

Omnipack Case Study – Competitive Intelligence Supported by Semantic Annotation

Tomáš Kliegr¹, Jan Rauch¹, Vojtěch Svátek¹,
Jiří Šplíchal², Tomáš Vejlupek²

¹Department of Information and Knowledge Engineering (DIKE),
University of Economics, Prague, W. Churchill Sq. 4, 130 67 Prague
{tomas.kliegr, svatek, rauch}@vse.cz

²Tovek spol s r.o., Chrudimská 1418/2, Prague 3
{splichal, vejlupek}@tovek.cz

Abstract. The importance of application of knowledge technologies in business practice is soaring. This trend is reflected in the curricula of university courses. This paper presents academic partnership project between DIKE and Tovek on the design and semantic annotation of business intelligence reports. The goal of the project is to create a shared knowledge base for members of the Omnipack Business cluster. This real-world task was transformed to two team case studies assigned at compulsory Knowledge engineering course taught by DIKE. The first case study encompassed writing a report evaluating strategic market position of a selected member of the cluster using the Porter's Five Forces model. In the second case, students were asked to annotate the report and extend the annotation ontology. The output of the project is a knowledge base, which was created as a result of a merge of the reports.

1 Introduction

University research has been paving the way for practical applications. Many enterprises have already become aware of the value that knowledge technologies can bring to their business and are now seeking professionals with the skills to design and operate knowledge intensive applications. It is again the role of universities to communicate to their students the importance and applicability of this field of study and equip them with the necessary skills to meet the expectations of the labor market. The academic partnership between DIKE¹ and Tovek described here is aimed at involving undergraduate students of applied computer science in solving a real world knowledge engineering case study focused on semantic annotation.

¹ Department of Information and Knowledge Engineering of University of Economics Prague

2 Semantic Annotation Case Study

The setting for the case study was to increase the competitiveness of members of the *Omnipack* cluster, which is a group of companies operating in the packaging industry based in the Hradec Králové, Pardubice and Vysočina regions that cooperate to achieve synergies such as better access to research and expensive machinery².

Tovek Company was contracted by the Omnipack cluster to improve sharing and exchange of knowledge within the cluster. One of the means was the design of a shared knowledge base describing the position of each of the cluster members on the market. DIKE as an academic partner in the project contributed to this knowledge base with the outcomes of team projects assigned in the introductory knowledge engineering course 4IZ210.

In Winter Semester 2007/2008 there were 8 course groups each with roughly 20 students. Students formed 38 teams of size approx. 4. Each team was assigned one company from the cluster to write a report about. The reports followed the *Porter's Five Forces* model [1], which is a business methodology for a qualitative evaluation of a company's strategic position. In accordance with this model, the student reports addressed the following issues: threat of new entrants, bargaining power of buyers, threat of new substitute products, bargaining power of suppliers, and rivalry of existing firms. Aside from resources available on the web, the use of proprietary sources such as interactive databases (Proquest 5000) and Anopress were a mandatory part of the assignment. Several teams experimentally evaluated company's web presence with tools such as *Alexa* or *Way Back Machine*³. The output of the first task was one report per team (6000 words).

Next, the reports were annotated. The initial ontology for annotation was designed by Tovek, but the students were asked to extend it by new concepts, relations and optionally introduce new types of relations. Students then tagged parts of the report – usually on paragraph level – with concepts from the ontology. One text fragment could be tagged by multiple concepts. Both ontology design and tagging were performed in the *Tovek Topic Mapper* tool (see Figure 1 for screenshot).

Students were shown how the labor-intensive annotation task can be automated with the ontology learning and evolution tool *DINO* developed [2] at DERI Galway. *DINO* was used for demonstration purposes only as it currently lacks support for Czech and is not interoperable with other tools in the project⁴. This tool is tuned for ontologies in Semantic Web languages (RDFS/OWL), not for topic maps.

² The definition of cluster requires its members to be competitors based in the same territory, which Omnipack meets.

³ www.Alexa.com gives estimate of a number of visitors of a website, *WayBackMachine* available at Archive.org provides a snapshot of the website in various time in the history.

⁴ *DINO* is a Semantic Web not a Topic Map application.

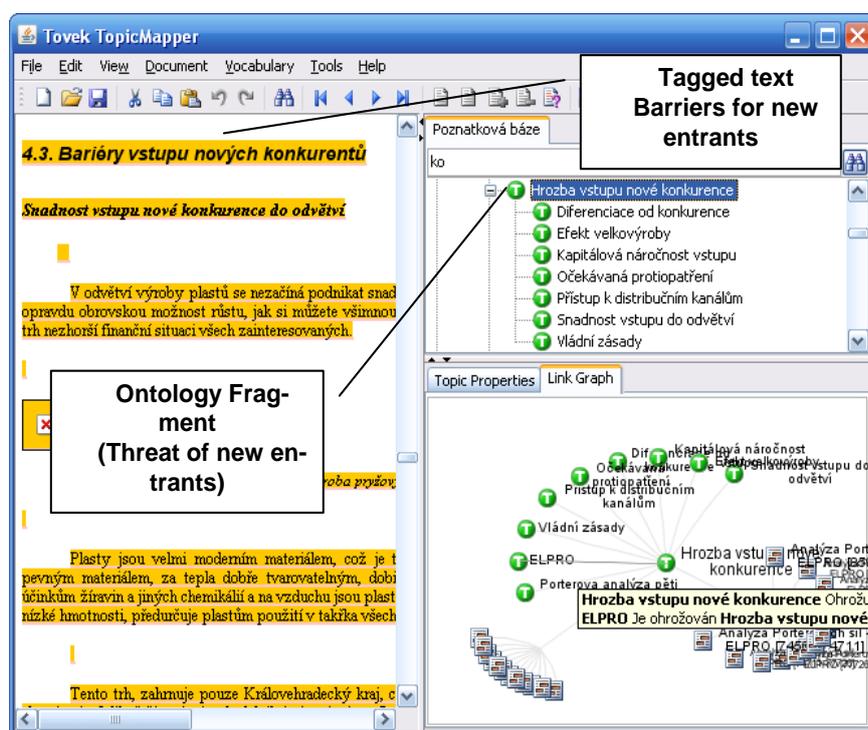


Fig. 1. The report being annotated (Left), ontology (Upper Right). The Link Graph (Lower Right) shows relations between concepts in the ontology and annotated fragments of the text.

The desired knowledge base was created by merging the annotated reports. The merger was carried out in the *Ontopia Knowledge Suite* tool⁵, which also provides a platform for demonstrating how this kind of knowledge representation can facilitate information retrieval from the knowledge base.

3 Software Support for Semantic Annotation

Tovek Topic Mapper (TTM) is a Java application used for semantic annotation and ontology design in the project. It was originally developed to enhance interoperability in the exchange of digital learning resources in different repositories across the EU⁶. TTM is compatible with the Topic Maps (TM) standard ISO 132 50. In TTM both the document and the vocabulary⁷ are saved as a topic map (XTM format), the mapping information denoting which text fragment is tagged with which concept is stored as a topic map as well. TTM was designed specifically for semantic annotation; its main

⁵ <http://www.ontopia.net/solutions/products.html>

⁶ Within the Calibrating eLearning in Schools project IST-028205 [3].

⁷ TTM uses different terminology than is common in the Semantic Web community. In TTM, a vocabulary roughly corresponds to what is usually called ontology.

benefits are clean user interface and ease of use. These features make TTM a good educational aid.

Omnigator, a reference implementation of Ontopia Knowledge Suite Navigator Framework⁸, was used in the project for knowledge base management and answering of user queries. *Omnigator* is a free stand-alone generic topic map browser. Students can thus experiment with it also outside the university computer labs. Its generic nature and extendibility is hoped to endorse student interest in using it further, especially in their bachelor and diploma theses.

4 Conclusion

The undergraduate course:

http://pes.vse.cz/main.php?action=PRED_INFO&id_predmetu=4IZ210&lang=en
Information and Knowledge Processing has been taught at DIKE for more than a decade. Including the semantic annotation case study to its curricula allowed undergraduate students to get hands-on experience with a relatively simple knowledge representation format (Topic Maps); those interested can then deepen their knowledge in an elective graduate course on Semantic Web technologies also offered by DIKE. Teaching knowledge technologies (although applied) to undergraduate students is in-line with world trends; a similar course *IT 499 Applied Knowledge Technologies* is already taught e.g. at George Mason University at Fairfax. This shift stems from the fact that the state of the art in knowledge technologies is evolving as is the scale and nature of its applications in business.

Acknowledgement

This research is supported by grant project GACR 201/08/0802 "Aplikace metod znalostního inženýrství při dobývání znalostí z databází"

References

1. Porter, M.E.: *Competitive Advantage*, The Free Press, New York (1985).
2. Nováček, V. et al: *Semi-automatic Integration of Learned Ontologies into a Collaborative Framework*. In *Proceedings of IWOD/ESWC 2007*. Springer Verlag, Innsbruck (2007)
3. Calibrate Project Deliverable 1.1. Supporting tools integrated into LRE portal version 3, Available online: http://calibrate.eun.org/shared/data/calibrate/deliverables/D3p3_v2.pdf

⁸ A complete software solution for building applications based on the Topic Maps standard.