From Pieces of Knowledge towards a Bigger Picture: How Can the Process be Supported

Tanja Urbančič [1,2]

[1] University of Nova Gorica, Vipavska 13, 5000 Nova Gorica, Slovenia

[2] Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia

E-mail: tanja.urbancic@p-ng.si

Abstract

In this paper we deal with the problem of information that is dispersed and is growing so fast that it is difficult to connect it 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 one is finding complementary pieces of knowledge from literature which supports hypothesis generation by a well-definde computer supported method. The second one is sharing and upgrading knowledge in collaborative settings, which. although well elaborated in technical aspects, still has many non-tecnical issues to be solved.

Keywords: *Knowledge management, education, data mining, networked organizations*

Povzetek

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 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 interdisciplinary approach and a lot of collaboration across different borders. Vast amounts of data and knowledge are available nowadays, but sometimes it seems that their value is not fully recognized and they are not used efficiently, mainly due to the fact that they are dispersed and can not be easily connected due to their quantity and diversity. Knowledge technologies provide tools and techniques that can help in bridging 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 that support the process have started. One of the important issues is incorporation 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 the emphasis on data mining and decision support. What is also important for the spread of such methods, is incorporation of these topics into study programs for non-computer science students. Among them, several interesting examples combine technical contents with various application 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 them efficiently in problem-solving in different professional fields. Interaction of technical, economical, social, environmental and other aspects has to be understood and different pieces of knowledge have to be combined into a meaningful view on the whole. Here, the approach of using well defined methods for well structured problems usually doesn't work. As nicely described 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 in which the capability to see the whole can be developed. Similar problem appears also within disciplines that from outside seem to be coherent, but yet they are devided in separate subfields that too often don't have enough intersection and communication. One such example is medicine. For example, autisms is investigated in the framework of behavioral psychology, genetics, biochemistry, brain anatomy and physiology, but there is lack of studies that would connect diverse findings into a coherent picture [Belmonte et al., 2004].

Another aspect of putting pieces of knowledge together is combining knowledge and information available at different locations through collaborative settings. This brings new challenges, non-technical being often more limiting than technical ones. Therefore it is not surprising, that training in communication, networking and team-working 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, educational institutions themselves became 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 in 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. In the continuation, we focus on the process of complementing knowledge in a collaborative setting. Although at the moment, the first aspect is mainly associated with knowledge discovery in science and the second in professional collaborative worksettings such as virtual enterprises, we strongly believe that they are relevant for education in different ways, as already indicated above. Besides more direct implications, we must also think about how future generations will be educated and trained in order to manage to deal with these issues successfuly.

2. Finding complementary pieces of knowledge from literature

The amount of information available on-line is growing with an enormous speed. A good example is Pubmed, United States National Library of Medicine's bibliographic database that covers more than 15 million citations and increases for more than 1500 complete references daily. It is obvious that no human expert can manage this stream of newcoming information without a suitable computer support.

Since expert fields are getting more and more specialized, scientists and other professionals tend to function in more or less "closed" sets of specialized professional literature with, in general, not many cross-references to other research or professional communities. On the other hand, problems of society of today are becoming more and more complex. Many phenomena, such as complicated disorders or diseases, can only be understood when different partial findings are put together and every support in this process is very welcome due to the problems mentioned above.

A simple, but extremely powerful method was proposed by Swanson [1990]. If a phenomenon C is to be explained and if there is a hypothesis that C is connected with an agent A, C may be from one field

of expertise, and A from another. In this case, very often literature about A and literature about C don't have any intersection. Swanson suggested to find a bridging term B that can be found in literature about A as well as in literature about C. If a closer look at such appearances in literature show 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 C and Magnesium played the role 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 that connected A and C via terms 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 a statement that calcium channel blockers can prevent migraine attacks. (Calcium channel blocker plays role B in this case.) Similarly, he connected the facts that stress and type A behavior can lead to body loss of magnesium, while stress and type A behavior are also associated with migraine. He also found in the magnesium literature that magnesium has anti-inflammatory properties, and in the migraine literature, that migraine may involve sterile inflammation of cerebral blood vessels. In this way, he found 11 pairs of documents that were, when putting 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 to that time that 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 from huge textual databases. In our paper [Urbančič et al., 2007] we mention several researchers that followed his idea and applied tit 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 comments that success depends entirely on the knowledge and ingenuity of the searcher. In our method, to explain a phenomenon C, we are looking for interesting rare terms in literature about C. Then we inspect literatures about these rare terms and look 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 by using Swanson's ABC model. We tested our method (called RaJoLink) in the autism domain and found previously unknown relations that were evaluated by a medical expert as interesting and promising on their way towards 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 with no geographical boarders, resulting in eBusiness, eMedicine, 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 specific form of networked organizations in a virtual enterprise [Camarinha-Matos and Afsarmanesh, 2003] in which a group of organizations or individuals join voluntarily 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. Having such a possibility greatly increases possibilities of choice since one can choose their coworkers across organizational and geographical boarders, having in mind in the first place their competences to accomplish a task. 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, also knowledge of members should be shared and combined in order to enable successful functioning of the network as a whole.

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 make it available to the partners of the network as well as to the external audience. This functionality is obtained through web systems and can be successfully achieved on the basis of available techniques and tools (Jorge et al., 2003). So, in principle, could we now work successfully with anybody in the world?

Things are not so simple. Kling and Lamb [2000] draw attention to the fact that interorganizational computer networks are also social networks in which relationships are complex, dynamic, negotiated and interdependent. They claim that needed organizational changes when "going digital" are often neglected and refer to them as to "hidden costs of computing".

We experienced this in a virtual enterprise SolEuNet (*http://soleunet.ijs.si*) in which 12 academic and business partners from 7 European countries joined their forces with the aim of offering their data mining and decision support expertise to the European market. Collaborative work of geographically dispersed teams had very well established Internet support and infrastructure [Jorge et al, 2003], the participants were professional experts and really devoted to the project. However, since engineering side of the project did not have a suitable counterpart at the organizational side, the organizational model evolved through different stages mainly on the basis of lessons learned during the project, including "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 more flexibility, as in the final model, every partner of the network was given the opportunity to be the net broker in particular projects. This resulted in enhanced choice for project partners and consequently in less tensions between them. The need of additional efforts in knowledge management [Smith and Farquhar, 2000] had to be fullfilled since in such a model, information and knowledge needed for a 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 modeling 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 have to face a psychological challenge of shifting from the culture of the enterprise and the motivation of the individual towards the network culture in which sharing knowledge brings advantages, not danger – providing that IPR issues are properly handled. The concept of network intelligence as a capability of going beyond the fixed individual identity by dialogue, mutuality and trust [Palmer, 1998] is an unavoidable counterpart to technological preconditions of networked organizations. Building this kind of intelligence is a long-lasting, but very important process that should be strongly encouraged in the society, starting with education system, which, unfortunately, still strongly favorizes individual competition over cooperation.

5. Conclusion

As knowledge is becoming one of driving forces of our economy and also of our 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 very dispersed. In this paper we want to present some technical promises to bridge these problems, and at the same time point out some non-technical issues that are connected with it.

We also present our belief that education programs should and could support the process by preparing next generations for the use of knowledge technologies, and at the same time by educating them for work in collaborative settings needed. Both is unavoidable for solving complex problems of today and of tomorrow.

Acknowledgement

This work was partially supported by the Slovenian Research Agency program Knowledge Technologies (2004-2008).

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, Vol. 27, No. 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, vol. 14, 121-136.

Jorge, A., Bojadžiev, D., Alves, M.A., Stepankova, O., Mladenić, D., Palouš, J., Flach, P., Petrak, J., (2003): Internet support to collaboration: A knowledge management and organizational memory view. V: Mladenić, D., Lavrač, N., Bohanec, M., Moyle, S. (ur.) (2003): 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, 247-259.

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

Lavrač, N., Urbančič, T., (2003): Mind the gap: Academia-business partnership models and e-collaboration lessons learned. V: Mladenić, D., Lavrač, N., Bohanec, M., Moyle, S. (ur.) (2003): 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., maj 2007, vol. 37, no. 3, 429-439.

Mladenić, D., Lavrač, N., Bohanec, M., Moyle, S. (ur.) (2003): 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) Number 4, 294-307.

Pubmed: Overview (http://www.ncbi.nlm.nih.gov), 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 , vol. 15, 199-204.

Urbančič, T., Štepankova, O., Lavrač, N., (2006): Enhancing human choice by information technologies. V: Proceedings of the seventh international conference on human choice and computers (HCC7), IFIP TC 9, Maribor, Slovenia, September 21-23, 2006 (J. Berleur in sod., ur.), 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. V: Artificial Intelligence in Medicine LNAI 4594 (R. Bellazzi, A. Abu-Hanna, J. Hunter, ur.) , Springer, 217-226.

Strokovni življenjepis

Tanja Urbančič *is Dean of the School of Engineering and Management at the University of Nova Gorica. She is also research fellow at the Department of Knowledge Technologies at the Jožef Stefan Institute. She received her PhD in Computer Science from the Univeristy 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 a co-author of several book chapters and journal papers in IEEE Transactions on SCM, Engineering Applications of Artificial Intelligence, Journal of Intelligent and Fuzzy Systems and others.*