

Topic Maps for repositories

Kal Ahmed

Chrystal Software, Slough, United Kingdom
kal_ahmed@chrystal.co.uk
<http://www.chrystal.com>

Abstract:

No abstract was provided for this paper.

Introduction

This paper discusses the potential application of Topic Maps as an interface to a multi-user document repository; presents some possible implementation approaches to creating Topic Maps for a repository; and finally demonstrates some graphical tools for Topic Map navigation and creation.

Current repository navigation techniques

Hierarchical browsing

A common feature of nearly every repository system, is the use of a hierarchy of nested containers for organising and navigating through content. Typically the structure of any given sub-tree of a hierarchy is defined by a single user and used by all other users with interest in the content stored therein. Users of such a hierarchy are therefore constrained by the structure imposed by the administrator. For small repositories, or repositories used by a single person, such an organisation may work well - most users are relieved of the task of organising the content and need only to learn where to look for the items of interest ¹ .

As a corpus increases in size, it gets progressively harder for a user to learn the structure of the hierarchy unless it matches the way in which that user mentally organises or works with the content. Of course, not all systems are rigidly controlled by a single administrator - many systems provide the freedom for users to create and manage a sub-tree of the hierarchy - but this simply leads to more confusion - without a pre-arranged classification system, any coherence in the organisational structure is lost as multiple organisational criteria are squeezed into a single system.

Searching

Browsing is not the only way to find content in a repository - most repositories also support searching. Most often this is provided for the benefit of those who consume the data, rather than for those responsible for creating and maintaining it - giving users a way to completely bypass the structural organisation of the data. However, the results of the query can only be as good as the query itself. An under-specified query will result in an unmanageably large set of hits, and an over-specified query might miss a piece of relevant content. Furthermore, a query across repositories of differing types requires that those repositories define a common set of meta-data with commonly agreed semantics for a combined search to return meaningful results.

¹ *A hierarchical organisation of content also eases administration of access controls, as access controls applied to a container node can be inherited by all of its contents. The use of a Topic Navigation Map for organising data in the repository does not cause security problems, as users will still only be able to access that content which the standard access-control lists of the underlying repository grants them access to.*

Using Topic Maps for repository navigation

Associative browsing

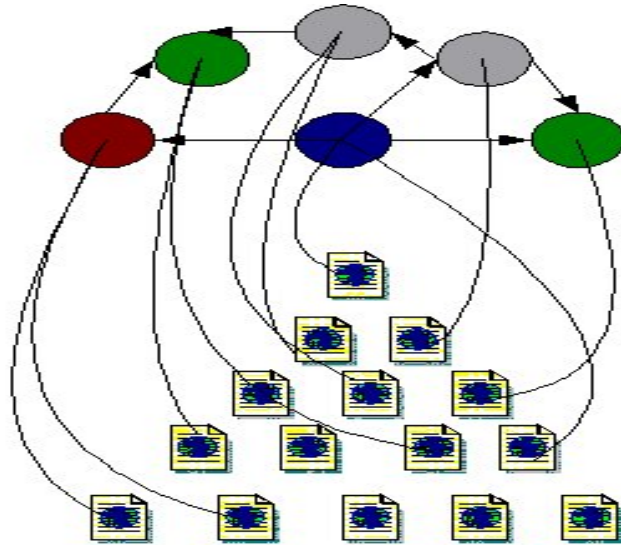


Figure 1. Associative browsing with a Topic Map

Topic Maps provide the casual browser of the repository, with a richly cross-linked structure over the repository content.

Topic occurrences create 'sibling' relationships between repository objects. A single object may be an occurrence of one or more topics, each of which may have many other occurrences. When a user finds/browses to a given repository object, this sibling relationship enables them to rapidly determine where there are other objects regarding the same topic as the current one. Topic associations create 'lateral' relationships between subjects - allowing a user to see what other concepts covered by the repository are related to the subject of current interest and to easily browse to them. Associative browsing allows an interested data consumer to wander across a repository in a guided manner. A user entering the repository via a query might also find associative browsing useful in increasing the chance of serendipitous discovery of relevant information.

Topic Map querying

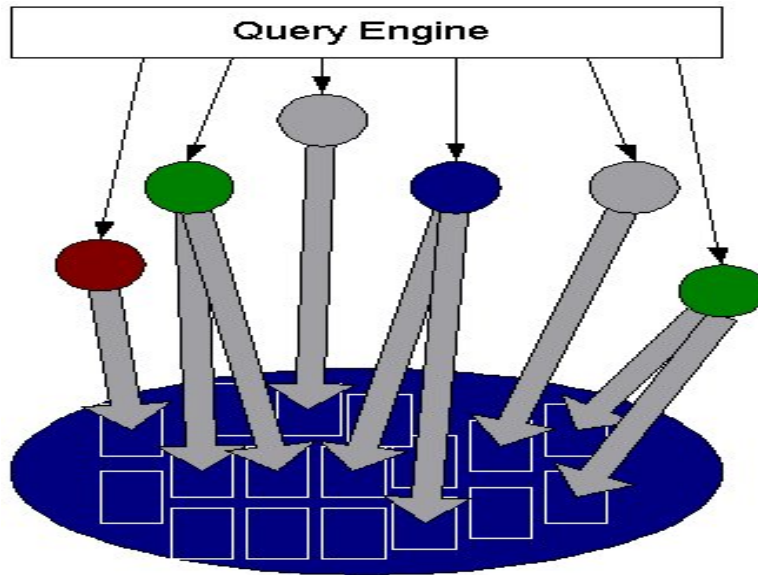


Figure 2. Topic Map querying

A Topic Map can be used to provide a useful higher-level abstraction across one or more repositories. Topic Maps provide a number of useful features for query-based access to the repository:

Topics can be used to group together repository objects which relate to a single abstract concept. Each repository object may be defined as an occurrence of the topic. Occurrences may be assigned a role, defining the relationship with the parent topic. These typed relationships mean that a user may first query on a concept and then rapidly narrow the size of the results set by occurrence role.

Scopes can be used for specifying query domains - enabling a user to easily narrow or broaden the size of the query set.

Facets can be used to provide indexing outside of the underlying repository. Facets enable indexing of repository contents where the underlying repository itself does not. Facets also enable multiple repositories to be indexed with a common set of meta-data, enabling meaningful cross-repository querying.

Topic cartography

The creation of useful Topic Maps should become a prime concern to the creators of large corpora. Strategies for the creation of such maps would be driven by the requirements of the Topic Map user and by the constraints of the authoring environment. Broadly speaking, 3 types of Topic Map can be identified:

1. System Topic Map
2. Semantic Topic Map
3. User-Defined Topic Map

The system Topic Map

The System Topic Map is a Topic Map which represents the structure of the underlying repository. Characteristics of repository objects are directly mapped to Topic Map constructs - these include such characteristics as the location of the object and object meta-data. Such a mapping could be made dynamically by an agent interposed between the topic map engine and the underlying repository and may be combined with other topic maps on-the-fly by a processing application.

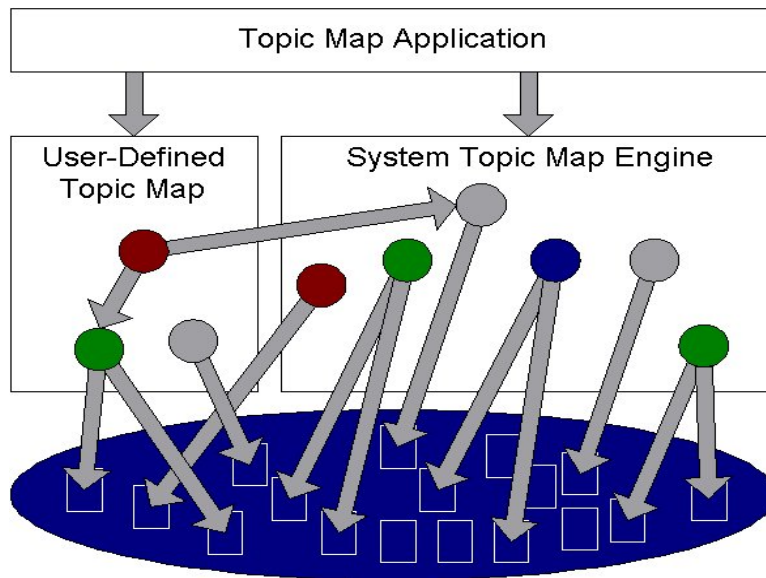


Figure 3. The system Topic Map

A key use of a System Topic map would be in creating a bridge between the repository and the Topic Map environments. It may be easier for someone used to navigating through the repository directly to get used to a Topic Map view of a repository if there are 'landmarks' which map directly to the underlying structure. Where time and effort has been spent in creating a hierarchical organisation of data in a repository, the System Topic Map provides a portable means for capturing the result of that effort. Many organisations have already made the choice to store portable data (XML, SGML etc.) but a move to a new repository can lose all of the effort and knowledge encapsulated in the organisation and repository-level meta data associated with the content.

A further use for a System Topic Map is in combining multiple repositories of the same type into a single 'virtual' repository which can be browsed seamlessly. A single topic map application could communicate with and merge the output of multiple system topic map engines.

The semantic Topic Map

The Semantic Topic Map is generated by automatically extracting meaning from the content of the repository and representing the connections made by analysis of that meaning as a Topic Map. Whereas in a System Topic Map the topics represent repository objects; in a Semantic Topic Map the topics represent concepts described by one or more repository objects.

Content analysis may be simply driven by such characteristics as document structure, meta-data or contained hyper-links. Typically a well-marked up, well cross-linked corpus will generate a good Semantic Topic Map.

A more complex approach, might make use of linguistic analysis to extract meaning from the textual content of documents.

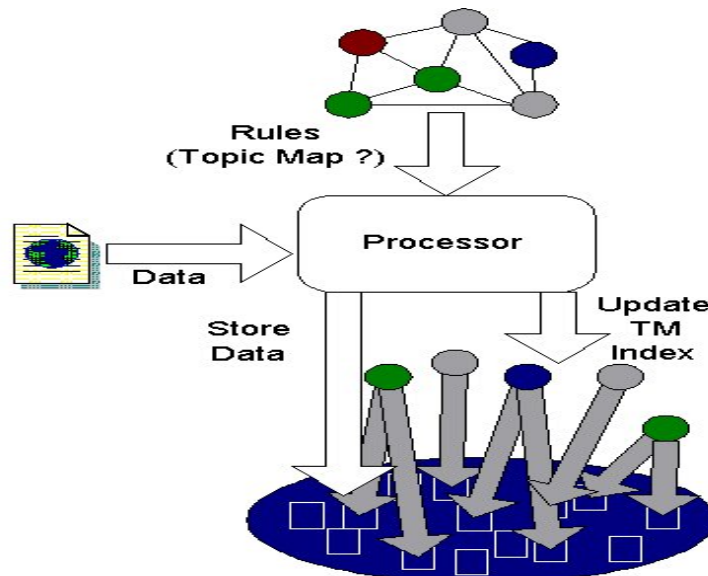


Figure 4. Semantic Topic Map generation

While it is possible that some semantic analysis could be done 'on-the-fly', the processing overhead of some of the more advanced forms of analysis might make Semantic Topic Map generation an asynchronous process. The rules used for the semantic analysis are, in themselves, an important form of knowledge as they encode the way in which the relationships between repository objects are inferred from their content by users of the repository. Some semantic information may be encoded as associations between topics or as topic-occurrence relationships. Other semantic information may be extracted which applies to just a single repository object and this information may be represented using a facet.

When used to generate an index for a corpus, a semantic Topic Map provides indexing features above and beyond those of a standard static index. Scopes provide a means of quickly creating domain-specific indices - combining multiple domain-specific indices on demand would enable each user of the same corpus to create their own personalised index of that corpus. Facets provide easily searchable meta-data. Associations provide rich, typed cross-linking between conceptual areas.

The user-defined Topic Map

The User-Defined Topic Map provides an individual with a means of creating their own perspective on a set of data. User perspectives may be:

- | | |
|------------------|--|
| Organisational | The perspective maps the repository to enhance location/retrieval of data. |
| Knowledge-driven | The perspective adds value to the data by asserting associations between repository objects based on some deeper understanding of the concepts represented by those objects. |
| Task-driven | The data is organised according to the user's work processes. |

User-defined Topic Maps have potential application in 3 areas:

- Individual Workspaces
- Shared Workspaces
- Knowledge Management

Individual workspaces

Using a Topic Map to create an individual workspace gives the user a means of better managing access to frequently used documents and to organise data in multiple ways. Topic Maps can be used to create logical paths from an abstract concept to a specific document in a way which more closely matches the way the user thinks. Tools are needed to make the construction, maintenance and navigation of such Topic Maps as easy as possible and to integrate as tightly as possible with the day-to-day tools and processes. Topic Maps allow a user to relate single data instances to multiple subject areas - such as a standard text referenced from multiple projects. Topic Maps also give the application the freedom to link to resources in other tools (such as email, PIM systems and remote documents) - enabling the user to pull information from many disparate sources into a single coherent set for their use.

Tools are already available that aid in this form of personal organisation. Topic Maps may be used as an interchange format between such products and/or platforms - for example moving my mind map from my PC to my Palm and back or creating a 'mobile' workspace on an Internet-accessible site that can travel with me.

Shared workspaces

Shared workspaces enable users to share knowledge by communicating to each other the associations and relationships between data instances. Multiple Topic Maps may be combined with relatively little effort, to quickly generate a composite view of the same data set. Topic Maps created by individual users can thus be shared across an organisation, enabling many other users to gain the insights and benefit from the knowledge encoded in the Topic Map. As with any Topic Map application, data instances may be in a repository or located elsewhere within or outside the organisation - as long as it can be addressed in some way.

When user share their workspaces, Topic Map merging rules and applying additional scoping using added themes can be used to ensure that the perspective of different people are combined only to the degree desired by the end-user.

Knowledge management

Topic Maps can be used to encode ontologies prepared by one or more subject matter experts. Such a map may be used simply to transfer an ontology from one tool to another, or as a 'publishing medium' for an ontology. A topic map engine combined with other analysis tools (such as linguistic analysis tools) could be used to automatically annotate documents according to a given ontology and record the resulting annotation as a Topic Map. Again, Topic Map merging rules could be used to generate composite or comparative views of the same data set using different ontologies or analysis methods.

Topic Map GUIs

Topic Maps may be statically published (as a subject index, for example) or more dynamically displayed to the user. For the types of 'workspace' applications described above, an intuitive GUI is a key requirement

for success. Topic Maps enable users to create large quantities of meta-data and highly interconnected sets of data. The challenge for a GUI is to present this graph and the associated meta-data in a readily interpretable manner.

Data visualisation techniques are gradually entering the mainstream. As graphics hardware prices continue to fall and new software becomes available, building a compelling Topic Map GUI is becoming easier.

Topic Map GUI approaches

Topic Maps are essentially interconnected graphs with (potentially) many dimensions of meta-data. There are a number of approaches to the visualisation of such data already in the commercial domain:

Hyperlinked-Trees A graph can be interpreted as a hierarchical tree relationship with additional hyper-links between nodes. Topic Maps support this type of visualisation due to the hierarchical relationship between topics, associations and types and also the containment nature of the relationship between topics and occurrences and associations and association roles. Standard GUI tools are capable of displaying this kind of tree. For example STEP's on-line topic maps (<http://www.topicmaps.com>). Distorted tree visualisations such as InXight's Hyperbolic Tree Browser (<http://www.inxight.com/demos/ht/index.html>) enable a larger proportion of the hierarchy to be displayed in the same amount of screen real-estate as the traditional tree browser.

Graphs Graph visualisation displays the Topic Map as a set of interconnected nodes. A static graph visualisation simply displays the nodes with their interconnections. Dynamic graph visualisations limit the scope of vision of the user to the node of interest and all nodes within a certain distance. As the user shifts focus from node to node, the display of the graph changes interactively. This form of visualisation enables all of the connections in the Topic Map to be more equally displayed, rather than making a dominant hierarchical relationship and at the same time avoids overwhelming the user with the quantity of information contained in the Topic Map. Mind mapping tools such as the Brain from Natrifical (<http://www.thebrain.com>) use this form of display quite effectively.

Landscapes An interesting data visualisation technique is to display interconnected information as a landscape, assigning coordinates to topics according to their interconnections and height to coordinates according to the degree of relevance or the degree of convergence of multiple topics. This is the approach used by Cartia's ThemeScape product (<http://www.cartia.com>) to create the NewsMaps web-site (<http://www.newsmaps.com>) - not a Topic Map application, but the potential is there.

Worlds

The data model of Topic Maps seems to lend itself well to the construction of three-dimensional spaces. Topics may be assigned coordinates in 3D space according to specific characteristics. A static 3D world enables the user to 'fly-through' the Topic Map; to learn the 'lie of the land'; to meet other user's browsing through the same map or even to bookmark frequently visited locations. A dynamic 3D world would respond to the user's movements, bringing 'most relevant' topics nearer and moving others further away as the user's focus changes. Three-dimensional worlds are already implemented as glorified chat-rooms (<http://www.activeworlds.com>), perhaps Topic Maps provide a framework for putting these worlds to serious use.

Conclusion

While the current focus for Topic Navigation Maps is on the creation of static publication indexes, there is significant scope for the use of the Topic Navigation Map standard in 'indexing' more dynamic data and to provide an organisational construct on top of one or more repositories.

Topic Map meta-data and facets provide a means of creating a common index across multiple repositories, allowing searching and browsing applications to treat many disparate repositories as a single virtual repository. Topic Map merging and scoping rules facilitate the sharing of individual Topic Maps, allowing users to benefit from the knowledge of others.

To move forward in the use of Topic Maps for these kinds of applications, development of compelling visualisation techniques is a must. Fortunately the tools to build these visualisations are becoming readily available and standard home and business hardware is already capable of advanced visual display which would have been prohibitively expensive only three or four years ago.

Author

Kal Ahmed

Chrystal Software

Postal Address:

53-61 Windsor Road

SL1 2EE Slough

Berkshire

United Kingdom

Telephon: +44 1753 559512

Fax: +44 1753 511955

E-mail: kal_ahmed@chrystal.co.uk

Web: www.chrystal.com

Web: www.techquila.com

Kal Ahmed — Kal Ahmed is a Solutions Architect with Chrystal Software Inc. where he is responsible for the design and implementation of custom solutions built on the Astoria and Eclipse structured document management systems. Kal also runs techquila.com, a web-site of tools, links and information relating to Topic Maps.

Kal has presented Topic Map visualisation tools to the Meta-Structures'99 conference, the UK SGML Users Group and most recently at XML'99.