National Centre For Software Technology
Bangalore

An Intelligent Topic Map Tool A Propos Semantic Web

A White Paper Submitted to

INVITED SESSION

CONTENT- AND SEMANTIC-BASED INFORMATION RETRIEVAL

IN CONJUNCTION WITH

THE 6TH WORLD MULTI-CONFERENCE ON

SYSTEMICS, CYBERNETICS, AND INFORMATICS

(SCI 2002)

Sheraton World Resort, Orlando, Florida, USA

July 14-18, 2002

By The Semantic Web Group, NCST, Bangalore

Written By:
Sireesha Ammanamanchi
Raman R.K.V.S.

An Intelligent Topic Map tool - a propos Semantic Web

Index:

1.Abstract:

This Paper proposes a novel approach for creation and Manipulation of topic maps. Topic Maps can be ornately used for representing any type of intricate structures using the three fundamental concepts of Topics, Associations and Occurrences. We put forward in this paper, a non-technical introduction to an intelligent Topic Map Creation and Manipulation tool that can be extensively used for Knowledge Representation, Knowledge Management, Navigation, Retrieval and Knowledge sharing using topic maps. We also institute the idea of using Topic Maps a propos the "Semantic Web" – which is construed to be the " Web Of Future".

2.Introduction to Topic Maps:

XML Topic Maps (XTM) is a product of the TopicMaps.Org Authoring Group (AG), formed in 2000 by an independent consortium named TopicMaps.Org, originally chaired by Michel Biezunski and Steven R. Newcomb, and co-chaired by Steve Pepper and Graham Moore.

Knowledge Management is one of today's buzzwords. Topic Maps are new ISO Standard for describing knowledge structures and associating them with information resources. The concept of topic maps allows the user to organize and represent knowledge in a form understandable by the computer. As Steve Pepper quoted - *if, as one writer claims [Ruggles 97], "knowledge management covers three main knowledge activities: generation, codification, and transfer", then topic maps can be regarded as the standard for codification that is the necessary prerequisite for the development of tools that assist in the generation and transfer of knowledge.*

3.Topic Maps A Propos Semantic Web:

Semantic Web is deemed to be the "Web Of future". For Tim BL, the father of the web, ***something has semantics when it can be processed and understood by a computer***. And the underlying concept of the Semantic Web is to represent information over the web in a machine processable form. The web is the most opulent repository of information in the history of information technology, but most of the information is not properly structured and till now, the only purpose of the information over the web had been just display. And now, enriching the concept of the WWW by adding semantics to it, and thus hoisting the Semantic Web would most fundamentally require information to be represented in well-defined, structured form so as to be machine-understandable and machine-processable.

In this scenario, using topic maps a propos the Semantic Web is fairly appropriate for the following reasons:

1. Topic maps can be optimally used to organize and retrieve online information.
   Topic map in its most basic form is a collection of topics and associations where associations connect topics together. A topic in its most generic sense can be anything in the universe. Anything that a user is interested in can be a topic. And basically a **'topic'** in a topic map represents a **'subject'** that we are talking about. A topic map is a standardized syntax in which any subject can be treated as a topic and relationships between these objects are **'associations'**. A topic in a topic map can represent any notion, an idea, a thing, a person and so on. An association is a link element that asserts a relationship between two or more topics. Further a topic may be linked to one or more information resources that are deemed to be relevant to the topic in some way. Such resources are called **'occurrences'** of the topic. There are other concepts like scopes, facets etc. which make topic maps more handy and efficient in knowledge representation and management.
2. Topic maps are powerful to represent knowledge and comprehensible and simple to work with. Topic maps are capable of supporting complex interrelationships within and among the concepts of various fields. Topic maps are simply documents and anyone can author a topic map to represent or share their knowledge. Further topic maps enable multiple alternative models of knowledge domains to coexist and to work together in a way that had not been available before.
3. Topic maps provide a logical base for semantic services. Some services required for semantic addressing can be stated as follows:
   i. Given an address, traversal to the addressed piece of information.
   ii. Given a piece of information, generation of an address for it in the terms desired.

iii. Given two addresses, determination of whether they address the same piece of information.

Topic maps can work in any combination of address spaces.

4. According to Charles F. Goldfarb, they lead the XML paradigm forward opening the way for a 'global positioning system for the web'.
5. Topic maps have the forte in them to make the indexes of human knowledge more and more intelligible and comprehensible and making such indexes easier and easier to use.

A graphical representation of topic maps in relation to the semantic web can be given as follows:

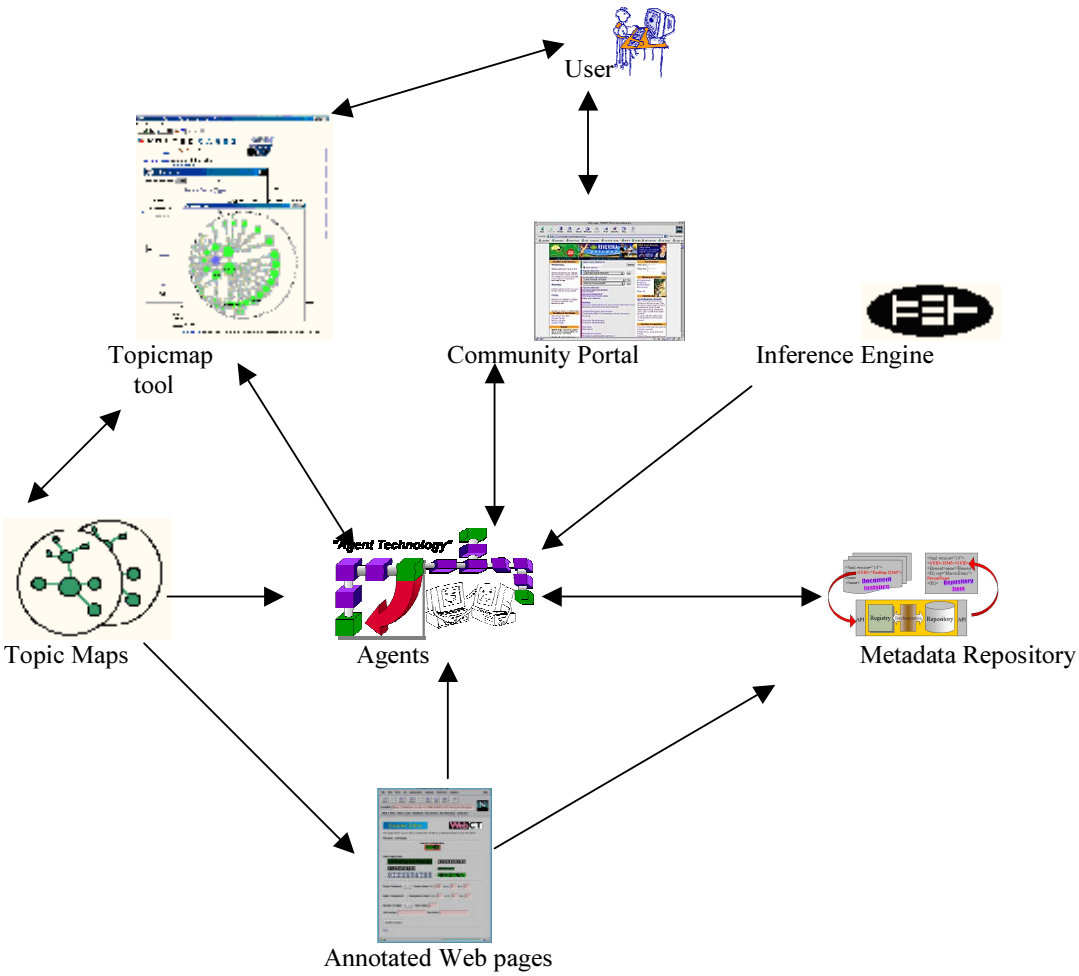

Fig 3.1 Topic Map tool in relation to the Semantic Web

4.An Intelligent Topic Map tool:

4.1 Introduction:

The topic map tool that we are going to develop can be extensively used to create, edit, save, search and query topic maps. Further, the user can navigate through the entire topic maps available using a navigation tool plug in and there will also be a browser plugged in to browse through the occurrences related to any topic in the topic map. Any and every community of the users can utilize it to its best – from the extreme of the users with high expertise in topic maps and Knowledge management and representations to the extreme of the novice users who just want to share their knowledge.

4.2 Previous Works:

While the thinking behind topic maps is mature, the proposed standard itself is relatively new (published in December 2000), and the commercial technology supporting the standard is still in its early

stages. The following can be quoted under the vendors who can be categorized as booming initiators that contributed towards topic maps so far:

- Ontopia's core product is the Ontopia Topic Map Engine. It is a generic implementation of the topic map standard, available as a software development kit. The SDK allows developers to access and manipulate the constructs found in topic maps (topics, associations, and occurrences).
- Empolis, a technology subsidiary of Bertelsmann, produces a product called K42. The K42 knowledge server is a software application developed on the basis of topic map technology. Based on open standards such as Java, XML, and XSL, K42 features a query language viz Topic Map Query Language for accessing the topic maps.

5. Features of our topic map tool:

5.1. User Interfaces:

The user interface is a very crucial and essential part in the design of any computer based information systems because it involves an online interactive communication between the user and the machine. And moreover, when the communication involves knowledge transfer, it is more vital that greater emphasis is given on providing an effective and efficient user interface. As stated earlier, this system is designed for all types of users, neophyte users as well as expert users. So, two types of user interfaces are provided. One is a Natural Language user interface and the other is purely technical user interface. Like wise, for querying the topic maps also, there are two interfaces provided, one is a natural language query interface and the other one being technical query interface.

5.1.1 Natural Language User Interface:

A novice user can work create, retrieve and manipulate topic maps using the natural language user interface that is provided with the system. The conversion of a natural language statement or a query will take place in two modules. The first module is the linguistic module and is essentially the linguistic interpreter that does the lexical analysis, Syntactic analysis and semantic analysis and converts the natural language statement into an equivalent logical form. The second module is the translator, which takes the logical form as input and does contextual analysis on it and converts it into the XTM format.

A graphical representation of the Natural Language User Interface can be shown as follows:

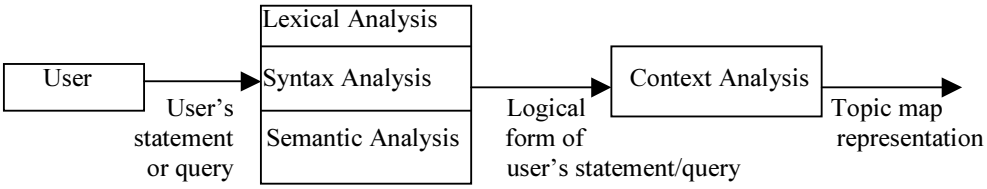

Fig 5.1 : An overview of the Natural Language Processing done with the system

The lexical analyzer and syntax analyzer and the logical form generator access the dictionary, which contains the domain dependent details and general syntactical feature details of words to assist in the lexical and syntax analysis. The lexical analyzer retrieves the word along with the syntactic feature, attributes of the word and passes the information to the syntax analyzer. In case the word is not available in the dictionary, it is passed on to morphological analysis, which extracts the root word from the derived word. It passes the root word along with any feature information conveyed by the original derived word back to the lexical phase. The lexical search for the word again commences and if the search is successful the dictionary attributes of the word, along with the features obtained from the morphological phase are attached to the word. The lexical analyzer allows a limited amount of spelling correction in each word by assuming the nearest dictionary item, after cross checking with the user. The syntax analyzer checks the syntactic structure of the sentence and gives as output, the syntactic structure along with the attributes of each word obtained from the dictionary during the lexical analysis phase. The logical form is so designed that it can be unambiguously interpreted and translated into the terminology of the topic maps.

The context analysis phase, takes the logical form as the input, interprets all the terms in it and translates it to the topic map terminology. For the translation of the logical form of the natural language sentence into the topic map terminology, the context analyzer is provided with a look up table, which consists of a mapping of the terms that are most frequently used in the natural language form to those in the topic maps. If there is a match found, the translation is successfully done, and if not, the context

analyzer allows a limited amount of correction in each term by assuming the nearest item after cross checking with the user. New terminology can be added to the look up table depending up on the number of times a phrase is used by the users.

For example, if the user gives an input like the following,

Michael Jackson is a Singer.

Here, the sentence is split into tokens and passed on to lexical and syntax analyzers. The tokens are identified along with the type. The result is given to the context analyzer, which looks up into the table of mappings, and interprets and translates this sentence, into the format Michael Jackson is the instance of the topic Singer and stores the knowledge.

Once the context analyzer translates the input of the user from natural language representation to the topic map representation, it can be stored in XTM format, or as an RDF schema* or in a database. Proper algorithms and rules would be used to identify the topics, instances, and associations etc. and to interpret the natural language sentences and translating these into XTM format.

The same method can be adopted in translating the Natural Language Queries into Topic Map Query Language (TMQL)* statements and displaying the results, and for interpreting the natural language search requests and displaying the search results.

5.1.2 Technical User Interface

The technical user interface of the system consists of several sophisticated forms for an expert user to create, retrieve and manipulate topic maps, by specifying the topics, associations, occurrences, scopes, facets etc. An expert user can query the topic maps using the TMQL statements or search for any information represented by the topic maps.

5.2 Architecture of the topic map tool:

The concept of topic maps, the distributed knowledge representation interface is in a number of ways similar to the human brain. The core analogy is the one between the associations between the topics and the associative memory. Associations between the topics in the topic maps are analogous to the associations between the concepts as they are stored in the brain. However the analogy can go much further including the processes of thought and learning.

The main idea is that different topic maps can be seen as nodes in a network connected by links along which many users travel in search for information/knowledge. The entire information provided in the topic maps can be organized in order to provide better access to the knowledge that they contain. This makes it possible to 'mine' the implicit knowledge distributed over millions of topic maps and associations between them and the way they are used.

This knowledge can be used to improve the quality of the results that users get in return for their queries, point them towards relevant documents, reorganize the web to make it more efficient, and even discover wholly new concepts and relationships. Thus, the collective intelligence of the millions of users and authors of web information may be leveraged, turning the web from a mere repository of information into an adaptive, self-organizing knowledge system that actively "thinks" in interaction with its users.

Such an intelligent web can be seen as the first step towards the "Global brain" which is likely to be the next step of evolution.

Our topic map tool will be built starting from this point of view, with the analogy between the associative memory of the human brain and the topic maps as the key strategy.

5.2.1. Functionality of our Topic Map Tool:

Using our topic map tool, a user can

i. Create a new topic map of his own choice with his own information set:

The user can create a topic map either using the natural language user interface or using the technical interface. In addition to that, the system would have the following functionalities

- Prompt the user as and when appropriate: For example, the system may understand that the user is referring to Michael Jackson, if he specifies a resource reference to Michael Jackson, which is already present else where, the system may intelligently prompt the user and check if he is referring to the same Michael Jackson.
- Allow the user to append or write notes to the existing topics or anything that is in a topic map.
- If the user is trying to create the topic maps in natural language format, the system has to understand or be able to analyze to the maximum possible extent what the user is trying to input.

* RDF – Resource Description Framework is a part of the W3C Metadata activity. The goal of this activity and RDF is especially to produce a language for the exchange of machine-understandable descriptions of resources on the web. For further reference http://www.w3.org/TR/rdf-schema.

* For further reference on TMQL, http://k42.empolis.co.uk/tmql.html

- Depending upon the path of traversal that is followed by the user in the topic map, the system tries to predict/foresee what the user is trying to represent and prompts the user to counter check if that is what is in his mind. A log maintained about the most frequently traversed paths determines this.

- The System will as well prompt the user to check the useful links about the topics he is creating. For example, if the user is trying to create a topic map on Computer Networks, the system may prompt the user with some important URL's that are related to computer networks, and allow the user to add some of them as references to the topics he creates under computer networks.
- Depending up on the name of the topic map, the system may try to understand which category that belongs and cross check with the user whether or not the topic map falls under that category and add it to that category, if the user agrees.
- Further the system displays a quick help to the user depending up on the location of the cursor or mouse pointer on the system.
- Adding intelligence to the tool would require many more functions to be visualized into a reality, and there will be many more value additions to the current design.

ii. Edit an existing topic map:

The user can edit an existing topic map by adding new topics, associations, occurrences, scopes etc. or by modifying the existing ones. While editing a topic map also, the system will have all the functionalities mentioned above.

iii. Retrieve information from the existing topic maps by querying/searching:

The user can retrieve the information either by searching the topic maps for a topic or by querying the topic maps for a particular topic or anything that meets the specified conditions.

iv. Merge two or more topic maps:

The user can merge two topic maps, either by subject if two or more topic maps have the same subject indicator or by name if two or more topic maps have the same name in the same scope.

v. Save the topic maps:

The user can save the topic maps in any of the three formats viz database, RDF schema, or an xtm document and converting from one form to another as and when required.

vi. Navigating:

There is a navigation tool plugged into the system, which allows the user to navigate through any number of topic maps. There will be a path of traversal recorded and displayed for the convenience of the user. If the number of the topic maps in the path of traversal exceeds a predefined limit, there will be a back button using which the user can always find out the previous topic map/topic visited in the traversal.

vii. A Plugged-in Browser:

There is also a browser plugged into the tool, which allows the user to browse through the occurrences and resource references specified by the user for a topic.

Initially though this is the design proposed for the intelligent topic map tool to be developed, there will be many value additions in course of development as well as after the development of the tool.

6. Conclusion:

In this paper, " An Intelligent Topic Map tool" has been discussed. The salient features of the tool are that it is used to represent and manage knowledge and thus it can be used extensively in conjunction with the concept of the Semantic Web. Further, it provides three different useful formats to save topic maps. Once developed this tool can be extensively used to create and manipulate the topic maps, by novice as well as expert users.

7. References:

- The standard itself (unofficial copy): http://www.ornl.gov/sgml/sc34/document/0058.htm
- http://www.infoloom.com
- Robin Cover's SGML/XML bibliography: http://www.oasis-open.org/cover/gen-apps.html#topicMaps
- ISO site: http://www.iso.ch
- HyTime Users' Group: http://www.hytime.org
- GCA conferences: http://www.gca.org
- www.topicmap.org

- www.topicmaps.net