Applications of topic maps in knowledge management systems

Article *in* International Journal of Electronic Business · January 2004

DOI: 10.1504/IJEB.2004.005883 · Source: DBLP

7 authors, including:

Josef Küng
Johannes Kepler University Linz
160 PUBLICATIONS 528 CITATIONS

A Min Tjoa
TU Wien
552 PUBLICATIONS 3,698 CITATIONS

Roland Wagner
Johannes Kepler University Linz
231 PUBLICATIONS 1,102 CITATIONS

Wolfram Wöß
Johannes Kepler University Linz
90 PUBLICATIONS 825 CITATIONS

Applications of topic maps in knowledge management systems

Franz-Peter Glawar*, Josef Küng, Thomas Luckeneder, Knud Steiner, A Min Tjoa, Roland R. Wagner and Wolfram Wöß

Institute for Applied Knowledge Processing (FAW),
Johannes Kepler University Linz,
Altenberger Straße 69, A 4040 Linz, Austria
E-mail: fglawar@faw.uni-linz.ac.at
E-mail: jkueng@faw.uni-linz.ac.at
E-mail: tluckeneder@faw.uni-linz.ac.at
E-mail: ksteiner@faw.uni-linz.ac.at
E-mail: tjoa@faw.uni-linz.ac.at
E-mail: rrwagner@faw.uni-linz.ac.at
E-mail: wwoess@faw.uni-linz.ac.at
*Corresponding author

Abstract: This paper gives a short overview not only of the basic concepts of topic maps, but also of the advanced concepts. This document discusses a meta-data structure, which allows making a topic map persistent within a relational database. To make integrity checks within topic maps possible, several constraint mechanisms are introduced. Additionally, for building world-wide internet directory services a multilanguage component is included. Finally, this paper deals with the advantages of topic maps for a conception and ongoing implementation of a knowledge management system in the legal domain.

Keywords: topic maps; knowledge management; organisational memory; knowledge representation; directory services; relational databases.

Reference to this paper should be made as follows: Glawar, F-P., Küng, J., Luckeneder, T., Steiner, K., Min Tjoa, A, Wagner, R.R. and Wöß, W. (xxxx) 'Applications of topic maps in knowledge management systems', *Int. J. Electronic Business*, Vol. 2, No. 5, pp.508–518.

Biographical notes: Franz-Peter Glawar is Research Assistant at the institute of applied knowledge processing (FAW), Johannes Kepler University Linz, Austria. He studied computer science and is now engaged in several research projects in the areas of information systems and semantic modelling.

Josef Küng is Associate Professor for applied computer science at Johannes Kepler University Linz, Austria. He studied computer science and has received his PhD from Johannes Kepler University Linz in 1994 for a thesis on regular neural networks. His current research interests are similarity searching, semantic based modelling and, semantic enriched queries.

Thomas Luckeneder is Researcher at the department for Medical-Informatics of Upper Austrian Research GmbH. He studied business informatics and has received his PhD being about topic maps and databases in 2002. His current research area is medical informatics.

Knud Steiner is Research Assistant and Project Coordinator at the Institute of Applied Knowledge Processing (FAW), Johannes Kepler University Linz, Austria. He studied business informatics and is now researching in the areas of semantic modelling and topic maps.

A Min Tjoa is a Professor of Software Technology at the Technical University of Vienna. He is teaching in the area of information systems and data warehouses. His current research interests are new database developments, data mining and data warehousing.

Roland Wagner is a Professor of Information Systems at the Johannes Kepler University of Linz. He is the director of the Research Institute for Applied Knowledge Processing (FAW). He is managing various industrial and research projects in the areas of information systems, knowledge-based systems and electronic commerce. He is teaching in the area of Information Systems and his current research interests are new database developments, object oriented database design, data mining and data warehousing.

Wolfram Wöß is Associate Professor for applied computer science at Johannes Kepler University Linz, Austria. After studying computer science and finishing his PhD thesis concerning rule based product modelling he is now researching and teaching in the areas of information systems, semantic data modelling and e-commerce. Beside, he has managed several research projects.

1 Introduction: the topic map model

The ISO standard *ISO/IEC 13250:2000 Topic Maps* [1] (‘Topic Map Standard’) defines an abstract model for semantically structured, self-describing link networks laid over a pool of addressable information objects (‘Topic Maps’) as a basic architecture within the following scope of applications:

- to qualify the content of information objects in order to enable navigational tools such as indexes, cross references, citation systems or glossaries
- to link pieces of information together for the purpose of enabling navigation between them, e.g. for creating thesaurus-like hypertext interfaces for corporate information and knowledge bases
- to filter views of information adapted for specific users and/or purposes, e.g. to define partial views depending on user profiles and/or knowledge domains
- to structure unstructured information objects or to combine unstructured information bases with structured ones in a way that an arbitrary structure is imposed without altering its original form.

The standard further specifies a serialisation syntax as the interchange format of topic maps based on the ISO standard *HyTime* [2] by means of a meta-document type definition (DTD).

1.1 Basic concepts

The basic concepts of the Topic Map Model are:

- *Topic and topic type*. Topics constitute the basic building block of the topic map model. A topic is an aggregate of topic characteristics (e.g. name characteristics) whose organising principle is a single subject. A subject, in turn, can be

  > "… any thing whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever …." [1]

  Topics are usually typed by other topics establishing a class-instance relationship between the typing topic and the topic being typed.

- *Topic occurrence and occurrence role type*. A topic occurrence represents a set of references to addressable information resources residing outside the topic map that are somehow relevant to a topic. Topic occurrences are usually typed by a topic indicating the sense in which they are relevant to a topic (occurrence role type).

- *Topic association, association type and association role type*. A topic association establishes a relationship between two or more topics. Whereas the meaning of a topic association as a whole is specified by its association type, the user-defined role played by one or more topics participating is specified by a set of association role types. Both the association type and association role types are also topics.

1.2 Advanced concepts

Other concepts defined by the Topic Map Standard expressing the power of the topic map model are those of:

- *Scope and theme*. A scope is a set of topics acting as themes used to control the user-defined validity of topic characteristic assignments. For example, scopes can be used to qualify topic name characteristics in multilingual topic maps.

- *Facet, facet type and facet value type*. Facets provide a means for annotating information objects pointed at by topic occurrences with simply structured meta-data (property/value pairs). Both properties and property values are expressed by means of topics (facet and facet value type).

- *Public subject*. A topic may have a public subject descriptor assigned in order to enable the recognition of semantically equivalent topics, if topic maps are being merged.

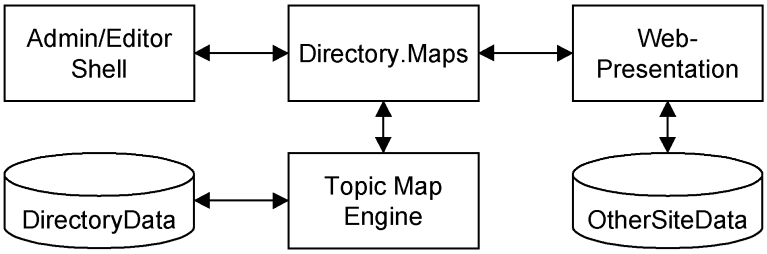
2 Application scenario I: a topic map based internet directory service

As discussed in Section 1, topic maps are being embraced by a wide number of organisations and companies. But there is still a lack of software that supports the topic map standard. In this paper, an internet directory service is introduced, which is based on a topic map using the advantages of semantic networks in this field. A key point for a useful and efficient application is the possibility to make the topic map persistent within a relational database and to establish integrity check mechanisms. Additionally, a multi-language component is important for providing worldwide access. Concept and architecture of the topic map based internet directory service are described in this section [3].

2.1 Components of the topic map based internet directory

In order to increase both maintainability and extensibility, the application for managing the content of the internet directory service is divided into six components (Figure 1):

Figure 1 Components of the topic map based internet directory [3]



- *Admin-/editorshell and web-presentation*. These two modules cover all necessary Tango-scripts for the admin-/editorshell and the web-presentation. Tango is a web-application software that enables dynamic generation of web-pages whose content results from a database query. In this application, Tango-scripts are not used for implementation of business logic of the topic map management and administration in general, but for implementation of the user and context dependent user interface. Database write operations are separated from these scripts and implemented as stored procedures within the Directory.Maps module. Data exchange is done by using well-defined interfaces with the advantage, that a module can be changed without affecting any other module.
- *Directory.Maps*. The essential business logic for the internet directory service is part of the Directory.Maps module. It is implemented in Oracle, mainly using stored procedures.
- *Topic map engine*. This module contains Oracle stored procedures with basic database access operations. Database operations are independent from the business logic in the Directory.Maps module.

- *DirectoryData and OtherSiteData*. DirectoryData includes the database schema for storing a topic map in a relational database, as well as the actual content of the internet directory service. The content is visualised via web-presentation module. In most cases, an internet directory service is embedded in a large and comprehensive website, which is characterised by heterogeneous data structures and a large database. Due to that fact, data concerning topic maps (DirectoryData) is separated from other data on this site. This kind of data is stored in OtherSiteData.

2.2 Three-layer architecture

To establish a well-structured architecture, topic maps are logically organised in three layers:

- Within the *meta-layer*, all topic map meta-types are specified. They are directly derived from the topic maps standard. Meta-types are used to control the internal business logic of the Directory.Maps module. Furthermore, the advanced concepts for topic maps [4] are also mapped to meta-types in this layer (see Section 2.3).
- The *schema layer* represents the schema of the internet directory service. The schema is primarily used to guarantee consistency and integrity of the content of the directory, which is stored within the instance layer. Consistency and integrity checks are implemented using constraints that are specified within the Directory.Maps module.
- The *instance layer* contains the actual content of the directory service based on the defined schema topics stored within the schema layer.

2.3 Meta-types and constraints

To make integrity checks within topic maps possible, several constraint mechanisms are introduced in this approach. The implementation of meta-types within the meta-layer is based on both, the basic concepts of the topic maps standard, as well as the advanced concepts for building topic map applications. In detail, the implemented and used meta-types are:

- *Topic type* (tt), *association type* (at), *association role type* (art), *occurrence role type* (ort).
- *Facets*. In this approach, it is distinguished between discrete and non-discrete facet types (ftd and ftnd), as well as discrete and non-discrete facet value types (fvtd and fvtnd). Discrete facet types and facet value types are supposed to be frequently used facets with a limited range of values. In contrast to discrete facets, the range of values of non-discrete facets is not limited. Non-discrete facet value types are arbitrary strings and mostly used only once. Both types are treated and processed as topics.

Furthermore, the introduced concepts in [4,5] resulted in the following meta-constraints. Each constraint is internally implemented as an association and belongs to the schema layer:

- *Association constraint* (ac). Within the schema layer, a concrete association constraint defines which association role types and which topic types (a regular topic of the associated topic type respectively) may be associated at the instance layer for an association type. For example, an association constraint for the association type 'contains' defines that it is only allowed to associate regular topics of the topic type 'country' within the association role type 'container'. In addition, it defines that it is only allowed to associate regular topics of the topic type 'state' within the association role type 'containee'.
- *Occurrence constraint* (oc). Analogous to an association constraint, an occurrence constraint determines which occurrence role type(s) are permitted to be assigned to regular topics of a specific topic type.
- *Facet constraint* (fc) *and facet value constraint* (fvc). According to the distinction of discrete and non-discrete facet types and facet value types, there is an analogous distinction between discrete and non-discrete facet constraints and facet value constraints. A facet constraint defines which facet types (regardless discrete or non-discrete) are permitted to be associated to a certain occurrence role type. Furthermore, a facet value constraint defines which facet value types are permitted to be associated with a facet type.

2.4 The multi-language component

The topic map standard includes a general concept for scopes use. Scopes are used to limit the validity of topics or the name characteristics of certain topics (e.g., the topic 'Paris' is only valid within the scope 'Countries' or a concrete base name of a topic is only valid within the scope 'Language: English').

Basically, all topic map elements with name characteristics are allowed to be used for multi-languages. In particular, this includes the following topics at the schema layer: topic type, association type, association role type, occurrence role type, discrete and non-discrete facet type and facet value type. It also includes the following topics at the instance layer: regular topic and locator-alias.

Editors, who are responsible for the content of the internet directory service, are supported in managing and maintaining different language versions of topics by the admin-/editorshell. Using scopes offers the advantage that each topic exists only once, but its name characteristics (which express its textual and also semantic representation) exist for each language scope. To achieve this, editors have to specify different terms according to the required languages at schema and instance layer for each topic.

2.5 User authorisation

Corresponding to the three-layer architecture of the main system, the authorisation system considers three different user groups:

- A *system user* is permitted to change (read and write access) each topic from each layer (meta-, schema- and instance-layer).
- An *administrator* has full access to topics at the instance- and schema-layer and read access, only to topics at the meta-layer. An administrator is responsible for the schema of a topic map.
- *Editors* have full access to topics at the instance-layer and read access, only to topics at the schema- and meta-layer. Editors are responsible for the content maintenance of the directory service.

The assignment of a user to a user group is the result of the login process that requires username and password.

3 Application scenario II: a topic map-based organisational memory system for the legal domain

3.1 Knowledge management issues in the legal domain

Knowledge results when information is combined with experience, context, interpretation and reflections and it can, therefore, be regarded as a high-value form of information that is ready to be applied to decisions and actions [6] and may be of several types [7]: knowledge about something is referred to as declarative knowledge, procedural knowledge is knowledge of how something is performed and knowledge dealing with why something occurs is called causal (or analytical) knowledge. With regard to one of the main issues of knowledge management, another significant distinction of knowledge is made between implicit (or tacit) and explicit knowledge [8]: implicit knowledge is personal knowledge that is embedded in individual experience and is therefore, not directly accessible to others. The most common way to share implicit knowledge is by means of highly interactive communication, since it can hardly be reduced to rules and recipes. In contrast, explicit knowledge is usually embodied in published resources such as research reports, conference proceedings, text books, journal papers and the like. Thus, it can be more easily stored, communicated and shared among individuals than implicit knowledge [9].

In the case of a law firm, for example, declarative knowledge is ‘knowledge of the law’: the legal principles contained in statutes, court opinions and other sources of primary legal authority. Procedural knowledge is knowledge of the mechanics of complying with the law’s requirements in a particular situation. Causal knowledge results from analysing declarative knowledge as it applies to a particular fact setting (e.g. conclusions reached about the course of action a particular client should follow in a particular situation). Whereas, most of the declarative and procedural knowledge is explicitly available, that, what actually makes up a successful lawyer, is his implicit knowledge [10].

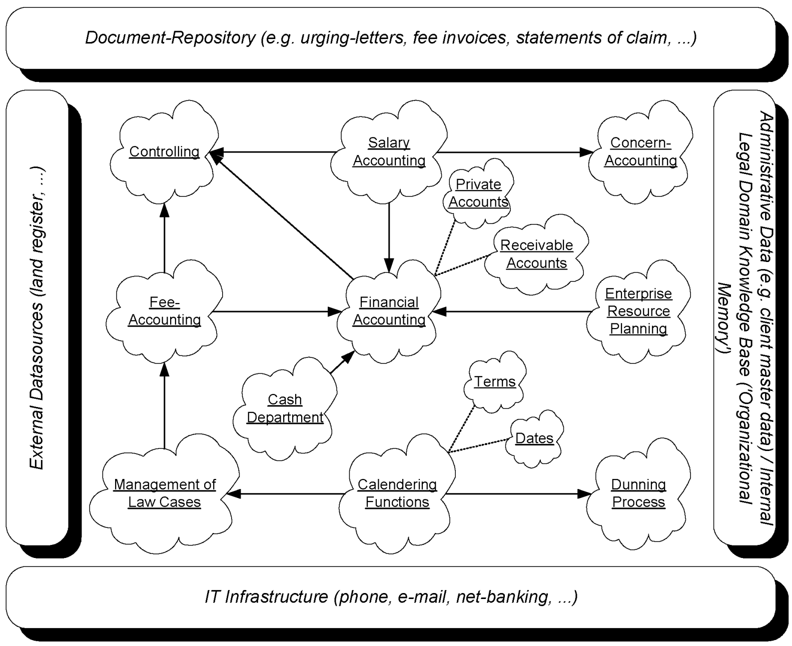
Therefore, the main issues with regard to a system supporting knowledge work in the legal domain are: The system must be able to [7]

- capture, represent and store both explicit and implicit knowledge
- categorise, link and index stored knowledge objects corresponding to knowledge units
- facilitate the retrieval of stored knowledge objects
- present the content of its knowledge base with sufficient flexibility to render it meaningful and applicable across multiple contexts of use.

3.2 Implementation and design issues

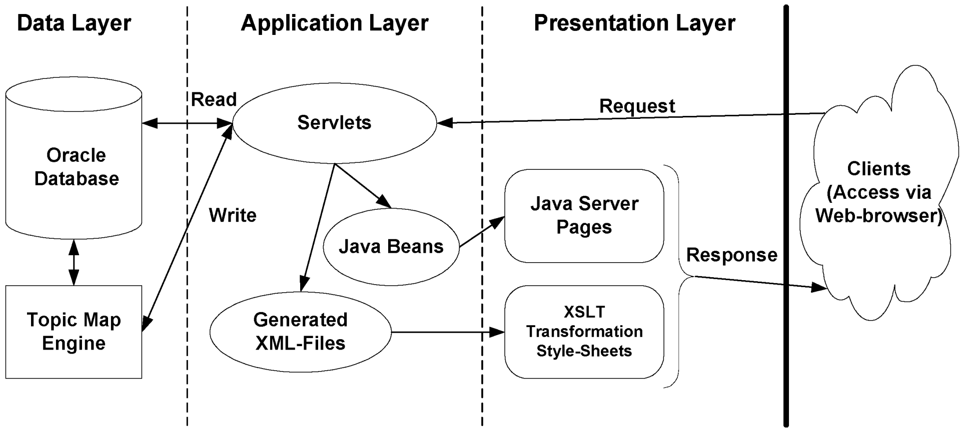
The presented Organisational Memory System for the legal domain is an integrated web-based information system, which comprises all workflows in a lawyer's office, as well as document management facilities for the maintenance of necessary documents (e.g., fee invoices, statements of claim). Despite of these core requirements, the system is influenced by a lot of basic conditions that have to be considered as well. Figure 2 shows all partitions of an integrated information system for the legal domain, covering the core modules, as well as the basic conditions [11].

Figure 2 Modules of a topic map-based OMS for the legal domain [11]



Considering the requirements mentioned above, the structure of the resulting topic map application is designed as a three-layer architecture illustrated in Figure 3. The data layer contains all necessary data stored in form of a topic map in a relational database (Oracle database). Furthermore, the topic map engine is also part of the data layer acting as a well defined interface and API for all write operations from the application layer to the topic map. Read access to the topic map is implemented via the SQL interface as select statements.

Figure 3 Architecture of topic map application [11]



The application layer actually covers the business logic for maintenance of the entire application. The most important part of this layer, are servlets acting as central entry point or for user requests. All necessary operations for a particular user request are instantiated from the appropriate servlet. Furthermore, all transformation processes from topic map data to aggregated XML files are also implemented within these servlets.

The dynamic generation of the user interface is done within the presentation layer. This layer consists of Java Server Pages and XSLT (eXtensible Stylesheet Language Transformation) style sheets for graphical presentation of generated XML files.

In addition to the presented three-layer architecture, we adopted the MVC (Model View Controller) design pattern in order to separate business logic, data and presentation logic [12], so that each part can be modified without affecting the other two parts (e.g. layout-adaption by a web-designer). Each implemented servlet acts as controller for the entire application using Java Beans (Model) for internal representation of topics. These instantiated Java Beans are used by Java Server Pages (View) to generate the appropriate user interface.

For example, the generation of invoices is a generation process, which consists of three steps and it is controlled via a servlet:

- Calculation model independent generation of the input-XML-file, on basis of the recorded internal activities and cash expenditures.
- The generated input-XML-file is transformed into an output-XML-file, which depends on the appropriate calculation model. The transformation is processed via XSLT style sheet, which represents a calculation model.
- The transformation from the output-XML-file to the desired output format is also realised with XSLT. Currently, our application supports two output formats: HTML for preview purposes and Adobe PDF (Portable Document Format) for printing invoices.

4 Conclusion and further research

In this paper, a meta-data structure is introduced, which allows making of a topic map persistent within a relational database. To make integrity checks within topic maps possible, several constraint mechanisms are introduced. Additionally a multilanguage component is included. The mentioned components are very important for the efficient implementation of internet directory services. Furthermore, the presented approach also focuses on the graphical design of the web presentation. The user interface has to correspond to the concepts of topic maps in order to allow appropriate navigation within the knowledge structure.

Further work will start with the implementation of a so-called 'Topic Map Engine' based on the introduced data model. Other fields of interest will be the usage of XML (Extensible Markup Language) and XSLT (eXtensible Stylesheet Language Transformation) for the calculation of appropriate fees on the basis of different calculation models, and the persistent implementation of topic maps in XML-databases.

Another focus of interest, will be the implementation of an improved graphical presentation of the topic map. That topic, which is currently visited by the user is positioned in the centre of this graphical representation and the semantically relevant topics are connected via lines, which length represents the semantic relevance.

Particularly, in connection with TIS (Tourism Information Systems), a further interest will be to develop a framework to cope with heterogeneous data sources. In a 'mediation layer', the heterogeneous data could be mediated with the help of a topic map in such a way, that the user will get access to these heterogeneous data easily [13].

References

1 International Organization for Standardization *(ISO) (2000) ISO/IEC 13250:2000 Information Technology – SGML Applications – Topic Maps*, Geneva.

2 International Organization for Standardization (ISO) (1997) *ISO/IEC 10744:1997 Information Technology – Hypermedia/Time-based Structuring Language (HyTime),* Geneva.

3 Luckeneder, T., Steiner K. and Wöß, W. (2001) 'Integration of topic maps and databases – towards efficient knowledge representation and directory services', *Proceedings of the 12th International Conference of Database and Expert Systems*, *DEXA*, pp.744–753.

4 Rath, H.H. (1999) 'Mit topic maps intelligente informationsnetze aufbauen – mozart und kugeln', *iX Magazin*, No. 12.

5 Rath, H.H. (2000) 'Making topic maps more colorful', *Proceedings of XML Europe 2000 Conference*, http://xtech.org/papers/xmleurope2000/pdf/s29-01.pdf (last visited August 21, 2003).

6 Davenport, T.H., De Long, D.W. and Beers, M.C. (1998) 'Successful knowledge management projects', *Sloan Management Review*, Vol. 39, No. 2, Winter 1998, pp.43–57.

7 Zack, M.H. (1999) 'Managing codified knowledge', *Sloan Management Review*, Vol. 40, No. 4, Summer 1999, pp.45–58.

8 Nonaka, I. and Takeuchi, H. (1995) *The Knowledge Creating Company*, Oxford University Press, New York.

9 Luckeneder, T. *et al.* (2001) 'Towards a topic map based knowledge management system for the legal domain', *Proceedings of the Third International Conference on Information Integration and Web-Based Applications & Services*, *IIWAS*, pp.217–227.

10 Edwards, D.L. and Mahling, D.E. (1997) 'Toward knowledge management systems in the legal domain', *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work: The Integration Challenge*, Phoenix, AZ, USA, November 16–19, 1997, pp.583–590.

11 Küng, J., Luckeneder, T., Steiner, K., Wagner, W. and Wöß, W. (2001) 'Persistent topic maps for knowledge and web content management', *Proceedings of the Second International Conference on Web Information Systems Engineering*, *WISE 2001*, December 3–7, 2001, Kyoto, Japan.

12 Gamma, E. *et al.* (1995) *Design Patterns – Elements of Reusable Object-Oriented Software*, Addison Wesley.

13 Ashari, A., Leewattanakit, W., Tjoa, A.M. and Wagner, R.R. (2001) 'A framework for integrating heterogeneous tourism information sources', in Sheldon, P.J., Wöber, K.W. and Fesenmaier, D.R. (Eds.): *Proceedings of the International Conference Information and Communication Technologies in Tourism, ENTER 2001*, *[ISBN 3-211-83649-7]*, Springer Verlag Wien New York, Montreal, Canada, April 24–27, 2001, pp.102–111.