From Fragments of Knowledge Towards a Bigger Picture: How Can the Process be Supported

Tanja Urbančič

University of Nova Gorica, Vipavska 13, 5000 Nova Gorica, Slovenia and the Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia, tanja.urbancic@p-ng.si

In this paper we deal with the problem of information that is dispersed and growing so fast that it is difficult to connect it together into a coherent picture as needed for complex problem solving. We present two examples and some methods that have potential to contribute towards putting pieces of knowledge together. The first consists of finding complementary pieces of knowledge in literature that supports hypothesis generation by a well-defined and computer supported method. The second one is sharing and upgrading knowledge in collaborative settings, which still has many non-technical issues to be solved, although well developed in its technical aspects.

Key words: Knowledge management, education, data mining, networked organizations

Kako od posameznih delov do celovitejše slike znanja

V članku je obravnavan problem hitrega naraščanja in raztresenosti informacij, kar otežkoča njihovo povezovanje v smiselno celoto, potrebno za reševanje zahtevnih problemov. Predstavljamo dva primera, ki prinašata obete v tej smeri. Prvi je iskanje komplementarnih kosov znanja iz literature, kar podpira generiranje novih hipotez z dobro definirano računalniško metodo. Drugi je izmenjava in nadgrajevanje znanja v sodelovalnih okoljih, kar je tehnično sicer že zelo dobro podprto, vendar pa prinaša s sabo še veliko nerešenih, predvsem netehničnih vprašanj.

Ključne besede: Upravljanje znanja, izobraževanje, rudarjenje podatkov, mrežne organizacije

1 Introduction

To cope with the increasingly complex problems of our society, it is important to take into account many different aspects and connect them into a bigger picture. This often requires an interdisciplinary approach and a lot of collaboration across different boundaries. Vast amounts of data and knowledge are available nowadays, but sometimes it seems that its value is not fully recognized and it are not used efficiently, mainly due to the fact that it is dispersed and cannot easily be connected due to quantity and diversity. Knowledge technologies provide tools and techniques that can help in overcoming these problems. Among them, data mining and decision support deserve attention due to a wide range of successful applications in different areas (Mladenić et al., 2003).

In the field of education, the importance of these trends has been recognized and changes supporting the process have begun. One of the important issues is the in-corporation of knowledge technologies into study programs. A review paper (Urbančič et al., 2002) analyzed data mining and decision support education based on investigations of the materials available on the world-wide web. They present several complete postgraduate programs with an emphasis on data mining and decision support. Also important for the spread of such methods is the incorporation of these topics into study programs for non-computer science students. Among them, several interesting examples combine technical content with various areas such as medicine, e-commerce and bioinformatics. In interdisciplinary study programs, students not only get information and knowledge from different disciplines, but also learn how to combine it efficiently for problem-solving in different professional fields. The interaction of technical, economical, social, environmental and other aspects must be understood and different pieces of knowledge must be combined into a meaningful view of the whole. Here, the approach of using well defined methods

for well structured problems usually doesn't work. As is described nicely in (Burns and Jordan, 2006), one of the most difficult things to teach is the defining of the problem itself. The paper also discusses alternative ways that the capability of seeing the whole can be developed. A similar problem also appears within disciplines that seem to be coherent from the outside, but are still divided in separate subfields that too often don't have enough intersection and communication. One such example is medicine. For example, autism is investigated in the framework of behavioural psychology, genetics, biochemistry, brain anatomy and physiology, but there is a lack of studies that would connect these diverse findings into a coherent picture (Belmonte et al., 2004).

Another aspect of putting fragmented knowledge together is combining knowledge and information available at different locations through collaborative settings. This brings new challenges, the non-technical being often more limiting than technical ones. Therefore it is not surprising that training in communication, networking and teamwork is also mentioned explicitly in the list of necessary improvements given in the proposal of the European Commission on how to modernize Europe's universities (Europa Press Release, 2006). At the same time, the educational institutions themselves have become very active in thematic networks where all these issues come across at a different level, but in a similar way.

In this paper, we deal with both aspects. First we present literature mining as a method that can provide very useful technical support when uncovering hidden connections in bibliographic databases. As such, it is a simple but powerful tool that helps in establishing bridges between disciplines and different professional communities. Then we focus on the process of complementing knowledge in a collaborative setting. Although at the moment the first aspect is mainly associated with the discovery of knowledge in science and the second with professional collaborative work settings such as virtual enterprises, we strongly believe that they are relevant for education in different ways - as already indicated above. Besides the more direct implications, we must also think about how future generations will be educated and trained in order to deal successfully with these issues.

2 Finding Complementary Knowledge in literature

The amount of information available on-line is growing with enormous speed. A good example is Pubmed, the United States National Library of Medicine's bibliographic database, which covers more than 15 million citations and increases by more than 1,500 complete references a day. It is obvious that no human expert could manage this stream of new information without the suitable support of a computer.

Since the expert fields are getting more and more specialized, scientists and other professionals tend to function in more or less "closed" sets of specialized pro-

fessional literature, in general without many cross-references to other research or professional communities. On the other hand, the problems of today's society are becoming more and more complex. Many phenomena, such as complicated disorders or diseases, can only be understood when different partial findings are combined and all support for this process is very welcome.

A simple, but extremely powerful method was proposed by Swanson (1990). If phenomenon C is to be explained and if there is a hypothesis that C is connected with agent A, then C may be from one field of expertise and A from another. In this case, literature about A and literature about C very often don't have any intersection. Swanson suggested finding a bridging term, B, which can be found in the literature on both A and C. If a closer look at such appearances in literature shows that A causes B and B influences C, this might support the hypothesis of A influencing C.

In one of his examples, Swanson was interested in the hypothesis that magnesium deficiency can cause migraine headaches. In this case, migraine played the role of C and Magnesium played the role of A. At that time, there were 38,000 articles about magnesium and 4,600 articles about migraine in Pubmed, with no direct evidence of any connections between the two. However, Swanson found several bridging terms B and more than 60 pairs of articles connected A to C via the terms of B. For example, in magnesium literature there is a statement that magnesium is a natural calcium channel blocker. On the other hand, in migraine literature there is the statement that calcium channel blockers can prevent migraine attacks. (the Calcium channel blocker plays role of B in this case.) Similarly, he connected the facts that stress and type A behaviour can lead to loss of magnesium in the body, while stress and type A behaviour are also associated with migraine. Also, in the magnesium literature, he found that magnesium has anti-inflammatory properties and, in the migraine literature, that migraine may involve the sterile inflammation of cerebral blood vessels. In this way, he found 11 pairs of documents that were, when put together, suggestive of and supportive for a hypothesis that magnesium deficiency may cause migraine headaches.

Swanson tested his method on different problems and actually found hypotheses, unknown at that time, which were later confirmed by clinical trials. His applications include connecting fish oil and Raynaud's syndrome, anticipating adverse drug reactions, etc.

We believe that this method provides a very valuable base for knowledge discovery in huge textual databases. In our paper (Urbančič et al., 2007), we mention several researchers who followed his idea and applied it to different problems. In the same paper, we also upgraded his method by proposing a new way for selecting hypotheses. The question we wanted to answer was: Being interested in phenomenon C, how do we find a candidate agent A as a potential cause of C? In other words, being interested in migraine, why did Swanson focus on magnesium and not on something else? Swanson is not very specific about this choice in his paper and he comments that success de-

pends entirely on the knowledge and ingenuity of the searcher. In our method, to explain phenomenon C, we are looking for interesting rare terms in literature about C. Then we inspect literature about these rare terms and look to see if they have any intersections via joint terms. From these joint terms, candidates for term A are selected and a hypothesis that A influences C is tested using Swanson's ABC model. We tested our method (called RaJo-Link) in the area of autism and found previously unknown relations that were evaluated by a medical expert as interesting and promising, helpful towards a better understanding of this complex phenomenon.

3   Sharing and Upgrading Knowledge in Collaborative Settings

New media and computer networks enable business, medicine, science, etc. to be done in a collaborative setting without geographical boarders - resulting in eBusiness, e-Medicine and eScience. Due to this development, networked organizations are becoming increasingly important. Their activities are facilitated by the use of shared infrastructure and standards, decreasing risk and costs. A virtual enterprise is a specific form of networked organization (Camarinha-Matos and Afsarmanesh, 2003) in which a group of organizations or individuals voluntarily join to share their knowledge and resources in order to better respond to a particular business opportunity through collaborative work, supported by information and communication technologies. This greatly increases the possibilities of choice since one can select their co-workers across organizational and geographical boarders, having their competences to accomplish a task at the front of the mind. The networked organization is as strong as it is capable to use potential of all its members and combine it into a successful way. As knowledge is one of the most important assets of a network, the knowledge of the members should be shared and combined in order to enable the successful functioning of the network as a whole.

A specific technical infrastructure is needed to support networked organizations in their activities. One of the important issues to be covered by this infrastructure is collecting dispersed information from partners, storing it in a consistent, understandable, computationally accessible and flexible way and making it available to the partners of the network and to the external audience. This functionality is available through web systems and can be achieved successfully using available techniques and tools (Jorge et al., 2003). So, in principle, could we now work successfully with anybody in the world?

Things are not that simple. Kling and Lamb (2000) pointed out that interorganizational computer networks are also social networks where relationships are complex, dynamic, negotiated and interdependent. They claim that the organizational changes required when "going digital" are often neglected and refer to them as the "hidden costs of computing".

We experienced this in a virtual enterprise SolEuNet (*http://soleunet.ijs.si*), where 12 academic and business partners from 7 European countries joined forces with the aim of offering their data mining and decision support expertise to the European market. Collaborative work by geographically dispersed teams had a well established Internet support and infrastructure (Jorge et al, 2003), the participants were professional experts and really devoted to the project. However, since the engineering side of the project did not have a suitable counterpart in the organizational aspects, the organizational model evolved through different stages mainly on the basis of lessons learned during the project - including the "discovery" of the danger of information asymmetries, the importance of the IPR issues and the key role of building trust among the partners of the network (Lavrač and Urbančič, 2003). The main direction of these changes was towards increased flexibility as, in the final model, every partner on the network was given the opportunity to be the net broker in particular projects. This resulted in enhanced choice of project partners and consequently in less tensions between them. The need for additional efforts in knowledge management (Smith and Farquhar, 2000) needed to be fulfilled since in such a model, the information and knowledge needed for the role of a net broker had to be organized, stored and maintained in a way that made it accessible to all partners (Jermol et al., 2004).

One of the lessons learned was that trust modelling and management should be a part of knowledge management when establishing and managing a virtual enterprise. A more detailed discussion with examples is provided in (Lavrač et al., 2007).

Partners in a collaborative setting must face the psychological challenge of shifting from a culture of the enterprise and motivation of the individual towards a network culture where sharing knowledge brings advantages, not danger – providing that IPR issues are properly handled. The concept of network intelligence as the capability of going beyond fixed individual identity through dialogue, mutuality and trust (Palmer, 1998) is an unavoidable counterpart to the technological preconditions of networked organizations. Building this kind of intelligence is a long-lasting but very important process, which should be strongly encouraged in society, starting with the education system, which, unfortunately, still strongly prioritises individual competition over cooperation.

5   Conclusion

As knowledge is becoming one of driving forces in our economy and society, it is important to support its development, sharing and use in an efficient way. This is difficult, due to the fact that information is growing so fast and that it is very dispersed. In this paper we presented some technical ideas to bridge these problems, and also pointed out some non-technical issues connected with it.

We also presented our belief that education programs should and could support the process by preparing the

next generations for the use of knowledge technologies and also by educating them for work in the collaborative settings required. Both are unavoidable for solving the complex problems of today and of tomorrow.
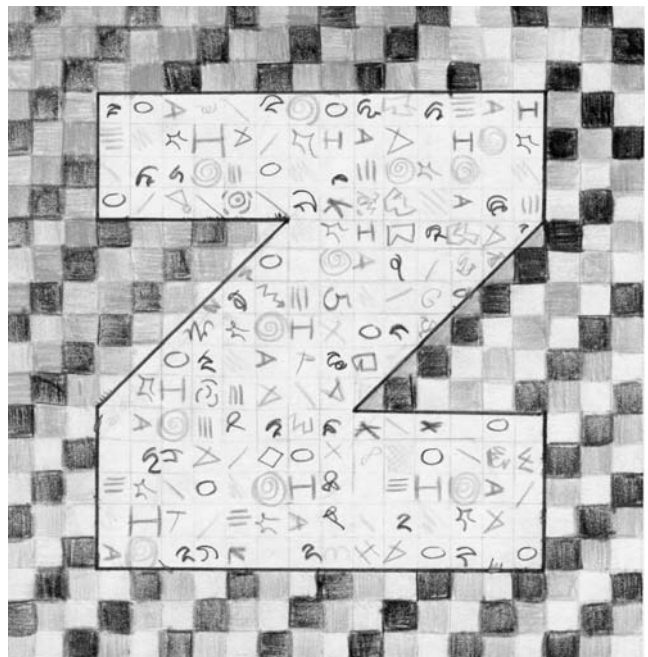
Acknowledgement

References

Belmonte, M.K., Allen, G., Beckel-Mitchener, A., Boulanger, L.M., Carper, R.A. & Webb, S.J. (2004). Autism and abnormal development of brain connectivity, *The Journal of Neuroscience*, **27**(42): 9228-9231.

Burns, A.B. & Jordan, T.M. (2006). Learning to see the (w)holes, *Industry & Higher Education*, **February 2006:**15-18.

Camarinha-Matos, L.M. & Afsarmanesh, H. (2003). Elements of a base VE infrastructure, *Journal of Computers in Industry*, **51**(2), 139-163.

Europa Press Release (2006). Europe's universities must be modernized, says Commission. *Industry & Higher Education*, **June 2006:**148.

Jermol, M., Lavrač, N. & Urbančič, T. (2004). Managing business intelligence in a virtual enterprise: a case study and knowledge management lessons learned, *Journal of intelligent and fuzzy systems*, **14**:121-136.

Jorge, A., Bojadžiev, D., Alves, M.A., Stepankova, O., Mladeni?, D., Palouš, J., Flach, P. & Petrak, J. (2003). Internet support to collaboration: A knowledge management and organizational memory view. In: *Data mining and decision support: integration and collaboration*, Edited by Mladeni?, D., Lavrač, N., Bohanec, M. & Moyle, S., The Kluwer international series in engineering and computer science, SECS 745, Kluwer Academic Publishers, Boston, Dordrecht, London, 247-259.

Kling. R. & Lamb, R. (2000). IT and organizational change in digital economies: A socio-technical approach. In: Kahin, B & Brynjolfsson, E. (eds.), *Understanding the Digital Economy – Data, Tools and Research*, The MIT Press.

Lavrač, N. & Urbančič, T. (2003). Mind the gap: Academia-business partnership models and e-collaboration lessons learned. In: Mladeni?, D., Lavrač, N., Bohanec, M., Moyle, S. (eds), *Data mining and decision support: integration and collaboration*. The Kluwer international series in engineering and computer science, SECS 745, Kluwer Academic Publishers, Boston, Dordrecht, London, 261-269.

Lavrač, N., Ljubič, P., Urbančič, T., Papa, G., Jermol, M. & Bollhalter, S. (2007). Trust modeling for networked organizations using reputation and collaboration estimates, *IEEE trans. syst. man cybern.*, Part C Appl. rev., **37**(3): 429-439.

Mladenić, D., Lavrač, N., Bohanec, M. & Moyle, S. (eds.) (2003). Data mining and decision support: integration and collaboration. In: Mladeni?, D., Lavrač, N., Bohanec, M., Moyle, S. (eds), *Data mining and decision support: integration and collaboration*. The Kluwer international series in engineering and computer science, SECS 745, Kluwer Academic Publishers, Boston, Dordrecht, London.

Palmer, J. (1998). The Human Organization, *Journal of Knowledge Management*, **1**(4) 294-307.

Pubmed: Overview, available from: http://www.ncbi.nlm.nih.gov, (12 January 2007).

Smith, R.G. & Farquhar, A. (2000). The Road Ahead for Knowledge Management: An AI Perspective, *AI Magazine,* **21**(4), 17-40.

Swanson, D.R. (1990). Medical literature as a potential source of new knowledge, *Bulletin of the Medical Library Association*, **78**(1), 29-37.

Urbančič, T., Škrjanc, M. & Flach, P., (2002). Web-based analysis of data mining and decision support, *AI Communications*, **15**:199-204.

Urbančič, T., Štepankova, O. & Lavrač, N. (2006). Enhancing human choice by information technologies, In: J. Berleur et al., (eds.), *Proceedings of the seventh international conference on human choice and computers (HCC7), IFIP TC 9*, Maribor, Slovenia, September 21-23, 2006 (IFIP, International federation for information processing series, 223, New York (NY): Springer, 255-264.

Urbančič, T., Petrič, I., Cestnik, B. & Macedoni-Lukšič, M., (2007). Literature Mining: Towards Better Understanding of Autism. In: Bellazzi, R., Abu-Hanna, A. & Hunter, J. (eds.), *Artificial Intelligence in Medicine LNAI 4594*, Springer, 217-226.

Tanja Urbančič is Dean of the School of Engineering and Management and Head of the Centre for Systems and Information Technologies, both at the University of Nova Gorica. She is also a research fellow at the Department of Knowledge Technologies at the Jožef Stefan Institute. She received her PhD in Computer Science from the University of Ljubljana. Her current professional and research interest is mainly in knowledge management, especially in its applications to education, medicine and public health. She is co-author of several book chapters and journal papers in IEEE Transactions on SCM, Engineering Applications of Artificial Intelligence, the Journal of Intelligent and Fuzzy Systems and others.

Nina Kočar