the fact that in the study of personal urbanosfera is not investigated. The influence of human behavior or social relations of himself, and therefore personal urbanosfera does not restrict the free will of a man.

Personal urbanosfera, as a special case of urbanosfera, retains all its properties. It enables the construction of models and systems in order to equalize the standard of living of various types of settlements, since the mathematical formalization of the concept of personal urbanosfera provides an opportunity to use computable criterion for assessing personal urbanosferas of different people.

In the example with the car, the objectives of personal urbanosfera are expressed in a variety $P = (p_1, p_2, p_3, p_4, ...)$, where p_1 – preservation

of functional optimum of a human while driving, p_2 – achievement of destination, p_3 – minimizing of the time spent on the trip, p_4 – minimizing of the consumption of gasoline.

Let us describe the purpose p_1. The current functional status of a person S is a reflection of the brain in a certain initial state, the characteristics of the current needs and the input impact on the time interval of adaptive behavior. In a formalized manner $S = f(N, I, ..., S_0, T)$, where N – need, I – external to the functional system of input effect, S_0 – initial state of a living system, T – time [4]. Having designated functional optimum of a human S_{opt}, we obtain that the purpose $p_1 = S_{opt} - S \to 0$, i.e. target p_1 is achieved when the difference between the optimum and current state of a person are close to zero.

In the general case, to the meaningful characteristics of the human body we refer those of adjustable available methods which have essential ratio to the survival (pulse rate, body temperature, etc.), so that significant changes of one of them sooner or later lead to changes of the others [5]. It is a common knowledge that as a signal of the functional state of human there are used figures of galvanic skin response, speech activity, blood pressure, vascular tone, pupil diameter, etc. [6][7]. Based on the data obtained from these signals one can determine the current state of the person. The most valuable among the informative characteristics in the specific case will be indicators of the functions of the body which carry the greatest load, so the choice of the signal about the functional status of a person for each activity must be selected separately. In this regard, it seems appropriate to search for new methods to obtain informative signals about the human condition, as well as the creation of combined dynamic methods of tracking [8]. Thus, from personal urbanosfera, the basic data for forecasting purposes of medical conditions are values of functional states of a human in different time periods, including the current functional states.

The authors suggest that the study of personal urbanosfera will open up new opportunities to optimize the resourcing of the population of urban areas, the emergence of new areas of research to improve the quality of life of human beings in urbanosfera will disseminate elements

of urban life in the rural areas by including remote rural settlements in urbanosfera cities that ultimately help to stabilize the processes of urbanization and rurbanization. In addition, personal urbanosferas will be a source of data about a person, on the basis of which the definition of medical conditions of the person and their prediction can be got.

The part of the urbanized area included within the urbanosfera interacting with personal urbanosfera of a man, is also a source of information that can improve the accuracy of prediction of medical conditions of the person. The efficient use of urbanosfera data depends on the level of urban environment automation. The most important role the data about the ecological state of the city streets and districts, adherence to urban ecology have played.

Having enough information about the urban area and personal urbanosfera of a person, you can diagnose and predict the health of the human condition, from the emotional states and ending with organ-specific diseases.

Data about urban urbanosferas include an information about population housing quality, environmental performance areas, the amount of resources consumed. Most of the above information may be obtained by evaluating the biometrics of the population and using data collection device that are installed on public transport.

Information about personal urbanosfera, except the above-mentioned one, includes information about the person, his movements, contacts, foodstuffs etc. To collect this information the mobile phone can be used as a hardware and its optional modules as software. Using this device you can get the following data:

1. about coordinates of the object and movements;
2. about body temperature and its change;
3. about emotional and physical state in terms of heart rate and galvanic skin response;
4. about the type and intensity of work.

Getting real-time data from city urbanosfera generates extremely large data sets (Big data). For their processing it is necessary to perform a quick search algorithms optimization to solve the problem. The

main way of quick search optimization is an acceleration of selection of the reference values, with respect to which there is a search. It requires finding patterns in the data collected at the stage of data collection, and based on them it will be possible to optimize the search algorithm of the support member, and hence the overall performance improvement of search algorithms and data sorting. The greatest interest is caused by anomalous values reporting failure of urbanosfera subsystems, which are an indicator of the extreme state of urbanosfera or pathological medical condition of the person.

Hereafter, in the case of the linearization of the problem of forecasting medical conditions, finite element method will be mathematically justified [9] and will significantly optimize the processing of data about urbanosfera and operations with them.

The main differences between urbanosferas are the goals of their operation. The goal of any urbanosfera is a multi-criteria vector, which can be degenerated into a vector of one element. Thus, the comparison of urbanosferas consists in comparison and finding dependency of interaction of vectors purposes. Most accurately the interaction of vectors is described in the mathematical theory of a vector field.

3 Conclusion

The use of concepts of urbanosferas, personal urbanosfera allow gradual increasing the level of automation of urban areas within a framework of a single concept. Data from urbanosferas and personal urbanosfera, allow us collecting large amounts of data on the basis of which the systems of forecasting medical conditions of the person will be built up. In this case, the personal urbanosfera is the source of data on the functional status of a person, his movements, interactions with other city urbanosferas. City urbanosferas are the sources of data on environmental conditions of the urban area, infrastructure development, urban environment, compliance with the principles of urban ecology territory.

According to the authors, perspective mathematical method for

processing data of urbanosfera is the finite element method, and for urbanosferas comparison high efficiency will be provided by the mathematical theory of the vector field.

References

[1] A. Krylov. *The methodology of research on engineering psychology and psychology of labor.* LA: LSU, 1974.

[2] R.M. Bayevsky. *Forecasting states on the verge of normal and pathological states.* M.: "Medicine", 1979.

[3] *World Habitat Day*, 2013. http://www.un.org/en/events/habitatday/.

[4] B.F. Lomov. *The activity of the human operator in the "man-machine".* Academy of Sciences of the USSR. 1975. Number 1. pp.51–50.

[5] W.R. Ashby. *The design of the brain.* Moscow: Izd. Lit. 1962.

[6] V.I. Medvedev. *The functionality of the human condition in the labor force.* A.B. Leonova. Moscow: Moscow State University Press, 1981.

[7] A.N. Lukyanov, M.V. Frolov. *Status signals of the human operator.* Moscow: Nauka, 1969.

[8] V.G. Abashin. *The adaptive multi-biometric mathematical model based on fuzzy sets.* The Federal Service for Intellectual Property, Patents and Trademarks. Certificate of state registration of the computer. Number 2011616318 from 11 August 2011.

[9] G. Strang, J. Fix. *The theory of finite element method.* 1977.

Valeriy Abashin[1], Aleksander Pilipenko[2]

[1]State University Education-Science-Production Complex, Orel, Russian Federation. Email: valeriy@abashin.ru

[2]State University Education-Science-Production Complex, Orel, Russian Federation. Email: nigyll@gmail.com

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Medical image ontologies

Natalia Bruc

Abstract: This paper presents an overview of medical image ontologies, namely 2 basic approaches to the creation of such ontologies are illustrated: graphical and on the basis of verbal annotations.

Keywords: ontologies, medical image, annotation.

1 Introduction

Ontology formally represents knowledge as a set of concepts within a domain, and the relationships between pairs of concepts. It can be used to model a domain and support reasoning about concepts [1]. Ontology is a way of detailed, formal representation of some area knowledge in the form of conceptual scheme, consisting of class hierarchy, slots, facets, instances, and interrelations between them.

Ontologies are used in artificial intelligence, semantic networks and software engineering as a form of representation of knowledge about the real world or a part of it. They become the privileged and the almost inevitable tool for representation and use of knowledge and data. This applies to many domains, especially to health care. Medicine is one of the most extensive domains for application of ontologies. In this article the approaches of creation of medical image ontologies are illustrated on concrete examples.

2 Medical image ontologies

Ontologies are most widely used in medicine. A special place here is given to image ontologies.

There are two approaches for describing images using ontologies:

1. graphical - information is given in numerical values of pixels;
2. on the basis of verbal annotations - the information is given in natural language.

Image ontologies presented in numerical values of pixels are such ontologies as: the Image Representations Ontology (IROn)[2], Ontology of Imaging Anatomy and Observations [3].

Let us describe the aforementioned ontologies.

IROn is an ontology of image representations for medical image mining. It is based on previous work – the BioTop [4] (it consists of 700 logical axioms and is compatible with the ontologies DOLCE, BFO and the OBO relation ontology). This ontology is constructed to support content-based mining of medical images, that "is realized by extending the concept of spatial relations from the 2D image space to multidimensional feature spaces" [2].

Daniel L. Rubin and his colleagues developed an ontology of image annotation and markup. In this ontology entities and relations for semantics of medical image pixel content are represented [3]. The ontology consists of "the *anatomic structures* visualized in images, the *observations* made by radiologists about images (e.g."opacity" and "density"), the *spatial regions* that can be visualized in images, as well as other image metadata." [3]. It is created using Protege ontology editor, in OWL-DL ontology language.

At this stage, we are most interested in the image ontologies, presented in a natural language, so more detailed we consider this approach.

As examples of image ontologies, presented in a natural language, let us give a brief description of several approaches.

In [5] it is presented a prototype of a database accessible through the network, to manage the images obtained during the experiments in biomedical research laboratory, in studies of the factors controlling the development of cataracts.

On the basis of the ontology being developed, which was created to describe the experimental data and protocols used in the laboratory, image storage allows members of the laboratory to group the images by several attributes. The use of ontologies for developing these and other tools facilitates interconnection between tools, and ultimately, data exchange with other researchers.

The Experiment ontology [5], which is created in Protege, formalizes the types of data related to biomedical experiment related to the cataract. Among these types of data there is a representation of a digital slit-lamp image taken in a few experiments. For example, the images created during the experiments of investigation of the role of SPARC protein in cataract show the stage of cataract development in experimental animals.

In [6] it is described the BioImage Database. BioImage Database is an ontology driven image database for multidimensional images of biological samples. Ontology, which is the central part of the database, is written in OWL-DL and developed using Protege OWL plug-in. All user interactions with the database mediate with this ontology. Manual metadata input is simplified by the dynamic creation of subordinated user interfaces from a simple web form underlying the ontology. An improved search interface allows the reliable search for relevant images.

The closest to our goals is the approach described in [7]. Project DICOM Ontology (DO) is elaborated in the Stanford Center for Biomedical Informatics Research (BMIR). DICOM Ontology (DO) is an ontology that will be a single common reference information model for the medical imaging domain [8].

This ontology is created for medical images obtained in a widespread format DICOM. DICOM (Digital Imaging and Communications in Medicine) is a standard of creation, storage, transmission and visualization of medical images and documents of the examined patients.

DICOM file consists of two parts: header and image. Let us explain that in the header there are described the patient's demographic data, information on the model of the device on which the study is carried on, the name of the medical institution, the data about the physician who carried on a study, the study type, the date of its carrying, etc.

The DICOM Ontology has the same structure that the DICOM standard and consists of a real-world model (describes patients, studies, images, and other features of medical imaging) and a DICOM entity model (describes connections between real-world entities and the classes that model the corresponding DICOM information entities).

There are such classes as Patient, Image, Real_World_Object, etc in this ontology. They are described by relevant slots, e.g.: class Real_World_Object has a subclass Patient with slots has_study and makes_visit. Each object in the real-world model contains a slot has_DICOM_IE, which links to the corresponding DICOM information entity that models the real-world object [7].

The DICOM Ontology was created with Protege ontology editor. The DICOM Ontology creators say that it was created to support the caBIG (The Cancer Biomedical Informatics Grid) initiative, and it may serve as a

foundation of systems to be implemented for cancer research and patient care.

3 Conclusion

The overview of medical image ontologies showed the existence of a variety of approaches to the development of such type of ontologies. We might note that among all the approaches most interesting and promising is The DICOM Ontology, working with widely known format of medical imaging DICOM.

References

[1] *Ontology.* http://en.wikipedia.org/wiki/Ontology_%28information_science%29

[2] http://www.debugit.eu/documents/ITAB_2009_IRON.pdf

[3] https://wiki.nci.nih.gov/download/attachments/42828938/Rubin-AAAI-AIM-2008_508_compliant.pdf?version=1&modificationDate=1353815142000

[4] S. Schulz, E. Beisswanger, U. Hahn, J. Wermter, H. Stenzhorn, A. Kumar. *"From GENIA to BioTop–Towards a Top-level Ontology for Biology*," 4th Int. Conf. on Formal Ontology in Information Systems (FOIS 2006), Baltimore, USA, 2006, pp. 103–114.

[5] Christine Fong, Cornelius Rosse, John I. Clark, Linda Shapiro, James F. Brinkley. (2004). *An Ontology-based Image Repository for a Biomedical Research Lab*. In Proceedings, MedInfo 2004, pages 1598, San Francisco, CA., http://sigpubs.biostr.washington.edu/archive/00000144/.

[6] C. Catton, S. Sparks, D.M. Shotton. (2005). *The ImageStore ontology and the BioImage Database: semantic web tools for biological research images*. 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT 2005), 30 Nov – 1 Dec 2005, London.

[7] *Informatics in Radiology: An Information Model of the DICOM Standard*, Charles E. Kahn Jr, Curtis P. Langlotz, David S. Channin, and Daniel L. Rubin, Published online before print October 27, 2010, doi: 10.1148/rg.311105085 January 2011 RadioGraphics, 31, 295-304.

[8] http://bmir.stanford.edu/projects/view.php/dicom_ontology_do_project

Natalia Bruc

Institute of Mathematics and Computer Science of ASM
E-mail: nataliebruc@yahoo.com

Modeling of Access Control System of Gas Transportation Enterprise Portals

Aleksandr Vladimirovich Demidov

Abstract

Restriction of access to the gas transportation system of management companies is a multilevel difficult integrable system. In this paper there was developed a formal model of access control in a network of corporate portals gas transportation enterprise.

Keywords: gas transportation enterprise; corporate portal; access control.

From the point of view of control, gas transportation enterprise is classified as complex and multi-layered one, and is characterized by geographically distributed structure, as well as the fact that economic activity is carried out in the special conditions, due to specific activities.

The relevance of automation of gas transportation company is associated with the complexity of the gas transportation system and the need to improve the reliability requirements of operation (aging significant part of fixed assets), economic efficiency, and the reduction in unit costs of production and transportation. Significant impact on the automation of technological processes and manufactures gas transportation company provides: liberalization of the gas market, increasing requirements for data protection processes and production, tightening environmental monitoring.

Currently, integrated enterprise management system is constructed as a multi-level hierarchical structure, clearly repeating the administrative subordination of dispatch services and production departments of enterprises.

According to leading experts and subject to geographically distributed structure of a typical gas transportation company, the most effective one from the point of view of the architecture of this complex is the ideology of corporate Web portal [1-3]. In this case, the corporate portal is an integrated, Web application class B2E, providing users (employees, customers, partners) a single point of access to the distributed information resources of the enterprise dedicated to them.

The multiplicity of users entails the *complexity of information resources management* of *enterprise portals*, as well as the *inability to determine the overall access control policy.*

Adding to the complexity is the fact that Web-portals transport enterprises have the variability of the structure and are implemented on different hardware and software platforms using heterogeneous and often incompatible, technologies. The unification scheme submission, storage, and access to the system in various businesses is extremely time-consuming and expensive way to solve the problem. This determines the need to build a subsystem of access to information resource network of corporate portals of gas transportation companies. The central task for this system is the task of access control.

A general concept of networking enterprise portals was previously developed in [4]. Consideration of the above features was carried out to develop a model of access control in a network of portals of corporate gas transportation companies. Formally, it is defined as follows:

$$M = \langle U, D, P, R, S, Z, W, F \rangle,$$

where
$U = \{u_1, u_2, \ldots, u_{uc}\}$ – set of access subjects (users), when uc – the number of access subjects. Subjects performed multiple information exchange within the allowed protocols (HTTP/HTTPS). Each subject has a set of unique (within the user domain) identifiers $I = \{i_1, i_2, \ldots, i_{ic}\}$ – set of identifiers, when ic – the number of identifiers. Identifiers are used for unambiguous identification of subjects among all elements of the set U. With the help of multiple authenticators A it is possible to confirm that the subject is the one it claims to be,

hence $A = \{a_1, a_2, \ldots, a_{ac}\}$ – the set of authenticators, iif ac – the number of authenticators. The following functions were introduced: $user : I \to U$, $id : A \to I$.

$D = \{d_1, d_2, \ldots, d_{dc}\}$ – set of user domains, when mc – the number of user domains. Each user u_j is included in the user's domain d_i, i. e. $u_j \subset d_i$, for $j \in \{1, \ldots, uc\}$ and $i \in \{1, \ldots, dc\}$. The following functions were introduced: $dom : Z \to D$, $authorization : w_j \times \{permission(u_k) | u_k \in U\} \times U \to \{ok, access\ denied\}$.

$P = \{p_1, p_2, \ldots, p_{pc}\}$ – set of permissions, when pc – the number of permissions. This set contains all possible access permissions of portal resources. The following relations were introduced: $UP = U \times P$ – relation specifying the correspondence between the subjects and the access permissions; $PH \subseteq P \times P$ – a partial order (hierarchy) on the set of the access permissions, denoted by "$\succeq$". The following functions were introduced: $permission : U \to 2^P$ – function, which assigns the set of access permissions to the subject u_i: $permission(u_i) \subseteq \{p | (\exists p_0 \succeq p) \wedge ((u_i, p_0) \in UP)\}$.

$R = \{r_1, r_2, r_3, r_4, r_5\}$ – the set of roles within the system, respectively, "unauthorized user", "authorized user", "portal administrator", "user". The following relations were introduced: $UR = U \times R$ – the relation specifying the correspondence between subjects and roles, $RP = R \times P$ – the relation, an allocation of the roles and permissions, $RH \subseteq R \times R$ – a partial order (hierarchy) on the set of roles, denoted by "$\succeq$". The following functions were introduced: $role : S \to R$ – the function that assigns one of the roles allowed to R to the session – s_i.

$S = \{s_1, s_2, \ldots, s_{sc}\}$ – session identifier, when sc – the number of such session IDs. The following functions were introduced: $sid : S \to I$, $suser : S \to U$, $sauth : S \to A$.

$W = \{w_1, w_2, \ldots, w_{wc}\}$ – the set of Web-portals, when wc – the number of Web-portals. Each element of this set is a map of Web-portal, presented in a tree – $w_j = (V_j, E_j)$, where V_j – addresses of Web-portal of resources defined as the maximum prefix;

c severed part pointing to the protocol URI (for example – *portal.dom/chapter1/section/page.html*), and E_j – is a set of edges that define the adjacency resources. The contiguity of resources is determined as follows: $(u, v) \in E_j$, if the full address of the resource v is the prefix of the resource u and/or vice versa.

$Z = \{z_1, z_2, \ldots, z_{zc}\}$ – the set of access servers, when zc – the number of access servers connected in the network. An access server can be serviced by several Web-portals, i.e. $z_j = \langle addresses_j, portals_j, d_i \in D \rangle$, where $addresses_j$ – addresses of access server needed for administrative purposes, $portals_j$ – one-to-one correspondence between the portals managed at the access server and addresses of interfaces to which they are connected: $portals_j = \{(w_i, int_k) | (i \leq wc) \wedge (int_k \in InterfacesSet)\}$.

$F = \{user, id, permission, suser, sid, sauth, dom, authorization, role\}$ – the set of the main functions of access control system.

It should be noted that in order to control, the adaptive model may be made in additional limitations on the combination of components, such as the lifetime of the session, the user data connection between the components, etc.

The introduced formal definition (analytical model) of the system elements are the basis for the establishment of formal methods of organization of information exchange within the network enterprise portals of transport enterprises.

References

[1] H. Collins. *Corporate Portals: Revolutionizing Information Access to Increase Productivity and Drive the Bottom Line.*, AMACOM, 2000, ISBN: 0814405932.

[2] D. Sullivan. *Proven Portals: Best Practices for Planning, Designing, and Developing Enterprise Portals.* Addison-Wesley Professional, 2003, ISBN: 0321125207.

[3] J.C.C. Terra, C. Gordon. *Realizing the Promise of Corporate Portals: Leveraging Knowledge for Business Success*. Butterworth-Heinemann, 2002, ISBN: 0750675934.

[4] S.A. Lazarev, A.V. Demidov *The concept of construction of a control system of an information exchange in the network of corporate portals*. Information Systems and Technologies, vol. 4 (60), (2010), pp. 123–129 – (in Russian).

Aleksandr Vladimirovich Demidov

Affiliation/Institution State University – ESPC, Russia, Oryol,
E–mail: a.demidov@ostu.ru

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Tools of exploratory data analysis for ultrasound investigations

Irina Frinea

Abstract

This paper makes introduction to exploratory data analysis and tools of exploratory data analysis for ultrasound investigations.

Keywords: Exploratory Data Analysis, Data Mining, Confirmatory Data Analysis.

1 Exploratory Data analysis

In statistics, exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics [1]. EDA was created and named for the American statistician John Tukey.

According to [2] EDA is the opposite of "Confirmatory Data Analysis" that aims statistical hypothesis testing, calculation of confidence intervals.

In EDA, as defined by Tukey, focus is on visualization, grouping (clustering) and detection of abnormalities are seen as exploratory techniques, these two are distinct subdomains of Data Mining, beyond exploratory analysis [3].

In [4] it is stipulated that EDA techniques are used in order to:

- maximize intimate knowledge of the data;
- reveal basic structure;
- extract important variables;
- detect exceptional outliers / anomalies;
- identify fundamental assumptions to be tested afterwards;

- develop models simple enough;
- determine the optimum state of the parameters;
- suggest some hypotheses on the causes of the observed phenomena;
- available data suggest suitable statistical techniques;
- provide knowledge for subsequent collection of data for research or experimentation.

2 Tools of exploratory data analysis for ultrasound investigations

One of the most widespread diagnostic methods based on medical imaging is the ultrasound investigation. Being noninvasive, relatively low cost in comparison with computed tomography or magnetic resonance investigations, this method obtained a routine character. On the other hand, as a result of these investigations a considerable stock of data is accumulated. These data have an inestimable value for research and training in the domain of ultrasound diagnostic.

Our goal is development of a toolkit which can store, classify and analyse the ultrasound investigations data to identify a priori unknown relationships between ultrasonographic signs and organ pathologies from hepato-pancreatic-biliary area.

As a rule, the physicians from their practice know a lot of factors influencing appearance and progress of diseases or disorders of the organs. There are mathematical methods that determine the interaction of these factors, evaluate degree of their influence on the investigated phenomenon.

Tools, which we propose to draw up, in its pilot version will include: database of investigation results, the set of modules for correlational analysis and multiple regression, module for generating reports, system interface etc.

As a basis for our research SonaRes system [5, 6] serves, with the following modules that we will use in developing of exploratory data analysis procedures:

- The Knowledge Base where data about pathologies presented as a formalized description of the scope and process of decision making are stored.
- Information about each patient that has been investigated (age, rural or urban living environment, living and working conditions, etc.).
- Sessions database that stores complete information about each investigation session. It also contains investigation reports (ultrasound bulletins) generated by the system, their final version (with some possible changes made by the physician), images corresponding to that investigation.
- Collection of annotated images related to pathologies from the knowledge base.

In order to analyze the diseases progress additional investigation might be necessary, such as biochemical analysis and others.

Let us see an example, that ilustrates, how these techniques can be applied. We examine the pathology "Vesicular biliary lithiasis. Soft calculus (pure cholesterol calculus)". It has the following characteristics: The gallbladder contents is abnormal. There are focal modifications there. They are represented as the image with the aspect of mass-like appearance, stable, with homogeneous structure. They are the echogenic structures. Neither acoustic shadow, nor posterior reverberation is formed.

After the selection of patients with pathology "Vesicular biliary lithiasis. Soft calculus (pure cholesterol calculus)" another classification is made according to their living environment. Now one can analyze the frequency of this pathology appearance and determine the factors leading to this pathology occurrence. Further, for patients with the same living environment we could analyze the presence of other pathologies and factors which give rise to them. Description of these relationships can help to establish possibility of occurrence of new pathologies or evolution of existing ones.

3 Conclusion

Exploratory analysis toolkit, the concept of which was discussed in this paper, is intended to process data which are obtained as a result of the ultrasound investigations and also of some additional data characterizing the patient's condition. Such a system will allow monitoring of the pathologies and forecasting their evolution.

References

[1] *Exploratory data analysis*, http://en.wikipedia.org/wiki/Exploratory_data_analysis

[2] John T. Behrens. *Principles and Procedures of Exploratory Data Analysis*, Psychological Methods, Psychological Method, Vol. 2, No. 2, 1997, pp. 131–160.

[3] Lucian Sasu. *Introduction to data mining, Course 3: Exploring Data*, Transilvania University of Brasov, 2012, pp. 1–64; http://www.slideshare.net/lmsasu/curs-3-data-mining. (in Romanian)

[4] Marina Gorunescu, *Exploratory data analysis methods*, Department of Computer Science, academic year 2012-2013, pp. 1–185 ; http://math.ucv.ro/ gorunescu/courses/EDA/curs1EDA.pdf. (in Romanian)

[5] L. Burţeva, S. Cojocaru, C. Gaindric, E. Jantuan, O. Popcova, I. Secrieru, D. Sologub, *SONARES – A decision support system in ultrasound investigations*, Computer Science Journal of Moldova, vol. 15, no.2 (44), 2007, pp. 153–177.

[6] S. Cojocaru, C. Gaindric. *Clinical Decision Support System SONARES*, International Journal Information Theories and applications, vol. 15, Nr. 2, 2008, pp. 137–143.

Irina Frinea

Institute of Mathematics and Computer Science of the Academy of Sciences of Moldova
E-mail: irina.frinea@math.md

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Reasoning on Semantic Sensor Streams for Smart City

Oxana Hotea, Adrian Groza

Abstract

We propose an architecture for knowledge integration and reasoning on real-time heterogeneous data sensors. The system is demonstrated in the public transport scenario. Sensor and urban ontologies are used together to fill the gap between low level sensor data and high level optimisation decisions.

Keywords: semantic reasoning, C-SPARQL, smart city.

1 Introduction

Traffic control systems represent one of the main application of the Internet of Things technology in the context of smart cities [3]. With the development of sensor networks and stream reasoning, valuable technological support for accessing rapidly changing data is provided. Yet, there is a lack of systems designed to manage them at the semantic level. The proposed reasoning algorithm aims to determine the number of people who would probably take a public means of transport, based on real-time data from sensor data.

2 System Architecture

The top level architecture highlights three layers: presentation, business and data layer (see Figure 1). The interaction is achieved by sending a request from the mobile phone application. The application

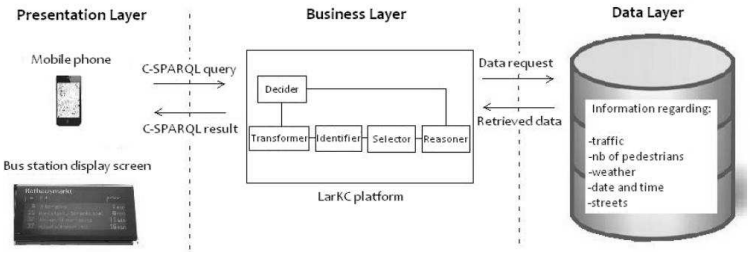


Figure 1. System architecture.

consists in a map where the current position of the user is indicated. Users can register their destination and can ask for routes or the time of arrival for a specific bus. For the business layer, we adopted the Large Knowledge Collider (LarKC), a platform for massive distributed reasoning that removes the scalability barriers of currently existing reasoning systems for the Semantic Web. Knowledge about street topology and amenities come from OpenStreetMap in XML format. The pedestrians' movement and their destinations are captured and recorded by the wireless network center. The weather and traffic information are gathered from sensors and cameras around the city. These heterogeneos data were integrated to optimise traffic.

Knowledge Representation. Data is collected from various types of sensors: temperature sensors ("Sensor T15 has a temperature reading of 17 C"), from the humidity sensors ("According to sensor H21, precipitation risk is 40 percent"), data regarding time and date ("Tuesday, April 15, 2012, 15:05"), traffic data from cameras and sensors located on route to destination ("Data from cameras C16 and C17 show a 65 percent congestion probability"), information about amenities in a particular area ("According to the information extracted from OpenStreetMap, there is a school in the given area"). We developed the sensor ontology for structuring sensor related knowledge. Figure 2 depicts the temperature and humidity sensor. Concepts are represented by ellipses, roles by arrow, and individuals by rectangles.

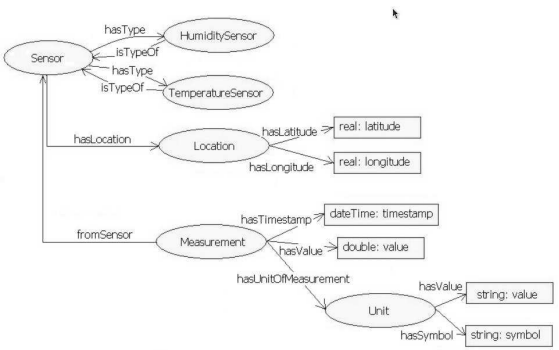


Figure 2. Part of the sensor ontology.

Street topology data was obtained using OpenStreetMap, which gives users the possibility to choose an area by its coordinates and export it in XML format. The information extracted from the XML files includes: knowledge about buildings, knowledge about streets and traffic area, and knowledge about transportation networks. The buildings ontology was used to classify buildings depending on the number of people who attend them. For example, schools and museums will be of type $FrequentlyAttendedAmenity$, given in description logic by $Museum \sqcup HighSchool \sqsubseteq Building \sqcap FreqAttendedAmenity$. Depending on the types of amenities in a given area, the decision of updating or not the public means of transport timetable will be taken.

Reasoning. Streams of data are continuously analysed in order to take decisions in real time. The continuous query in figure 3 counts the number of possible passengers in each station and returns those stations with more than 50 pedestrians.

3 Discussion and Conclusion

The association between data coming from semantic sensor networks and the existing data sources is mandatory in order to achieve high

```
REGISTER QUERY StationsWithManyPassengers
PREFIX s: <http/urbanontology/camera#>
SELECT DISTINCT ?station ?passangers
FROM STREAM <www.camerasensor/cam11.rdf>
     [RANGE 30 MIN STEP 1 MIN]
WHERE {?passengers s:identifiedIn ?station . }
AGGREGATE {(?passengers, COUNT, {?station} )
FILTER (?passengers > 50)}
```

Figure 3. Continuous processing with C-SPARQL.

quality process for decision making [2]. Analysing spatio-temporal dynamics of visitor movements at mass events exploits proximity-based Bluetooth tracking in [4]. In our case, different types of sensors are exploited in order to figure out the situation. In both approaches, one issue regards balancing behavioral privacy and information utility [1].

Acknowledgment. Part of this work was supported by the PN-II Romania-Moldova bilateral contract "ASDEC: Structural Argumentation for Decision Support with Normative Constraints", 2013-2014.

References

[1] S. Chakraborty, Z. Charbiwala, H. Choi, K. R. Raghavan, and M. B. Srivastava. Balancing behavioral privacy and information utility in sensory data flows. *Pervasive and Mobile Computing*, 8(3):331–345, 2012.

[2] S. Hasan, E. Curry, M. Banduk, and S. O'Riain. Toward situation awareness for the semantic sensor web: Complex event processing with dynamic linked data enrichment. In K. Taylor, A. Ayyagari, and D. D. Roure, editors, *SSN*, volume 839 of *CEUR Workshop Proceedings*, pages 69–81. CEUR-WS.org, 2011.

[3] D. Miorandi, S. Sicari, F. Pellegrini, and I. Chlamtac. Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, 10(7):1497–1516, 2012.

[4] M. Versichele, T. Neutens, M. Delafontaine, and N. Weghe. The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent festivities. *Applied Geography*, 32(2):208 – 220, 2012.

Oxana, Hotea[1], Adrian Groza[1]

[1]Department of Computer Science/Technical University of Cluj-Napoca, Romania
E-mail: Adrian Groza@cs.utcluj.ro, Oxana.Hotea@student.utcluj.ro

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Decomposing Morphological Rules of Polish for Adaptive Language Learning

Mykola Kostikov

Abstract

The paper is concerned with the approaches to decomposing learning material in intelligent tutoring systems. The principles for building a domain model for the intelligent tutoring system of Polish morphology are developed.

Keywords: adaptive learning, e-learning, grammar, intelligent tutoring systems, knowledge engineering, language model.

1 Introduction

The modern informatization of the education calls for developing different approaches to the teaching process. To benefit from applying information technology, we should change the structure of learning material. Methods of artificial intelligence help make the learning process adaptive and personalized, so as to define the individual problems and demands of each user and react to them accordingly.

2 Need for the language model

By now there is still a lack of language learning software for Slavic languages, and the existing ones have substantial drawbacks. The analysis of modern programs for Polish has shown that they often lack the feedback which is an indispensable condition of adaptive and personalized learning. Teaching Slavic languages in other Slavic-speaking countries also has its specifics, e.g. grammar becomes of much greater importance. Thus we should consider the peculiarities of the given grammatical system and develop a learning system based on a language model.

3 Decomposition of learning material

Decomposition of learning material may be viewed in two ways. First of all, we can divide it into topics, units, lessons and so on, design separate learning objects so as to meet the SCORM technical standards or other formal requirements. Such decomposition may be called structural.

It is also possible to decompose the knowledge itself by developing a hierarchy of learning elements (ontology of terms and notions, semantic network etc.). At first we define the basic elements of knowledge and relationships between them. Thus previously acquired knowledge becomes the basis for solving more complex problems. Intelligent tutoring systems implement this step-by-step approach, being able both to solve problems and explain this process to users thanks to the domain model which lies underneath and is stored as a knowledge base.

4 Decomposition of Polish morphological rules

The traditional approach to the morphology formalization implies using relational models. One of the co-developers of the Polish grammatical dictionary SGJP, M. Woliński notes that 'a relatively compact relational model can be used to describe Polish inflection in a uniform way' [1]. These models are useful for quick processing big amounts of data, e.g. in morphological analysis, but difficult to use in the language learning process due to its specifics. Relational models are based on inflectional patterns within which the inflectional rules are the same, thus the number of patterns may reach several hundred to take into account all the subtleties, such as phonological alternations. Yet the conditions for choosing a specific pattern aren't described in these models.

As one of the closest approaches to the 'intelligent' grammar formalization we consider the 'smart paradigm', a notion introduced by the co-developers of MOLTO — multi-language online translation tool. Smart paradigms are 'heuristic functions that take just one or a few forms of a word as their arguments and infer the complete inflection table of the word from them' [2]. Such meta-paradigm 'inspects the base form and tries to infer which low-level paradigm applies' [3]. Unfortu-

nately, the base form itself isn't enough to determine the inflectional pattern in many cases, so the smart paradigms use additional word forms provided by the users — linguists and speakers of the language.

Instead of this, we propose using additional parameters which are more likely to be known or found by the students, e.g. gender and animacy for nouns, which are presented in dictionaries or can be inferred from the context. They not only facilitate the morphological synthesis, but also make our problem solver algorithm more similar to human logic and thus easier to explain to the students. Furthermore, our main idea is to move away from processing a paradigm as a whole. We should analyze specific word forms and separate out the basic elements of morphological rules (adding certain endings etc.) and transformations (e.g. ń+<vowel>→ni, interchange t→ć), so-called elementary rules. We should also investigate which of them are common and different between various patterns and define the factors which impact on their choice. Then we may analyze where and how often each of them occurs in a certain vocabulary, select certain words for lessons, build the hierarchy of learning elements and logical learning sequences.

By now we have decomposed elementary rules for Polish nouns and tested the model on the 350 most frequent (according to [4]) nouns. The experiment has proved that such decomposition and using additional parameters help to resolve ambiguities in inflection in most cases.

5 Adaptive grammar learning

The presented model is being created as a part of the intelligent tutoring system for learning language in context. In theory, such model allows of using any word for generating learning examples and exercises, as it determines the inflectional rules based on the word itself and its general characteristics. The main drawback is the need for providing the additional parameters by the knowledge engineer and/or the user. Besides, grammatical exceptions should still be taken into account.

The feedback based on the model should ensure the personalization of the learning process in two aspects. First of all, we can analyze

the nature and reasons of the students' mistakes more precisely, as within the inflectional patterns model it's hard to detect and react to users' problems in applying specific components of a complex inflection process, e.g. interchange of consonants. Secondly, we can analyze the interaction between grammar and vocabulary and create individual learning plans based on the occurrence of certain rules in a given vocabulary (depending on user needs) and vice versa, since learning in context helps both to build grammar skills and learn words.

6 Conclusion

The described approach helps to model the inflection of Polish nouns. Applying it to other parts of speech and developing an intelligent tutoring system based on this model are the subjects to further research.

References

[1] M. Woliński. *A Relational Model of Polish Inflection in Grammatical Dictionary of Polish.* Human Language Technology – LTC, vol. 5603 (2007), pp. 96–106.

[2] A. Ranta. *How Predictable is Finnish Morphology? An Experiment on Lexicon Construction.* Resourceful Language Technology, 2008, pp. 130–148.

[3] G. Détrez, A. Ranta. *Smart Paradigms and the Predictability and Complexity of Inflectional Morphology.* Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012, pp. 645–653.

[4] J. Kazojć. *Słownik frekwencyjny leksemów V.06.2009*, 2009. http://www.slowniki.org.pl/slownik-frekwencyjny-leksemow.pdf.

Mykola Kostikov

National University of Food Technologies, Kyiv, Ukraine
E–mail: Mikolasz@bigmir.net

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Uncertain Reasoning Models Transformations For a Model Selection

Natalya Morosanova

Abstract

Uncertain reasoning models define the way to deal with uncertainty from various sources. The large number of different models results in the problems of their coupling and selection. These problems are addressed in the paper, and a method of model selection based on the transformation of the models is proposed.

Keywords: uncertain reasoning model, expert system, inference, isomorphism

1 Introduction

Choosing the way of representing uncertainty is a key problem in intelligent systems construction [1]. Moreover, the rapid development of the Internet provides the means for the integration of different systems. This integration is based on the transformation of one uncertain reasoning model into another one. Each uncertain reasoning model defines the value set [false ... true] which the proposition can take, and the rules of the uncertainty propagation during the inference process, usually, by the means of the operations defined on the said value set.

The problem of the uncertain reasoning models transformations is both of theoretical and practical interest. From the theoretical standpoint, discovery of the relations between different models leads to a better comprehension of their validity and design of generalizing models. As for the practical applications, the results obtained in the study of different models relation can be used to solve the problems of model

selection and coupling of different models in one system.

2 Uncertain reasoning models transformations

The problem of uncertain reasoning models transformation received attention during the last 30 years [2–5]. This problem consists in finding a transformation from one uncertain reasoning model value set to another one. It is required that the sought-for transformation possesses some properties, such as preservation of some special values in two value sets: true, false, unit element, etc., element order in two value sets and at least one pair of operations.

The results presented in the papers which address this topic can be divided into two categories: obtained transformations for the pairs of uncertain reasoning models and general approaches to the transformation construction.

The first category includes:

- transformation between certainty factor model and a special case of Dempster-Shafer evidence theory [2];
- transformation between certainty factor model and subjective Bayesian method [3];

The second category includes axiomatic [4] and algebraic [5] approaches. These approaches propose construction of a transformation with the use of an intermediate model. Axiomatic approach defines the testable and desirable properties (axioms) of the operations that deal with uncertainty. This approach defines the intermediate model as a model with a fixed set of axioms on a pre-defined value set (i.e. unit segment). Algebraic approach views an uncertain reasoning model as an algebraic system (i.e. an ordered Abelian group). It uses the algebraic morphism notion to construct the transformation between uncertain reasoning models.

2.1 Isomorphic transformations

A special case of a transformation is the isomorphism, which can be described with the following functional equation

$$h(f(x_1, \dots x_n)) = g(h(x_1), \dots h(x_n)), \tag{1}$$

where h is a transformation function, f and g are the operations defined in two uncertain reasoning models. It is worth mentioning that this type of transformations is the most widespread in the literature. Other types of transformations include generalization of operations and extension of the value set.

2.2 Functional equation solution for the addition

Let us consider a special case of Eq. 1 when one of the operations is an addition. When one of the operations is an addition, the Eq.1 can be rewritten as

$$\frac{h(x+y)-h(x)}{y} = \frac{g(h(x),h(y))-h(x)}{y}. \tag{2}$$

Then, passage to the limit $y \to 0$ implies that

$$h^{'}(x) = \frac{\partial g(z,y)}{\partial y} h^{'}(y)\Big|_{y=0,z=h(x)}. \tag{3}$$

If the solution for the Eq. 3 exists, it is a candidate for a solution of Eq. 1.

2.3 Example of certainty factors isomorphism

This method for finding a transformation function can be also used for the operations other than addition, but isomorphic to it through the known transformation, for example, a multiplication operation.

Let us consider an uncertain reasoning model selection when there are only inference examples in the following form:

$$if\ A\ then\ B\ with\ c, \tag{4}$$

and some (A, B, c) tuples are given. The goal is to propose an uncertain reasoning model which the given examples hold in.

The proposed method to solve this problem is to search for the model as a transformation of some well-known model. Let us use the

certainty factors model. Then the functional equation for the transformation is

$$h(xy) = g(h(x), h(y)), \tag{5}$$

where $g(x, y)$ is an approximation for an unknown operation, defined by points $g(A, c) = B$. This equation (Eq. 5) can be solved using the above-mentioned passage to a differential equation.

3 Conclusion

In this paper different uncertain reasoning models and their transformations are reviewed. The isomorphic transformations are discussed in greater detail, and their application to an uncertain reasoning model selection is proposed.

References

[1] H.J. Zimmermann. *An application-oriented view of modeling uncertainty.* European Journal of Operational Research, vol. 122, no 2 (2000), pp. 190–198.

[2] B. Grosof. *Evidential confirmation as transformed probability.* In Uncertainty in Artificial Intelligence, pp. 153–166, 1986.

[3] X.Luo, C.Zhang., H.F. Leung. *Information sharing between heterogeneous uncertain reasoning models in a multi-agent environment: A case study.* International Journal of Approximate Reasoning, vol. 27, no 1 (2001), pp. 27–59.

[4] Y. Cheng, R. L. Kashyap. *An axiomatic approach for combining evidence from a variety of sources.* Journal of Intelligent & Robotic Systems, vol. 1, no 1 (1988), pp. 17–33.

[5] P. Hajek, J. J. Valdes. *An analysis of MYCIN-like expert systems.* Mathware & soft computing, vol.1, no 1 (2008), pp. 45–68.

Natalya Morosanova

Lomonosov MSU, CMC department, PhD student
E–mail: nmorosanova@gmail.com

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Extraction of rules from the process of merging formal contexts

Daria Stelmashenko

Abstract

The formal context analysis theory [1] represents all the objects from the knowledge domain as sets of attributes, stored in a formal context. The procedure of merging several formal contexts [2] leads to the growth of the knowledge domain and makes the artificial model closer to the reality. During the merging procedure a number of pseudoconcepts, which have inconsistent attributes, are verified. Comparison of inconsistent sets of attributes with those that correspond to real formal concepts allows building a certain amount of rules. These rules can be used to expand the knowledge base of the FIACR system, described in [3].

Keywords: formal concept analysis, merging of formal concepts, information retrieval, rule generation, FIACR system.

1 Introduction

Formal concept analysis (FCA) [1] is a theory, which allows describing any object of the knowledge domain (KD) by a set of attributes. A formal context can be determined as a set of objects, representing a certain KD. The task of merging several formal contexts becomes a challenging problem because it allows to broaden the scope of the considered KD. The existing context merging procedures [4, 5] generate a global context with objects that have incomplete sets of attributes and thus are described incorrectly in the scope of the considered KD. The papers [2, 6] describe an alternative method for merging contexts, which generates a global context with full and consistent sets of attributes.

This means that the generated formal context correctly represents the corresponding KD and includes all possible objects of the KD that exist in the present attribute set. This condition makes it possible to use the global context as a basis for rule generation procedure, which is based on comparison of the objects of the global context and the pseudoobjects, which proved to be non-existent in the considered KD. A more detailed explanation of this process is given below.

2 Procedure of merging formal contexts

Let us call a class any nonempty set of objects of the formal context. All the common attributes of the objects of this class are called the intent of this class. A formal concept (FC) [1] can be defined then as a class of objects that includes the largest possible number of objects with the same intent. The set of attributes of the formal concept is called the intent of the FC and the set of objects is called the extent of the FC. In this paper we will equate the terms formal concept and intent of the FC. This assumption simplifies the contexts merging method, described in [2], and does not bring in any limitations to further discourse because any FC is uniquely identified by its intent.

The algorithm of merging two formal contexts, described in [2], is based on the process of merging the sets of formal concepts of these contexts. On the input of this algorithm the pairs of contexts, which together form the yet unknown global context, are given. For each pair of contexts a set of candidates into the FC of the global context is built by constructing all possible combinations of the FC of the given contexts. It is proved in [2] that if we iteratively build such sets of candidates into the FC of the global context, and each time intersect these sets and leave only their common elements, then at some iteration we will get the set of candidates that will consist only of the FC of the global context. Because the set of FC uniquely defines the formal context, the desired global context is found. The candidates into the FC of the global context, which didn't prove to be real formal concepts, are called pseudoconcepts [2, 6].

3 Rule generation procedure

Let us assume that the global context has already been built by the method, described in [2]. Let D be the set of the FC of the global context. Let us also assume, that during the iterative process of excluding the pseudoconcepts from the sets of candidates into the FC of the global context, briefly described above, all the pseudoconcepts were saved in the set R. This means that all the elements from the constructed sets of candidates into the FC of global context, which didn't prove to be the real FC of the global context, have been saved in the set R. Let us exclude all the elements from R, which fully lie in any element from D. This excludes the valid combinations of attributes from R and guarantees that the set R now consists only of the non-existent combinations of attributes. Each of these combinations may be regarded as a rule, stating that the corresponding set of attributes is inconsistent and that these attributes cannot simultaneously belong to the same object.

The retrieved set of rules is redundant as it may contain sets of attributes, which include some other sets of attributes from R. In this case only the shortest sets of attributes must be kept, as they automatically form a rule, according to which the longer set is inconsistent. Once all the elements of the set R, which include any other element from this set, are deleted, the set R will contain the full and irredundant set of rules. For example, if we consider two sets of attributes: A, $B \in R$, $A = (a, c)$, $B = (a, b, c, d)$, then $A \subset B$ and the set B should be excluded from the set of rules R, as it is uninformative, because if there is a rule, which states, that attributes a and c are inconsistent, then any combination of attributes, which includes (a, c) is also inconsistent.

4 Conclusion

The rules, retrieved as a side effect of the method of merging formal contexts, described in [2], verify the inconsistent sets of attributes, which cannot be superposed in the scope of the KD, corresponding to the constructed global context. These rules can be used in the knowledge

base of the FIACR [3] expert system shell, which implements direct monotonic inference, based on the rules from its knowledge base. The constructed rules can also be used in any other system or algorithm, which uses rule-based inference.

References

[1] B. Ganter, R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer (1999).

[2] D. Stelmashenko. *The algorithm for global context restoring*. A collection of articles by young scientists of MSU CMC faculty, vol 9 (2012), pp. 191–209.

[3] S. Solowiev S, J. Pechersky, G. Solowieva. *The Principles of Expert Recognition Systems Design*. Internanional Journal of Pattern Recognition and Artificial Intelligence, vol 3, No. 1 (1989), pp. 113–119

[4] P. Valtchev, R. Missaoui, P. Lebrun *A partition-based approach towards constructing Galois (concept) lattices*. Discrete Mathematics vol. 256, no.3 (2002) pp. 801–829.

[5] Li Guan-yu, Liu Shu-peng, Zhao-Yan *Formal Concept Analysis based Ontology Merging Method*. Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference, vol 8, (2010) pp. 279–282

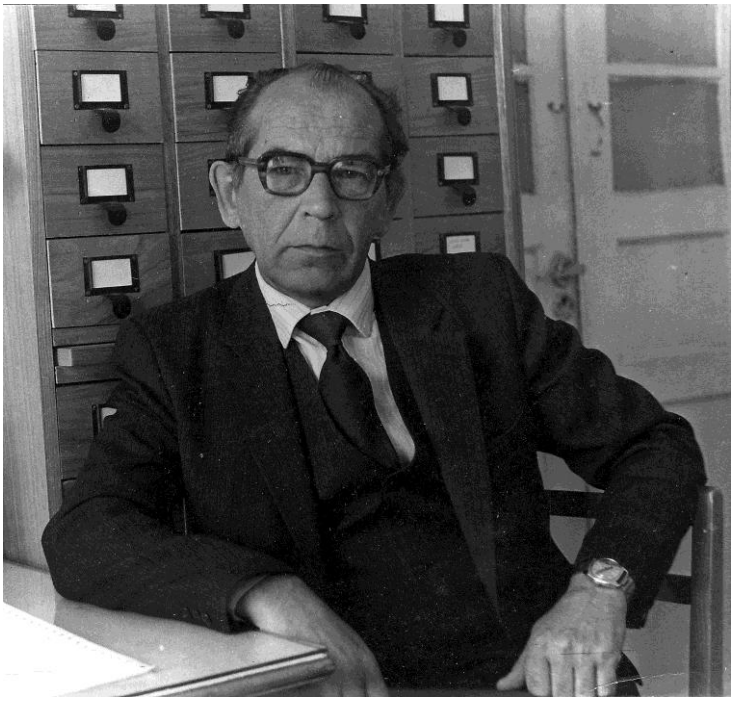
[6] S. Solowiev, D, Stelmashenko *Approaches to research on formal contexts* Journal of information processes, vol 11, No. 2 (2011), pp. 277–290.

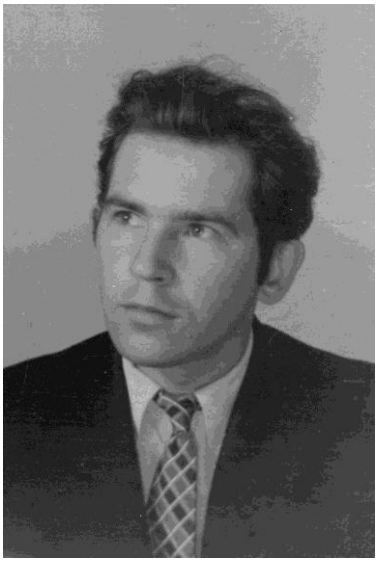
Daria Stelmashenko

Moscow State University
E–mail: dashanikolaeva@gmail.com

Part 4

IN MEMORIAM OF Yuri Nikolayevich Pechersky





Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Thankful memories of Yuriy Nikolayevich

Gennady Andrienko, Natalia Andrienko

We had first known about Yuriy Nikolayevich Pecherskiy and his laboratory at the end of 1980-ies, when we looked for an opportunity of post-graduate study. One of our friends, who attended a school organized by Yuriy Nikolayevich, was very impressed by the interesting topics and enthusiastically recommended us to get in contact with him. After several phone conversations and a personal meeting in Kishinev, we were invited to passing the entrance exams for the post-graduate programme at the Institute of Mathematics. After almost 30 years, we can say with certainty that this was a turning point in our academic careers and lives.

In 1988, we have entered the post-graduate programme at the Institute of Mathematics and started our studies in the laboratory of Mathematical Cybernetics lead by Yuriy Nikolayevich. Our scientific advisor was Sergey Solovyov. Over the three years of our study, we always felt the sympathy, support, and assistance of Yuriy Nikolayevich. His advices were invaluable for us as young scholars and just young people. Apart from the purely academic support, he helped us in the search for applied research, which enabled us to successfully cope with the known problems of the transition period.

Unfortunately, the collapse of the USSR in the early 1990s that resulted in very difficult conditions for the academic science and threats to its survival forced us to leave Moldavia for the region of Moscow and, later, to accept the invitation to work in Germany at the National Centre for Mathematics and Data Processing (GMD), now the Fraunhofer Institute IAIS. These decisions were very hard for us. We are sincerely grateful to Yuriy Nikolayevich for his support and understanding during this difficult period.

We remember our last conversation before our departure from Kishinev. Yuriy Nikolayevich was very hopeful that the situation would

change for the better and we would be able to come back. Unfortunately, our hopes did not come to reality. We often recall this conversation and, as before, feel deep gratitude to Yuriy Nikolayevich for his warmth and paternal care.

Yuriy Nikolayevich specially came to Moscow in 1993 to support us at the defense of the candidate thesis of Natalia Andrienko at the Moscow State University. Sadly, that was our last meeting.

Gennady Andrienko[1], Natalia Andrienko[2]

[1]Fraunhofer Institute IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
E-mail: gennady.andrienko@iais.fraunhofer.de

[2]Fraunhofer Institute IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
E-mail: natalia.andrienko@iais.fraunhofer.de

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

A Good Farmer Makes a Good Farm

Vsevolod Arnaut

Every person has its own specific path when shaping its personality, through different expected experiences, and also through people who have a big impact over us. One of the most important period that contributed to my professional development was when I started working at the Institute of Mathematics of the Academy of Sciences of Moldova, a place that I have been hearing of since I was a student due to its famous personalities in Mathematics, such as: academician Vladimir Andrunachievici, professor Valentin Belousov, doctors habilitatus Vladimir Arnautov and Iurie Reabuhin. Working at the Institute, I had the opportunity to know all the researchers and then I realized that even if they weren't globally famous, they were very appreciated in scientific circles, due to their wide experience, from which I would learn along the way.

After some changes in my job position that occurred because of my army enrollment for a period of 2 years I rejoined the work at the Institute of Mathematics of the Academy of Sciences of Moldova, but this time in the Laboratory of Mathematical Cybernetics. I had there the most pleasant, competent and responsible colleagues: Vladimir Levcenco, Vladimir Burdaev, Serghei Soloviov, Mihail Vexelman, Fiodor Frolov, Vasile Sârbu, Alexandr Andrunachievici, Slavian Guţu. We were all different, and only a strong personality like the Head of the Laboratory, Iurii Nicolaevici Pecerskii was the one who could build us as a team and lead us to great performances.

I will further try to highlight those professional and personal qualities which in my opinion had a big role in defining the managerial gift of the Head of Laboratory, Doctor Iu. N. Pecerskii.

1 Bringing in young researchers

An important indicator in a research group is the presence of a powerful core made of young researchers. Such a research polygon was the

Laboratory of Mathematical Cybernetics, in which I was working. It's quite surprising how the Laboratory of Mathematical Cybernetics could bring in so many young people, even if they did not have any previous interaction with Moldova. This was the case of Ghenadii and Natalia Andrienco, natives of Ukraine, who were enrolled on a postgraduate program at the Institute of Mathematics. They were very active in the laboratory's activities, and became very soon researchers in expert systems domain. During the same period, Ghenadie Ghincul, Alexandr Zincenco, Alexandr Savinov, graduates of prestigious Russian institutes, became also part of the laboratory. An important thing to be added is that Doctor Iu. N. Pecerskii always encouraged and supported young students from the State University, to do their internship in the Laboratory of Mathematical Cybernetics. This was a great opportunity for young researchers, but also a great way to test them before being employed at the lab. Having students enrolled in a post doctorate program at the lab, was also one of the most important priorities.

2 Culture

Coming from an intellectual family, he had a fine sense for all that is known as art. He read an imposing number of books, from different categories, such as science fiction, documentary literature and scientific literature. Besides this, he had a special passion for painting, his works being exposed in the Institute of Mathematics. Later, because of a very busy schedule, he had to give up his passion, but he continued to encourage all the cultural projects. This way, when the Academia Chapel choir was created, he was emboldening all the young people who loved music to participate, even if this was taking them 2 hours of their weekly work program.

One great quality that Iu. N. Pecerskii had, his communication skill, he was able to solve problems without the need of raising his voice, but with a lot of calm and tactics.

3 Research organization

During its existence the lab had several research directions. First of all the decisional models based on collective expertise should be mentioned. Dr Iu. N. Pecerskii has actively participated in the research and use of these

models, thus having published the work called Expert estimations in problems of technical-scientific programs maintenance. The work presented an investigation of the stages which lay at the basis of the collective expertise models and the ways of implementing each of the stages, based on technical-scientific programs example. The work Interactive system for collecting and processing expert information contains a detailed description of the use of collective expertise in scientific programs estimation, on the basis of a specialized system created for expertise organization. A series of applications have been elaborated which served as a basis for the implementation of all the stages of expertise for the evaluation of some alternatives in the scope of determining the best one. A series of applications have been elaborated which served as a basis for the implementation of all the stages o expertise for the evaluation of some alternatives in the scope of determining the best one. SIREX and KIOT are two of this kind of applications. Another productive direction was the problematic of expert systems. There have been elaborated expert systems prototypes on different topics. Thus the team managed by Mr. S. Soloviov, has created the expert system FIACR, oriented on tomatoes selection. Having a background on cluster analysis and using this experience, Mr. V. Burdaev has elaborated an expert system oriented on choosing the most adequate methods of clustering. Mr. V. Levcenco and Mr. A. Savinov have created the EDIP expert system which allowed the use of fuzzy knowledge.

The organization of scientific conferences was another good method to stimulate research. The annual conferences on artificial intelligence held in places like Casa Tehnicii, Vadul lui Voda, etc., had become a tradition. The conferences created the great opportunities on meeting new, important people in the field, all due to the merit of the Head the Laboratoty Dr Iu. N. Pecerskii.

4 Endowment of the Laboratory

Having a well equipped laboratory in the 80's, was a great deal. If compared with nowadays informational technologies, this might seem insignificant, however 20 years ago you couldn't even dream about it. The Laboratory of Mathematical Cybernetics was one of the very well

equipped labs at that time, having the first mini computer Iskra 226. Compared to contemporary equipment, it would look like a monster, but at that time it looked quite curious and cute. The next generation computers were, ES 1840 and Prawitz, and along with them the computer park has started to grow bigger and better with equipment such as the XT model, AT, and so on. Having all these devices, was a great deal for the laboratory, that being a huge step in researches and offering new possibilities in gaining contracts related to the research themes. All that was possible due to the perseverance of Doctor Iu. N. Pecerskii.

5 Science popularization

Another great quality of Doctor Iu. N. Pecerskii, was the ability of scientific results popularization. Not everybody could talk about scientific achievements so clearly that the audience would understand complicated aspects, but he had this way of simplifying and at the same time, keeping the essential of important results. It can be said that he was the first person in Republic of Moldova who talked about intelligent systems and about computers, making this information clear to people who had no previous background in this field. At that time, I was given the chance, together with Slavian Guţu, to translate the following books from the science popularization collection signed by Doctor Iu. N. Pecerskii: *Intelligent systems* and *Notes on computers*.

6 Voluntary work

It may not seem important at this moment. Nevertheless volunteering denotes the altruism and opening to the community. It's true that you don't get to be paid for that, but you certainly have a great feeling when you know you did something good and important for the society. During Soviet Union, Wall newspapers were a special way of reflecting the activities in many organizations. Many years Doctor Iu. N. Pecerskii was the most important figure in editorial body, due to his artistic and full of responsibilities spirit.

7 Material incentives

The 90's brought a big downturn period for the whole Academy, that including the Institute of Mathematics. Little, late salaries were

endangering the researches. Even in these conditions the Laboratory of Mathematical Cybernetics succeeded in signing contracts and in selling program products. All these were the merit of the managerial gift of Doctor Iu. N. Pecerskii, encouraging the lab's employees with remunerations as a result of their work.

8 Devotion and respect

Dr. Iu. N. Pecerskii had a great devotion for the Institute and a great respect for all the coworkers. The lab was his second home, since he was there from early morning to late at night, even on weekends or holidays. He has built a strong relation with his colleagues, always being there for them in their delicate or difficult situations, using his diplomatic and tactical manners to help them get through.

The whole activity of Doctor Iu. N. Pecerskii was directed to the development and prosperity of the Institute of Mathematics.

Vsevolod Arnaut

Moldova State University
E-mail: arnaut_s@yahoo.com

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

Scientist, Popularizer and Guide of Artificial Intelligence in Moldova

F.V. Frolov

It is always very difficult to recall a man which has gone early and which has been scientist, teacher, head and a personality for you. But I would like to make a low bow and to express my thanks to Yurii Nikolaevich Pecherskii. The first meeting in that remote 1978 in Kishinev and 10 spent years – it is a school which has left an indelible trace for all life.

Yurii Nikolaevich Pecherskii set up the laboratory of Mathematical Cybernetics in the Institute of Mathematics of the Academy of Sciences of Moldavian SSR and became its leader. In this laboratory young scientists had gathered, who took up one of the new and advanced directions in science – artificial intelligence. This direction of the laboratory united the works from development of informational system for science management for Science and Technology Department of State Planning Committee of MSSR – DIANA, robotics, recognition methods, up to leading directions in the field of methodology and expert systems creation. Under the leadership of Yurii Nikolaevich the specialists of the laboratory published hundreds of scientific articles, participated in many national, All-Union and international scientific conferences. Having an excellent style, Yurii Nikolaevich himself prepared and published a number of scientific monographs which became classical for us, collaborators of the laboratory.

Along with this, Yurii Nikolaevich became not only the leader, but also the popularizer of these directions in the republic. He organized edition, and became the author of a number of popular brochures for schoolchildren, one of which being "Intellectual systems". Yurii Nikolaevich had a gift to bring to schoolchildren the advanced directions in science in the field of artificial intelligence by easy, clear style. Just this

was not the end of his work with schoolchildren. Under his leadership for the first time in the Republic of Moldova there had been organized a school club of robotics. We managed to equip this club of robotics by electronic hardware and equipment, which was passed from the institutes of the Academy of Sciences. Participants of the club of robotics showed their exhibits at national and international conferences. The exhibit which was demonstrated in Bulgaria, was requested for usage at national enterprises. So, Yurii Nikolaevich found time for all this, considering the work with youth very important and necessary. It seemed that time became slower for him. He worked 18-20 hours a day.

I would especially like to point out the ability of Yuri Nikolayevich to gather scientists and specialists in different directions around him. He actively communicated with Ministry of Communications, Department of Science and Technology of the State Planning Committee of the Republic, institutes of the Academy of Sciences of the Republic, Kishinev Polytechnic Institute and other organizations not only in the Republic, but also in the Soviet Union. These are the leading centers in Moscow, Leningrad, Kiev, Minsk, Novosibirsk, Tomsk, Odessa, Kharkov, Vladivostok, Irkutsk and other cities. The organized by him in the 80th,

All-Union school-seminar “Logico-combinatorial methods in pattern recognition” under the leadership of D.A.Pospelov (at that time he was Deputy Chairman of the Committee on Artificial Intelligence of the USSR), became the place of meeting for more than 40 scientists from different scientific centers. This school-seminar was carried out in a beauty spot of Vadu lui Voda village at the bank of Dniester. This fact contributed to the active work in combination with excellent recreation.

Thanks to the efforts of Yuri Nikolayevich I had the luck to take part in the work of international fundamental laboratory on artificial intelligence in Bratislava. It was an excellent time of the science growth in the field of artificial intelligence systems creation: fifth generation computers, expert systems, et al. Today one can understand not only scientific importance of the work carried out at that time, but also the practical one.

When working today in atomic field in Open Joint-stock Company “Concern Rosenergoatom” and permanently meeting the problems of keeping and passing knowledge of the staff engaged in control of the work of reactors at atomic power stations, one can understand the role which the expert systems could play. Hundreds of instructions, directing documents with regulations on control of atomic reactor are still waiting today the implementation of modern expert systems at the background of creation of which Yuri Nikolayevich Pechersky had been, and who will remain the scientist, the teacher and an excellent person for me forever.

F.V. Frolov

E-mail: frolov_fv@mail.ru

Our Pechersky

G.Ginkul, S.Soloviev etc.

We[1] worked together in the Laboratory of Mathematical Cybernetics of the Institute of Mathematics. The head of the laboratory was Yuri Nikolayevich Pechersky (1934 – 2003). We remember Pechersky. Time has put almost everything in its place, leaving us with a moral duty to describe the magic of the person who was our chief.

Objectively, Y.N.Pechersky created a very successful research team that was able to creatively solve complicated problems of data analysis, artificial intelligence, expert systems and fuzzy sets. It is remarkable that this laboratory was a second home for us. We were eager to come there, we were eager to work there. Pechersky was the one who presided. He created – we used, he bestowed – we accepted, he supported us – we moved forward relying on his support.

Pechersky never rebuked anybody, he nurtured us. It was probably the main point of his life; he was very unique in this respect and he was greatly successful. Pechersky nurtured the laboratory, the staff working there and their scientific achievements. Eight defended dissertations[2] were his eight awards. Greatly he was gratified with our success! And it was him who gradually, gently, little by little made events move in the desirable way.

Pechersky never ordered and he hardly ever used imperatives. He usually expressed orders in the form of request: "It is necessary to give a lecture...", "Please, go into the preparation of..." It sometimes seemed to us that we were distracted from some important things, but some time later we understood the genuine conception and practical use of Pechersky's messages and instructions.

[1] S. Albu, G.Andrienko, N.Andrienko, A. Andrunakievich, V. Arnaut, V. Burdaev, M. Vekselman, G.Ginkul, S.Gutu, A. Zinchenko, M. Kolesnikov, L. Levchenko, A.Savinov, Y. Secrieru, S. Soloviev, G. Solovieva, V.Sirbu, F.Frolov.

[2] G. Andrienko, N.Andrienko ,V. Arnaut, G.Ginkul, A. Savinov, S. Soloviev, G. Solovieva, V. Sirbu.

Pechersky often started a conversation with the words "Yes, of course..." ,but it was just illusive figure of speech aimed to arise creative thinking of his converser. Frequently his phrase finished unexpectedly. He loved paradoxical word combinations and phrases – "He is in a good sense a bad person...".

Pechersky's incredibly wide circle of social contacts always had complex organization: absolutely different people firmly surrounded his inner center, where one can distinguish three circles. It seemed that if a person got into his field of attraction, he stayed there forever. The first two circles contained his course mates and friends of youth. Pechersky deeply loved them, was in correspondence with them, they sometimes met, participated in the joint projects and successfully accomplished them. From time to time we, the representatives of the third circle, were sent to members of the circle 1 and 2 to Pospelov[3], Zakrevsky[4], Yankovskaya[5] for studying, reporting, implementation.

We did not have problems with business trips at that time, but we lacked computers; we had to ask for computer machine time somewhere else and rely on our personal relations. In fact, our projects depended on relations between people which was a shaky basis for the long lasting projects. Nevertheless, Pechersky easily launched such projects. He really trusted people and liked to say "Human relations are the most important", "The main principal is to be human". Perhaps his strong conviction protected our work and allowed us successfully accomplish it. Of course, in this aspect, our relations with Pechersky outgrew the frames of employer–employee relations, but we never went beyond the bounds.

Proper principles demand sacrifices. We can only surmise how deeply Pechersky suffered when people he had brought up left the laboratory. According to the priorities of human relations you should not feel hurt or offended when you let a person go. He let people go and maintained good relations with them. How did he feel about that?

[3] Dmitry Aleksandrovich Pospelov, Computation Centre of the Academy of Science, Russia, Moscow.

[4] Arcady Dmitrievich Zakrevsky, Institute of Technique Cybernetics of the Academy of Science, Minsk.

[5] Anna Yefimovna Yankovskaya, State University, Tomsk.

In personal contacts Pechersky was always the exemplar of perfection. In a mysterious way, he could convert natural human indignation into irony, which simply made rudeness, swearing, and cursing unnecessary. His speech was completely free from any verbal rubbish. As a highly intellectual person, he did not need it at all.

Pechersky read a lot, wrote well, edited, corrected and even bound books. He systematically read fiction, periodicals, novelties in literature, classics. His particular favorite was science fiction. "You just haven't read good science fiction. You must read Lem."

He took seriously and profoundly scientific literature. Once a week he visited Academic library and thoroughly "percolated" all the novelties on his branch of study. His passion for reading coincided with a fine sense of language and perfect knowledge of grammar. Sometimes his passion for reading and sense of the language got into conflict. "I sometimes read a book, – Pechersky said, – turn over the pages, and have one thing in mind: it's a good page, there are no mistakes."

Pechersky wrote his articles and books in the most inappropriate hours, he used his best time for people: he solved social problems of the postgraduates, organized conferences, managed to get additional financing; he also issued a wall paper, organized chess tournaments and the center of robotics. As a person, he had never been in the solitude.

Pechersky died suddenly and unexpectedly, the funeral took place on February 4th. More than 100 people, his friends and followers gathered in dank and gloomy Kishinev to bid him a final farewell. A lot of people did not come to the ceremony that day, because they did not know about his death, or because they had dispersed all over the world. But no matter where we were on that sad Tuesday, it is our duty to forever cherish our students in memory of our

Yuri Nikolayevich Pechersky, Kishinev, “Doina”, sector 223.

G.Ginkul[1], S.Soloviev[2]

[1]ZAO Sberbank-Technologii, Moscow, Russia
E-mail: ghincul@hotmail.com
[2]Lomonosov Moscow State University, Moscow, Russia
E-mail: soloviev@glossary.ru

In memoriam, Yu. N. Pechersky

Vladimir Ivanovich Levchenko

My direct collaboration with Yuri Nikolayevich Pechersky began in 1983. At that time he was the head of the Laboratory of Mathematical Cybernetics at the Institute of Mathematics. Earlier I was doing pure theoretical research in the Department of Functional Analysis and practically was not familiar with applied research. Thus, the transfer to Yu. N. Pechersky was for me, in a way, a turning point as I had to learn a new domain from scratch.

We were connected with Yuri Nikolayevich not only officially, but were on pretty close friendly terms. For example, for many years we have to put out together the wall newspaper "Mathematics" that is a long-lost genre now.

In 1983 we, together with Yu. N. Pechersky, became the authors and compilers of the first information publication about the Institute of Mathematics with Computing Centre. The brochure was entitled "Order of the Red Banner of Labor Institute of Mathematics with Computing Centre of the Academy of Sciences of the Moldavian SSR". It provided all research directions being carried out at our institute at that time. This small blue book also contained information about scientific meetings and conferences, organized by the Institute; dissertations defended by the researchers and graduate students of the Institute; the index of brief personal data of the researchers; and even a dozen photos. As I recall, that such a structure of this brochure was proposed by Yu. N. Pechersky.

In 1977 the Institute of Mathematics organized in Chisinau the Soviet-Finnish Symposium "Automated design systems". The Laboratory of Yu. N. Pechersky in the person of a group of researchers

Levchenko V. I., Colesnicov A. E., Maxim M. S., Martynuk S. P., Cotelea V. V., Danilichenko A. F. gave two presentations at this symposium:

1) An Automatized System for Technological Preparation of Plug-In Units Production.

2) On Formation of Technological Sets and Their Dividing to Working Zones.

These presentations were based on works carried out by the Laboratory of Mathematical Cybernetics hired by the Vibropribor Plant and related to technological preparation of the production of printed circuit boards on conveyor.

In the early 1980s the Science Department of the State Planning Committee of the Moldavian SSR, at that time headed by Rotari E. T., expressed the need for a sufficient objective assessment of scientific and technical programs, available then in our republic. Historically, the Laboratory of Yu. N. Pechersky began to work in this research direction. One of the first works of this cycle was a short article entitled "Personnel subsystem and questionnaire of scientific institutions within the DIANA system", published in 1983. In the same way, it was suggested to use, as the most objective, the methods of group expert estimation of scientific and technical programs. The group of researchers, in many respects under the leadership of Yu. N. Pechersky, has published a number of papers directed towards the developing automated systems for acquisition, analysis and processing of expert information on complex objects and, in particular, on scientific and technical programs and research teams. The result was the DIANA system – Automated system for management of science and technology development, already mentioned above. This system was implemented on computing equipment available then, namely on personal computers ISKRA-226.

As an exotic episode I can recall a short-time contact of Yu. N. Pechersky with medicine. In 1990 the abstract entitled "Automated dialogue systems for alcoholism screening" (co-authors: Hotineanu M.

A., Levchenko V. I.) was published at the All-Union Conference on Narcology, held in Dushanbe.

In the early 1990s the Laboratory of Yu. N. Pechersky maintained sufficiently fruitful contacts with the Institute of Engineering Cybernetics of the Academy of Sciences of Belarus, in particular, with the Corresponding Member of the Academy of Sciences of Belarus A. D. Zakrevsky. One of the achievements of this collaboration was "Instrumental expert diagnostic system EDIP, based on finite predicates" (Zakrevskij A. D., Pecherskij Yu. N., Levchenko V. I.). As a continuation of this research direction, the researchers of the laboratory published papers on matrix representations of fuzzy knowledge using predicates, which has applicability in expert systems of diagnostic type. Among the publications of this series the following articles may be mentioned: "Matrix representation of fuzzy knowledge in the expert system shell EDIP", published in Romania in 1991, and "The representation of fuzzy knowledge", published in Japan in 1992. Based on an expert system with fuzzy knowledge base the demo was developed in the Laboratory, showing how one can control real systems using fuzzy instructions. As the object of control the so-called inverted pendulum was taken, that is, a pivot hingedly fixed with the lower end to a cart that can move to left/right. The aim of the control system was not to allow the pivot to fall, by moving the cart to the left or to the right. There is no problem with building such a system, using the equation of fall of the pivot. The situation is complicated, if the random forces are applied to the pivot. However, if you look how such a problem is solved by a human, holding a stick on his finger and balancing easy enough without letting it fall down, then we can see that it is not about to solve any equations. The rule that guides the human in this case is simple – "Move in the direction where it falls! If it falls fast, move fast!" In the developed expert system the knowledge base consisted of such instructions, which were not many. At the same demo it could be seen how the quality of control changes, if some of rules are removed from the knowledge base.

By this time, the Laboratory headed by Yu. N. Pechersky has been already renamed into the Laboratory of Artificial Intelligence Systems.

At that time in the Laboratory two groups, concerned with similar directions, but having different approaches, became formed. What both groups had in common was that they addressed research directions related to expert systems, which are a part of artificial intelligence.

At that time the Laboratory of Yu. N. Pechersky organized small but very useful symposia on expert systems, usually held in the Vadu lui Voda village in spring or early summer. The Laboratory maintained good contacts with Computing Centre of the Academy of Sciences of USSR, in particular, with the Laboratory of Artificial Intelligence, headed then by D. A. Pospelov. Also there were bilateral relations with the Institute of Control Sciences of the Academy of Sciences of USSR.

Speaking of Yuri Nikolayevich Pechersky as a scientist, it should be mentioned that he did a great work on science popularization. Having a literary talent, he has written several popular science books on themes close to the research of his Laboratory. Pechersky's writing is accessible and at the same time very scientifically accurate that undoubtedly sets his books apart from the popular literature of this direction.

Communication with Yuri Nikolayevich both in official and everyday life has always given a great satisfaction. Those who knew him, probably, first of all remember his perfect sense of humor, erudition and tactfulness. For a long time I had to share with him the office – we have worked at the tables standing side by side. I recall that once coming to the work early in the morning, I have found him doing something inappropriate for a head of laboratory – he was carefully washing and polishing the window glass. Answering my puzzled unspoken question, he replied unexpectedly and shortly: "I cannot fly away on professional business trip. Every time I look at the window, there is always non-flying weather!"

V. I. Levchenko

July 10, 2013

Proceedings of the International Conference on Intelligent Information Systems
IIS'2013, August 20-23, 2013, Chisinau, Republic of Moldova

My teacher and friend Yu. N. Pechersky

Xenia Naidenova

"And I am working, and I am being carried away,
And I am writing about robots…"

The first time I got to know Yuri Nikolayevich Pechersky in 1977. It was when me, Tanya Cheboksarova, and our teacher Nikolay Georgievich Boldyrev, all three form Leningrad, attended the First All-Union School-Seminar "Logico-combinatorial methods in pattern recognition". This school was held in Moldova on October 2-6 in the Vadu lui Voda village at the "Nistru" recreation centre. I will never forget this school.

Figure 1. Kiev, 1984, Yu. N. Pechersky and X. Naidenova

First, this school has been a significant scientific event. The fact is that in those years, pattern recognition was addicted to probabilistic methods. In logico-combinatorial methods, a connexion between

automation of discrete automata and modeling of cognitive functions, such as inductive inference of logical decision rules and their application to the description of objects, classes, and pattern recognition, was outlined. In essence, it was the beginning of modern powerful scientific domains, such as Machine Learning and Data Mining.

Second, the wondrous atmosphere had reigned at this school – interesting and profound reports, free scientific discussions in the open air, informal friendly intercourse of scientists. I remember that it was just this very school, where D. A. Pospelov within the cultural and educational program gave us a stunning lecture "Games People Play" by Eric Berne's book with the same title. This book was translated into Russian only many years later, in 1992.

It seems to me that there was not a single person at this school who would not fall in love with somebody or would not found a lifelong friend. The soul and the main organizer of this stunning scientific event was, of course, Yuri Nikolayevich Pechersky. Outwardly he was extremely modest, laconic, but also extremely responsive and attentive to people. He did not try to stand out somehow, to draw attention to himself. But, nevertheless, he did it, because he was carried away by scientific problems, he responded enthusiastically to the ideas of his colleagues and students, was ready to support all of them.

I do not remember how, but it seems to me that our friendship had begun from that memorable school. Already in the next year, 1978, Yuri Nikolayevich Pechersky invited me to visit him on summer holidays with my little son, also Yura. It was a very big help for me.

I had not yet defended my dissertation, I earned not enough money, and my vacation was short, so summer was the most difficult time for me in the sense of providing the child with a normal vacation.

I remember gratefully how we lived as Yura Pechersky's guests, at that time in still a pretty happy family situation. There was everything: tourist mini-hikes with campfire and tent, stay-at-home holidays with welcome of guests, and even home seminars. Yuri Nikolayevich, doing any daily household tasks (and sometimes there was a lot of them – cleaning, cooking, washing, garden, and dog Kent), was wholly occupied with a scientific problem. Not by chance I have written in the epigraph the words "And I am working, and I am carried away...", because Yuri

Nikolayevich was just full of passion in work. He was easily carried away, and being carried away, tried to get involved in the problem as deep as possible, sometimes even seemed otherworldly to a certain extent. And, that's important, he was worried very much about the result, he needed to do any work flawlessly, and if some circumstances prevented it, this led him to almost physically suffer.

Also in other, very difficult years of his family life, Yuri Nikolayevich remembered about my summer problems and offered me his assistance, for example, to come for vacation in Sevastopol through his friends from Tomsk. Though, I did not go to Sevastopol.

But I used every moment of meetings with Yuri Nikolayevich, to learn, learn, and learn, read his books (in manuscript), asked questions.

The years from 1978 to 1986 were very fruitful for Yuri Nikolayevich as a writer. In 1978 his book "Theoretical and graphical methods in pattern recognition" («Теоретико-графические методы в распознавании образов», in Russian) was published. Methods of graph theory in pattern recognition and their application in paleontology - this is the direction established by Yuri Nikolayevich. In the same year the books "Robots among people" («Роботы среди людей», in Russian), published in 1979, and "On the path to artificial intelligence" («На пути к искусственному интеллекту», in Russian), published in 1981, were drafted. This last book appeared in print a year before the book "Fantasy or Science?" («Фантазия или наука?», in Russian) by D. A. Pospelov with the same subtitle "On the path to artificial intelligence", and possibly anticipated the latter.

Re-reading these books now, one is surprised at how exactly and deeply there are touched the social, philosophical and moral problems, which are posed for human society by the development of intelligent artificial systems and robots. There are still no solutions for these problems, and the questions only multiply. I like very much the conciseness of many definitions to which Yuri Nikolayevich comes. For example, he suggests considering that in artificial intelligence the subject of research is intelligent systems, regardless of their physical nature or, for example, the distinction of two types of dialogue - equal and unequal. An example of the latter is human-computer communication.

Erudition, mastery of popular exposition (but without loss of complexity), philosophical comprehension of problems are inherent in the author's writing style of Yuri Nikolayevich Pechersky.

And the recognition has come. That is what Yuri Nikolayevich himself wrote to me in one of his letters: “To my surprise, the book about robots has been met well. The whole edition has been sold in 5 days. Leonya Sushchenya (good man!) has sent an inconveniently enthusiastic review. Our president and my director have spoken well too. The final blow has been delivered by the Copyright Committee, through which Czechoslovakia asked for my consent to translate and publish the book in their country. Hoo-ee! I have agreed.”

In 1986 his book “Pattern recognition systems in nature, science and engineering” («Распознающие системы в природе, науке и технике», in Russian) was published. This book is the most considerable work of Yuri Nikolayevich. I don’t know if this book is being republished, but it is still not outdated.

Here I do not list all the books written by Yuri Nikolayevich. It would be well to create a full list of his works. Maybe such a list already exists? But he could have more monographs and books, nevertheless. Why it did not happen?

In the Laboratory of Mathematical Problems of Management, Yuri Nikolayevich led a whole scientific direction aimed at elaboration of computer-aided data processing systems. In particular, there was set a task to create a computer-aided dialogue system for management of science and technology development in the Republic of Moldova. No more and no less. Maybe the best expert system in the Soviet Union, FIAKR, was developed just in Moldova, under the leadership of Yuri Nikolayevich. All the powers Yuri Nikolayevich gave to the institutional work, managing staff of his group. He was worried about their work as about his own. Organization of numerous conferences, meetings, performing scientific-research work, reporting, and the like, prevented, I think, the possibility to give all his powers to his scientific interests. As far as I understand, he never refused to help someone, any sort of work he performed with the maximum efficiency. Once he wrote to me:

"... But I especially liked frankly saucy plans to put point-blank the question about work in the organization. Well done! SCIENCE CAN NOT BE SERVED BY HALVES and the others are not an example here."

Besides the huge pressure of organizational work, he was affected by the lack of real understanding, encouragement and immensely difficult circumstances of his family life.

Any sort of work in his life Yuri Nikolayevich performed with the maximum efficiency. Here is his story about how he had cured his dog Kent:

"We, Kent and I, remained alone until about August 25, and not in the best time of his life. The dog fell ill hard before Alla's departure. He busied himself with plucking out the hair and wounding everywhere his teeth could get. The likely reason is that the mistress every day while walking had washed him in the sewage from the meat-packing plant. I delivered a massive return blow in the form of high-calorie diet (like in a restaurant) and insertion into the dog of very possible vitamins in the form of vegetables and powders. I neglected my work for a week and walked with Kent during 4-6 hours a day. Now he is coming to himself. "

Drawings of Yuri Nikolayevich convey pretty well his subtle emotional traits, his inside lyricism. Unfortunately, I have only few of them. Two "quick" pencil sketches you can see in the following illustrations (Fig. 2, Fig. 3).

Figure 2. Field flowers

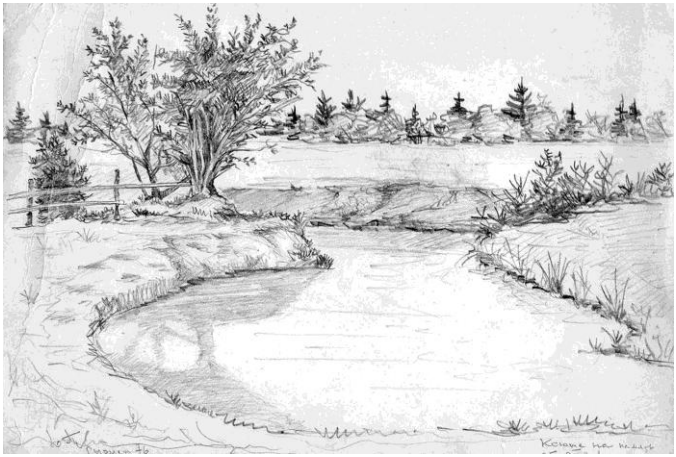


Figure 3. A small lake

But the enthusiasm of Yuri Nikolayevich was manifested in sports also. In May 1980 he wrote to me:

"I am okay in sports. In May I travelled to Riga, to the All-Union Academiada on tennis. It was hard, because in my class I was playing with three Masters of Sport. I was the last but one. Nevertheless, in this season I, probably, will get the First-Class. However, it is difficult with trainings since time constraints. Only consolation - I physically exhaust the adversaries who are 15-17 years younger."

Yuri Nikolayevich was the perfect, responsive, charming, deeply honest person, and a real scientist. He had an outstanding talent as teacher. His students and colleagues will better tell about this. But he was completely defenseless and unarmed in front of betrayal and human baseness. He only spread his hands and said: "Well, how can that be?", and incredibly suffered under it. I am very sad that he is not with us today.

Xenia Naidenova

Military Medical Academy, Saint-Petersburg, Russian Federation
E-mail: ksennaidd@gmail.com

April 20, 2013

Proceedings of the International Conference on Intelligent Information Systems
IIS2013, August 20-23, 2013, Chisinau, Republic of Moldova

Yuri Pechersky, the outstanding scientist and organizer of All-Union workshop COLOM, a remarkable person and friend

Anna Yankovskaya

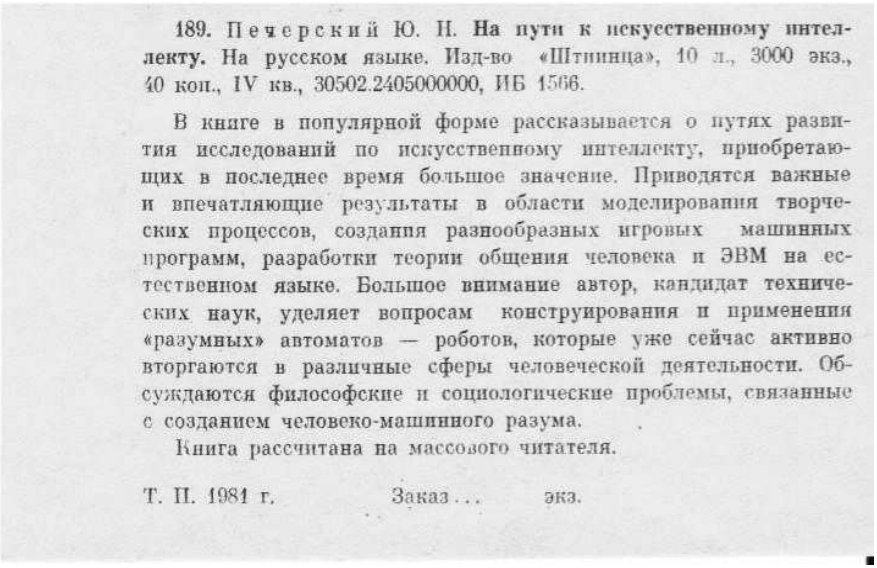
Many schools people founded,
But by reason alone
We felt really confounded
Till we opened COLOM!
Yes, we needed it badly,
For it can't be denied
That to do something ably
Isn't easy inside.
Just to slave away inside
Will not make you grow wise.
So we saw the new days in
Outside, at sunrise.
Leaving sometimes the hubbub
And the dust of the town
We escaped to the forest
And there touched the ground.
May the wind swing us softly
On the Dniester blue waves,
And the papers on posters
From green trees show new ways.
May the sun pour its sunshine
Or the rain drop from clouds,
But the friendship of young days
Forever abounds!
And from "classes" to "classes"
Of this wonderful school
We remember its fun things
And the serious too.

I have been acquainted with Yuri Pechersky since student days, when I studied at the Radiophysical Department (RPD) of Tomsk State University (TSU). Yuri graduated from this department in 1958 and worked at the basic research laboratory of calculating devices of TSU Siberian University of Physics and Technology. This laboratory was a base for students specializing in computers. He supervised graduate theses of RPD students and taught students to solve chess problems (end game) on the Central Computer "URAL". As a research employee of the laboratory, Yuri took a position of the secretary of the Scientific Seminar chaired by our teacher Arkadij Zakrevskij. In 1961 Yuri Pechersky moved to Yerevan and soon afterwards to Chisinau where he worked as head of the Laboratory of Intelligent Systems at the Institute of Mathematics and Informatics of the Academy of Sciences of the Moldavian SSR, later the Republic of Moldova.

His publications in Russian and international journals both in Russian and English testify to fact of Pecherskys recognition as an outstanding scientist. Books by Pechersky, such as "Theoretical and Graphical Methods in Pattern Recognition" (1978), "Robots among People" (1979), "Designing Transmitting Processor" (1980), "On the Way to Artificial Intelligence" (1981), "Models of Robotic Production" (1985, with co-authors), "Recognition Systems in Nature, Science and Technique" (1986), "System and Theoretical Programming" (1987), "Optimization and Data Processing" (1987), "Intelligent Systems" (1988), "Sketches on Computers" (1989), "Mathematical Modeling and Optimization" (1989) etc. are highly interesting in scientific terms. The publication "Yuri Pechersky is My Teacher and Friend" by Xenia Naidenova speaks about a high academic level of these books.

I would like to note that Yuri Pechersky was a perfect scientific expositor of science. He contributed much to the creation of the website topic "Chronicle of Events".

Pechersky used to announce about book releases to his friends by sending them information cards. An example of these cards sent to me is shown in Fig. 1. It should be noted that the edition of these books was usually large. For example, the edition of the book presented in



189. Печерский Ю. Н. **На пути к искусственному интеллекту.** На русском языке. Изд-во «Штиинца», 10 л., 3000 экз., 40 коп., IV кв., 30502.2405000000, ИБ 1566.

В книге в популярной форме рассказывается о путях развития исследований по искусственному интеллекту, приобретающих в последнее время большое значение. Приводятся важные и впечатляющие результаты в области моделирования творческих процессов, создания разнообразных игровых машинных программ, разработки теории общения человека и ЭВМ на естественном языке. Большое внимание автор, кандидат технических наук, уделяет вопросам конструирования и применения «разумных» автоматов — роботов, которые уже сейчас активно вторгаются в различные сферы человеческой деятельности. Обсуждаются философские и социологические проблемы, связанные с созданием человеко-машинного разума.

Книга рассчитана на массового читателя.

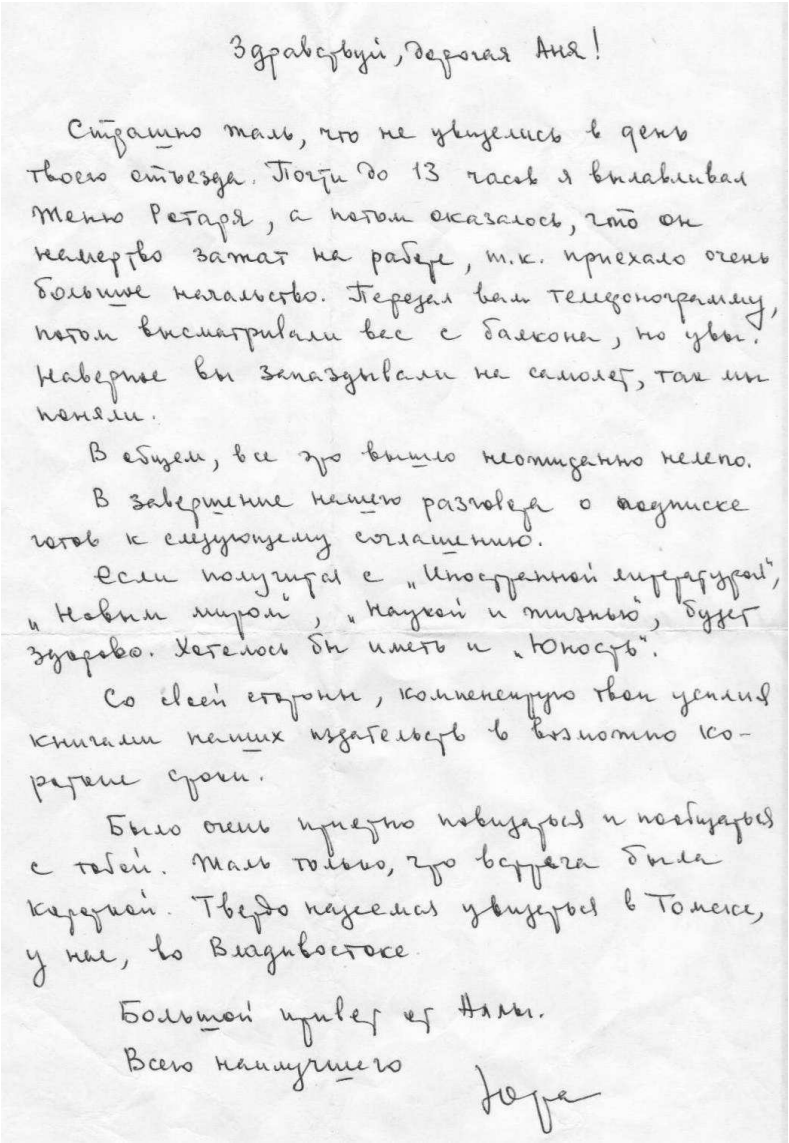
Т. П. 1981 г. Заказ... экз.

Figure 1. Information card from Yuri Pechersky.

Fig. 1 was 3000 copies.

Yuri often presented and sent books and journals to me in Tomsk written not only by him but also interesting scientific books and journals of other authors. A striking illustration of the books and journals exchange is the following letter from Yuri shown in Fig. 2.

For me, Yuri has always been a great friend. It was often that during the excursions that he organized I lost my handbag, and he always found it. One day when I was traveling to school by bus from the airport to Chisinau, the amber from one of my earrings fell and was lost. That amber was very dear to me. I had to remove the second earring. When I told Yuri about it, he reassured me that it would be found. And when I returned back to the airport by the same flight, a woman who sat next to me, said, 'What a pity that someone has lost this amber', and showed it to me. I immediately pulled out from my handbag two earrings, one of which had no amber. We both were happy. Upon my arrival I called Yuri and thanked him for the words



Здравствуй, дорогая Аня!

Страшно жаль, что не увидемся в день твоего отъезда. Почти до 13 часов я выкликивал Женю Ретаря, а потом оказалось, что он намертво застат на работе, т.к. приехало очень большое начальство. Передал вам телеграмму, потом высматривали вас с балкона, но увы! Наверное вы запаздывали на самолет, так мы поняли.

В общем, все это вышло неожиданно нелепо.

В завершение нашего разговора о подписке готов к следующему соглашению.

Если получится с "Иностранной литературой", "Новым миром", "Наукой и жизнью", будет здорово. Хотелось бы иметь и "Юность".

Со своей стороны, компенсирую твои усилия книгами наших издательств в возможно короткие сроки.

Было очень приятно повидаться и пообщаться с тобой. Жаль только, что встреча была короткой. Твердо надеемся увидеться в Томске, у нас, во Владивостоке.

Большой привет от Ани.

Всего наилучшего

Юра

Figure 2. A letter from Yuri.

that he reassured me with.

I knew that Yuri was a talented painter, and I admired his creativity. He told me that he started drawing only at the retirement age, and I hoped that when I reached that age, I would also begin to paint. Unfortunately, it has not come true.

At the COLOM (combination logical methods) school I have got a lot of real friends, namely: T. Cheboksarova, X. Naidenova, G. Korneva, who are close to me in spirit and in terms of scientific activity. Especially, I remember the night stories by G. Korneva about Marina Tsvetaeva, whose biography she knew very deeply. Those stories fascinated me, and we could not sleep all night.

Many All-Union and international conferences and international workshops on combination logical methods in recognition problems conducted under the leadership of Pechersky were organized by the Institute of Mathematics together with the Computational Center AN MSSR, the Moldovan Republic Regional Department 'Artificial Intelligence' of the System Analysis Committee of the Presidium of the USSR Academy of Sciences. The information about the first seminar held between 2 and 6 October 1977 in Vadu-lui-Vode (Moldova SSR), was published in the booklet Information Materials: Cybernetics 4 (104), 1987, Moscow, released by the Scientific Council on Complex Problem Cybernetics of the USSR Academy of Sciences.

Dmitry Pospelov, the outstanding scientist in the field of artificial intelligence, has played a great part in the creation of the COLOM School. He made stunning reports not only in the area of this school but also in other subjects. For example, he gave a lecture based on the book by Eric Berne 'Games that People Play'. Dmitry Pospelov was a connoisseur of art and gave a lecture with a slide show based on the compositions of Amaravella. It should be noted that one-fifth of the compositions was donated to D.A. Pospelov from Peter Fateev (Amaravellas member).

Representatives from dozens of research institutions and universities in Moscow, Vladivostok, Kazan, Kiev, Leningrad, Minsk, Novosibirsk, Riga, Tomsk, Chisinau etc. participated in the workshops.

Among the active speakers, the reports should be noted by D.A. Pospelov (Moscow), A.D. Zakrevskij (Minsk), Yu.N. Pechersky and A.V. Karelina (Chisinau), L.A. Rastrigina (Riga), N.G. Bondariva, X.A. Naidenova, T.N. Cheboksarova, G.N. Korneva (Leningrad), A.E. Yankovskaya (Tomsk), V.V. Nikolaev, Yu.S. Moskalenko (Vladivostok), V.V. Aleksandrova, N.D. Gorsky, A.O. Polyakov (Leningrad), V.I. Gorodetsky (Leningrad), E.V. Mazur (Kazan), N.D. Vashenko (Kiev). It should be noted that the geography of the students attending the workshops presented the best cities of the USSR.

Discussions about the prospects of the development of the combination logical approach to solving recognition problems were held at each workshop. Decision-making recommendations provided theoretical and practical results in the area of forecasting, recognition, and diagnostics of objects using combination logical methods. Thus, at the first workshop the recommendations highlighted the necessity of the wider research conducted into mathematical apparatus, discrete devices theory, coding theory, graph theory, discrete mathematics.

It is worth noting that some workshop meetings were held not only on the premises of the camping site "Nistru" but also on the beach of the river Nistru, where particularly the problem of COLOM pattern recognition was disclosed rather intensively. Many useful reports were made. Particularly, students were interested in information about the breakfast, lunch and dinner time. Despite the blinding light and the scorching sun the listeners did not spare the speakers and asked tricky questions. When there was a lack of posters that were hung on branches, the speaker drew in the air, and students recognize a particular image, and this was somewhat exciting. Nevertheless, all students passed the tasting exam.

Not only the COLOM workshop and other conferences were organized by Yuri, but he also helped his friends implement the results of their research. For example, systems "Abiturient-T" and "Student" (Yu. Komarov, A. Yankovskaya) developed and operated in TECI (TSUAB), were put into operation in two or three schools of Chisinau with a help from Yuri Pechersky.

It is difficult to enumerate all that Yuri Pechersky has done.

Yuri trusted people very much. Sometimes it landed him in trouble, but still his trusting people helped him in difficult situations. In the 2000s, Yuri was seriously ill. He asked me to help financially in connection with the operation. The difficulty was that the dollars – not rubles – were needed, and one had to deliver them to Chisinau as soon as possible. A lot of ways had been considered so as to ensure a money transaction from Moscow, Minsk, and other cities. But Yuri found the way to transact dollars. It was not only me who sent the money, but also his friends whom I addressed. Thus, people who helped were V.P. Tarasenko, F.P. Tarasenko and his childhood friend L.S. Lyahovich, and others. Unfortunately, we could not save Yuri.

I am sure that Yuri's friends will also write about him, namely: A.D. Zakrevskij, A.A. Utkin, X.A. Naidenova, and many others.

Yuri will remain in our hearts and memories forever.

Anna Yankovskaya

National Research Tomsk State University; Tomsk State University of Architecture and Building
E–mail: ayyankov@gmail.com

AUTHORS INDEX

Table of contents

Part 1. Invited papers

Part 2. Regular papers

Part 3. Students' papers

Part 4. Dedicated to Iu.N.Pecherskiy