

Topic Map technology - the state of the art

Graham Moore

STEP UK Ltd., Swindon, United Kingdom

gdm@stepuk.com

<http://www.stepuk.com>

Abstract:

Topic Maps are being embraced by a wide number of organisations throughout the world. Companies and individuals have realised how the power of Topic Maps can help them solve their information problem. However, in order to make the vision a reality there must be software that supports the Topic Map paradigm. This paper presents a look at Topic Map technology, asking questions about what it should, could and does do. It presents the cutting edge of Topic Map development.

This paper does not focus on one Topic Map technology, rather it identifies key functionality drivers such as Topic authoring, Topic Map merging and illustrates the ways different technologies have tackled these problems.

This paper dives under the hood and looks at some of the implementation issues of building Topic Map technology, issues in this area are things such as the object model design, exposed interfaces, Topic Map storage and searching.

The other key aspect of this presentation is the analysis of how different Topic Map technologies are being, or could be used, in the construction of information systems. This analysis will provide a template for the construction of other Topic Map systems and provides real world scenarios of the technologies in use.

Introduction

Topic Maps are being embraced by a wide number of organisations throughout the world. Companies and individuals have realised how the power of Topic Maps can help them solve their information access problems. However, in order to make the vision a reality there must be software that supports the Topic Map Standard. This paper presents a look at Topic Map technology, asking questions about what it should, could and does do. This paper provides a cutting edge insight into Topic Map software development. However, given the non-static nature of technology the related presentation will present the ideas that are on the edge at the time of the presentation.

This paper looks at the implementation issues of building Topic Map technology. This technology is one that support the Topic Map lifetime, from creation and authoring through maintenance and delivery and onto evolution. It focuses specific aspects of these stages such as Topic Map merging and the import and export of Topic Maps. In addressing all these issues it compares and contracts object model design, relational support, Topic Map storage, searching and what interfaces to expose to developers.

In this paper we also present how different Topic Map technologies are being, or could be used, in the construction of new information systems. This analysis will provide a template for the construction of other Topic Map systems and provides real world showcase scenarios of the technologies discussed.

Topic Map technology

Topic Map import/export

The first aspect we cover is the Topic Map import/export mechanism. We start here as the standard dictates that to be a conforming Topic Map application the software must be able to read and write a Topic Map instance that adheres to the DTD architecture. This DTD is defined in the standard in terms of the HyTime architectural form. There is no definition, within the standard, that restricts or defines the kinds of operations that can be performed on an internal data structures used to represent Topic Maps. So the process can be seen as:

1. Process valid Topic Map instance and create internal topic map representation

2. Manipulate internal representation
3. Export manipulated data model in a form that adheres to the Topic Map DTD

Due to the work happening with XTM, the XML Topic Map initiative, Topic Map software will be expected to be able to work interchangeably with either syntax.

As the Topic Map model uses Topics in various roles, import mechanisms need to be able to deal with forward references. Using an algorithm that scales to work with very large Topic Map instances is a requirement for industrial strength Topic Map software. Currently both one pass with stub topics and two pass solutions have been implemented.

Topic Map merging and internal representation

Topic Map importing leads us on to two other interesting areas, the internal representation and the issue of topic map merging. We will first cover the internal topic map representation. There are a number of approaches that have been investigated so far the main two being object and relational.

The object approach requires that as structures in a Topic Map instance are processed by the import mechanism that objects relating to each construct be created. The classes used to construct the object model are TopicMap, Topic, Occurrence, TopicAssociation, TopicAssociationRole, Name, Facet and FacetValue. As instances of these classes are created they are associated together to give a complete representation of the topic map objects and their relationships with each other. In order to make this model persistent, it is necessary to commit these instances to some form of OODB (Object-Oriented DataBase). Interestingly, the classes within the topic map model can be considered in terms of classes used in an abstract linking model. For example, Topic, TopicAssociation and Facet are subclasses of Link and Occurrence, TopicAssociationRole and FacetValue are subclasses of Anchor. While debate goes on as to whether TopicAssociations should be Topics etc, lessons have been learnt in the construction of linking software, such as GroveMinder, (<http://www.techno.com>) and X2X, (<http://www.stepuk.com>). Building these technologies has provided useful insight into the interfaces to expose and how to persist hundreds of thousands of link structures in a reliable and scaleable way.

Taking the relational approach requires the creation of tables such as Topic, TopicAssociation and TopicAssociationRole. However, it also requires the construction of many join tables. Join tables are required to model the associations between different constructs, e.g. that a topic contains N occurrences, or that a particular Topic characteristic is in a given scope. It has already been shown that the relational model supports Topic Map queries and it will be interesting to see how these two approaches evolve.

The second aspect related to import is the notion of Topic Map merging. Given that a TopicMap instance has already been processed by some given Topic Map software it will be necessary at some point to merge in another Topic Map. The standard provides some guidelines and constraints to be used when merging maps. These include the Topic Naming constraint and the concept of Identity. It has been found that implementing these two mechanisms is a trivial activity. However, user requirements are such that more control over this merging process is desired. This is understandable as a Topic Map represents a commitment in time and is the encapsulation of corporate knowledge. Thus, Topic Map software must provide a user interface to allow the controlled merging of Topic Maps and for the parameterisation of the process. This parameterisation would allow control over what kinds of merges happen automatically and which are referred to the user for authorisation. In addition, the software itself will evolve such that it can make inferences as to which two topics are 'the same topic'. For example, based on fuzzy logic, such as the nature of a Topic's relationship

with other topics, the software could proffer that a given topic has a 75% chance of being the same topic as one in the map being merged.

Topic Map authoring

We have discussed the import mechanism and touched on user requirements in terms of merging. Here we discuss the Topic Map authoring process and how software can support it. The Topic Map authoring process can itself be divided into a number of areas. The first area to address is the automatic creation of a Topic Map from a given information set. This process will create an initial set of topics with occurrences. These two tasks are dependent on the amount of information available within the information set but can produce useable results. The most challenging aspect facing Topic Map software is the creation of strongly typed topic associations. The best results of automatic generation come from heavily marked up data, and especially well marked up indexes. This is because professional indexers have had the opportunity to classify and associate information.

Experience has shown that the automatic generation of Topic Maps is a useful first step in the construction of a production topic map. However, the real value of a Topic Map comes through the involvement of people in the process. Topic Map software will support this by allowing browsing of any existing map and then the easy creation of new topics, topic associations and occurrences. More evolved software will utilise open systems to allow the user to browse a range of repositories that could contain topic occurrences.

As the Topic Map model is so general creation software will most likely employ constraint based components to ensure that Topic Map authors are guided in their work. In addition to this, software may provide a mechanism that can offer a list of topics that may be the same as one that a user wants to create as new. It is these mechanisms that will help to ensure the quality of the Topic Map.

Topic Map delivery in real world systems

We have discussed the import mechanism, the different kinds of internal representation and the ways in which Topic Maps are created and maintained. Finally, we discuss Topic Map delivery. Topic Maps are intended to enable people to have better access to information. Continuing in this spirit, the delivery of Topic Map information must be enabled over a variety of mediums including web, EJB and WAP. Topic Map systems that are flexible and dynamic are more likely to be able to deliver in these different environments. Taking the web as an example a Topic Map server will service user requests. It will deliver using HTML or simple applets the ability to navigate the Topic Map and view occurrences. In a WAP environment the Topic Map can not only be used for the rapid and focused navigation of large information sets it can also be used to select resources that have been created for delivery in the smaller bandwidth WAP environment.

Topic Map software is being used both as dynamic information server and as tool for batch processing information sets in conjunction with the Topic Map data. Topic Map software should be versatile enough to support both modes of operation. Given this it can be seen that a Topic Map system will be a significant piece of any corporate information infrastructure serving the needs of a variety of users and processes.

Conclusion

This paper has given a brief insight into some of the issues and approaches taken in the design and construction of Topic Map software. It has discussed software with regards to the Topic Map creation process, its ongoing maintenance and its delivery. We have highlighted different approaches to the internal representation and at each stage identified where the technology will be heading. The Topic Map paradigm will make a significant impact on the information systems we use, there has already been much progress in the construction of Topic Map software and it will continue to realise the power of the paradigm.

Bibliography

- [1] Steve DeRose, David Orchard, Ben Trafford, Eve Maler (Editors), XLink Working Draft W3C Working Draft 19-January-2000, <http://www.w3.org/TR/WD-xlink-20000119>
- [2] Michel Biezunski, Martin Bryan, Steve Newcomb (Editors), ISO/IEC 13250 Topic Maps
- [3] Charles F. Goldfarb, Steven R. Newcomb, W. Eliot Kimber, Peter J. Newcomb (Editors), ISO 10744 HyTime 2nd Edition

Author

Graham Moore

Chief Technical Officer

STEP UK Ltd.

Postal Address:

Unit B, Dorcan Complex

SN3 5HQ Swindon

Wiltshire

United Kingdom

Telephon: +44 (0) 1793 485465

Fax: +44 (0) 1793 485451

E-mail: gdm@stepuk.com

Web: www.stepuk.com

Graham Moore - Graham Moore is Chief Technical Officer for STEP UK and is responsible for the development of skills and technology that are used in the construction of information systems. Graham is the development lead responsible for the realisation of the X2X XLink product from STEP UK and has been developing Topic Map software for the past twelve months. Graham has worked in the SGML and XML industry for several years, including working for Database Publishing Systems Ltd. Graham is also studying for a Phd in Distributed Information Management at Southampton University, where the current focus is on the symbiosis of metadata and linking.