Literature review: Automated knowledge generation in Topic Maps

Q.H.J.F. Siebers

Univeristy of Maastricht, Morpheus Software Maastricht

Abstract. This document reviews three publications that discuss or propose a way of automating knowledge generation in Topic Maps. The presentation of P. Kruijsen offers the use of inference rules as the bases of logical deduction. The paper of Q. Siebers further expands this idea to an implementable solution. The presentation of L. Garshol uses logical induction by means of statistical analysis of keywords.
The inference rule method has a strong logical bases. The statistical method has a strong automatization level.
The combination of these methods – and with others – might produce better results for automated knowledge generation. More research is needed in this field.

1 Introduction

Topic Maps[1] [4] provide a new solution for knowledge management and storage. Because it's based on the ideas of mind-maps and semantic networks, it provides the possibility of logical analysis.

This paper reviews three publications that are involved in the research into automated knowledge generation in Topic Maps. The first, 'Using PSI's in inferencing' [5], is a presentation given at TMRA'05 [2] [6] by Peter-Paul Kruijsen. The second publication, 'Implementing Inference Rules in the Topic Maps Model' [10], is a paper presented at TMRA'06 [7], and published by the University of Maastricht, by Quintin Siebers. The third publication, 'Automated Classification' [1], is a presentation by Lars Marius Garshol at Topic Maps 2007 [11].

All three publications propose a method of extracting new knowledge from a topic map, semi- or fully automatic. Also, all three publications aim there results at the same software framework and query language this framework uses.

1.1 Contents

The remainder of this review is structured as follows. Section 2 describes the publications in more detail. Section 3 compares the publications and the relations between them. Section 4 discusses the implications these publications present.

[1] When the ISO model is referred, we use the term 'Topic Maps', when an instance of the model is referred, we use the term 'topic map'.

[2] TMRA: 'Topic Maps Research and Applications'

2 Detailed descriptions

In this section the three publications will be further explored.

2.1 Using PSI's in inferencing

This publication describes a question made by P. Kruijsen to L. Garshol about using PSI's in inference rules. Enabling this would create the possibility for interaction between topic maps with different ontologies (Interoperability). Eventually the main question in the presentation becomes:

> 'Can inference rules be placed within a topic map?'
>
> *Kruijsen, P. [5]*

At the end of the presentation, a list of requirements is presented that would be needed to support storage of inference rules within a topic map.

At the time of the presentation, the implementation of inference rules within the framework was based on them being outside of a topic map. The use of inference rules is unique in the market, because of the choice of query language.

The presentation presents the idea that an inference rule can be inferred during use of the topic map, to create – and reveal – new knowledge.

2.2 Implementing Inference Rules in the Topic Maps Model

This paper was written as a bachelor thesis. It supplies a guideline for a framework developer on how to implement inference rules storage within a topic map.

The paper does not specify a specific research question, but mentions what the goals of an implementation of inference rules within a topic map should be. The paper takes these goals as a basis for argumentation.

The definition of an inference rule states that they are a derivation of knowledge. This means that applying the rules to a topic map would create new knowledge.

After the introduction into Topic Maps and inference rules, the paper starts by discussing requirements of an implementation of inference rules within Topic Maps. These requirements involve syntaxes, in-memory storage, a listener model and a way of controlling cached facts.

Next, the paper supplies the reader with some consideration points for when creating an implementation. These points are aimed to fit into the listener model requirement.

The interoperability section discusses the usability of inference rules in practical situations.

Finally, the paper concludes that any implementation would require fine tuning to match requirements. It also states that more research is needed in order to discover the potential in regards to automatic derivation .

2.3 Automated Classification

This publication is a presentation of a new library which was added to Ontopia's [8] framework, the OKS [2]. This library makes it possible to scan texts in documents and extract keywords from them.

The presentation states two methods of extraction: LSA and Extraction. LSA stands for Latent Semantic Analysis, and uses vector spaces and linear algebra to compare complete documents. This creates a clustering of documents that are alike. The Extraction method counts the times a keyword occurs in the text, and gives it a relevance score.

The Extraction method can be used to extract new keywords from the documents, which can then be added to the topic map. This will in time create a repository of key words used in all documents linked tot the topic map.

In the conclusions of the presentation, it is stated that the current implementation is for assisting in manual classification. However, it also states that more research could lead to fully automated classification (and thus creation of new knowledge). It even mentions that it might be possible to eventually extract ontology from texts. This is meta knowledge and would therefor influence everything else in the topic map.

3 Comparison

It's quite clear from the detailed descriptions of the publications that Q. Siebers based his paper on the presentation of P. Kruijsen. All the requirements specified by Kruijsen are used in the paper, except the use of constants in the inference rule head. Also, the paper works out the interoperability issues mentioned by Kruijsen in more detail.

The presentation from Kruijsen and the paper from Siebers both address the issues of extending the topic map automatically by using the possibilities of the query language as a basis. This language is a logical language, based on Prolog [9, 3]. The rules, as proposed, define a new type of association based on existing types of associations. When assumed they are logically correct, they will generate a logical consequence (the new knowledge) . This process is called logical deduction[3].

The presentation of Garshol on the other hand, bases the extraction of new knowledge on the use of statistical methods. The LSA method tries to cluster documents that resemble each other, and the Extraction method uses statistical rules on top of keyword counting to determine the usability and generality of the keywords. This process of selecting a word as a prominent keyword on the amount of times it occurs can be viewed as logical induction[4].

[3] Wikipedia: Deductive reasoning is the kind of reasoning in which the conclusion is necessitated by, or reached from, previously known facts.

[4] Wikipedia: Induction (...) is the process of reasoning in which the premises of an argument are believed to support the conclusion but do not ensure it.

Another difference between Kruijsen and Siebers on the one side, and Garshol on the other side, is the origin of the new knowledge. The origin in the publications of Kruijsen and Siebers – the origin of the inference rules that is – are mainly users or developers of topic maps. The knowledge is deducted from this origin. On the other hand, Garshol's method extracts new knowledge from texts automatic. In this case, the origin of the knowledge is the automatic process of scanning documents.

4 Implications

Both methods, inference rules and statistical extraction, provide new knowledge to a topic map. However, both methods have their downsides.

The inference rules are based on logic, and process rules to gain more knowledge. In this lies its flaw: without the rules, no new knowledge can be deducted. In both publications about the use of inference rules to automatically generate knowledge, there is no mention of a way to automate the process of making the rules.

The statistical extraction functions without such a set of rules – except for the linguistical rules – to deduct the new knowledge. Instead, it completely works on the principles of induction. This might work fine when wanting to extract keywords, but it will be hard to use it to extract logical statements automatically.

The combination of the two methods discussed might combine the strengths and defeat the weaknesses. Also, the use of – and combination of – other reasoning techniques might reveal an complete and optimal automatization of knowledge generation. The reasoning techniques 'abduction' and 'defeasibility' might be such methods.

4.1 Conclusion

Neither publication proposes the perfect solution to automatically generate knowledge. More research is needed to analyze other methods of reasoning, and combinations of these, for finding a more suitable method. The goal of future research should be to find a method that can automatically generate new knowledge on every knowledge level within Topic Maps.

References

1. Garshol, Lars Marius (2007). Automated Classification. http://www.topicmaps.com/tm2007/papers/Garshol_presentation.pdf.
2. http://www.ontopia.net/solutions/products.html (2007). The Ontopia Knowledge Suite. http://www.ontopia.net.
3. International Organization for Standardization (1995). Information technology - Programming languages - Prolog. Technical Report ISO/IEC 13211, International Organization for Standardization, Geneva, Switzerland.
4. International Organization for Standardization (2000). Information technology - SGML Applications - Topic Maps. Technical Report ISO 13250, International Organization for Standardization, Geneva, Switzerland.
5. Kruijsen, P. (2005). Using PSI's in inferencing. http://www.informatik.uni-leipzig.de/~tmra05/PRES/PPKb.ppt.
6. Maicher, L. (2005). TMRA'05 - International Workshop on Topic Maps Research and Applications. http://www.informatik.uni-leipzig.de/~tmra05/.
7. Maicher, L. (2006). TMRA 2006 'Leveraging the Semantics'. http://www.informatik.uni-leipzig.de/~tmra/2006/.
8. Ontopia (2006a). Ontopia: The Topic Map company. http://www.ontopia.net.
9. Ontopia (2006b). tolog. http://www.ontopia.net/topicmaps/materials/tolog.html.
10. Siebers, Q.H.J.F. (2006). Implementing Inference Rules in the Topic Maps Model. http://www.fdaw.unimaas.nl/education/bachelor/conference/Implementing%20Inference%20Rules.pdf.
11. Topic Maps 2007 (2007). Topic Maps 2007 - The First International Topic Maps User Conference. http://topicmaps.com/tm2007/.