Information management - Topic Maps visualization

Benedicte Le Grand
Laboratoire d'Informatique de Paris 6, Universite Pierre et Marie Curie, Paris, France
Benedicte.Le-Grand@lip6.fr
http://www-rp.lip6.fr/~blegrand

Michel Soto
Laboratoire d'Informatique de Paris 6, Universite Pierre et Marie Curie, Paris, France
Michel.Soto@lip6.fr
http://www-rp.lip6.fr/~ms

Abstract:

Topic maps - the new ISO standard ISO-IEC 13250 - provide a bridge between the domains of knowledge representation and information management. Topics and topic associations build a structured semantic link network above information resources. Our research aims at visualizing this semantic layer efficiently, which is a critical issue as topic maps may contain millions of topics and associations. This paper is divided into 3 parts. First, we depict briefly basic topic maps concepts. Then, we review a few graph visualization techniques. Finally, we describe the visualization tool we developed at the Laboratoire d'Informatique de Paris 6 and study how this tool may be used - and enhanced - for topic maps visualization.

Introduction: basic topic maps concepts

According to , topic maps - the new ISO standard ISO-IEC 13250 - will become the answer for organizing and navigating through large and continuously growing information pools. They provide a "bridge" between the domains of knowledge representation and information management. This standard defines both an abstract data model and a serialization syntax to represent knowledge structures and to link them to information resources.

Figure 1 describes topic maps basic concepts: topics, occurrences of topics and relationships (associations) between topics.

For example, consider an information pool consisting of conference material. These resources have different roles (occurrence roles) - articles, videos, charts, call for participation, etc. Examples of topics are XML Europe, Paris, France. Topics also have different types - conference, city, country, etc. - and associations exist between them: XLM Europe takes places in Paris, Paris is in France. In this example, there are two types of associations - "takes place in" and "is in".

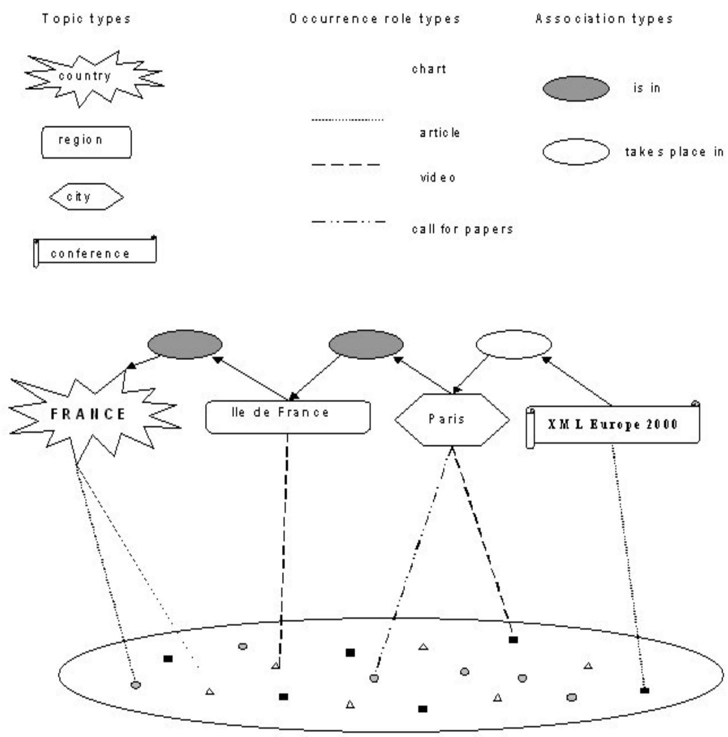

Figure 1. Topic maps basic concepts

As shown in this diagram, a topic map is divided into a topic domain (consisting of topics and associations) and a resource domain.

Topics and topic associations build a structured semantic link network above information resources (topic occurrences). This network allows an easy and selective navigation. However, if topic maps are very large, users may have problems to understand them and find relevant information. Thus, it is necessary to represent them efficiently.

In this paper we will focus on the visualization of the "portable semantic network" and disregard the information resources layer. We thus need to represent topics and topics associations. In the future, we will also represent occurrences (i.e. the links between individual topics and the occurrences of information about those topics).

The portable semantic network can be viewed as a graph in which topics are nodes (vertices) and associations are arcs (edges), both of which are typed. Visualizing this semantic layer may be a critical issue as real topic maps will consist of millions of topics and associations.Therefore, it is interesting to study how existing large graphs visualization techniques may be applicable to topic maps.

Existing graph visualization techniques

NicheWorks is a 2D interactive visualization tool for the investigation of very large graphs that cannot be represented on one static display. NicheWorks allow users to examine a variety of nodes and edges attributes in conjunction with their connectivity information. Parts of a graph may be shown or hidden using interactive manipulation of views of node and edge attributes; nodes and edges have different colors and shapes according to their attributes. Figure 2 and Figure 3 are examples of NicheWorks visualizations.

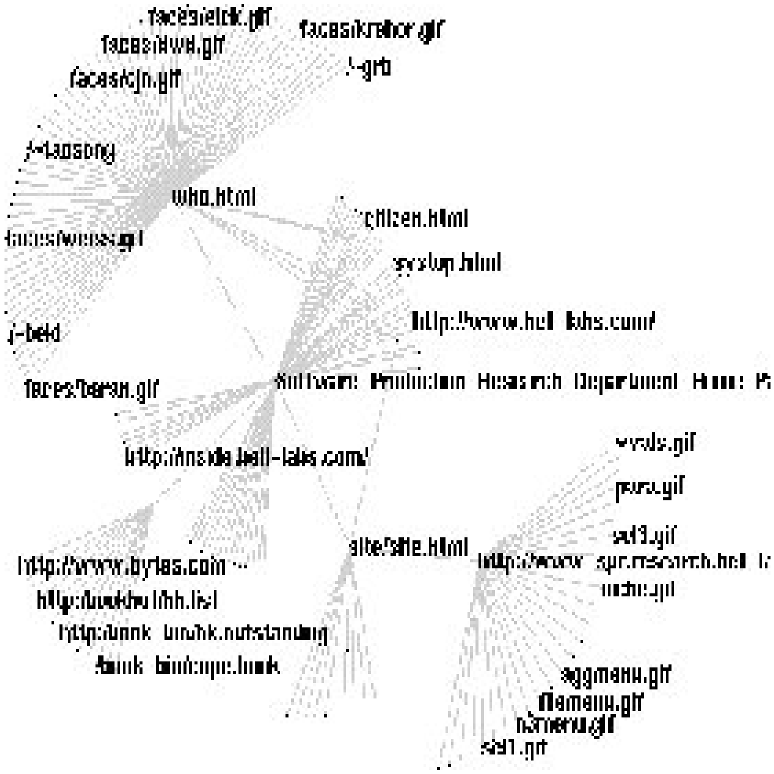

Figure 2. Example of graph visualization with NicheWorks

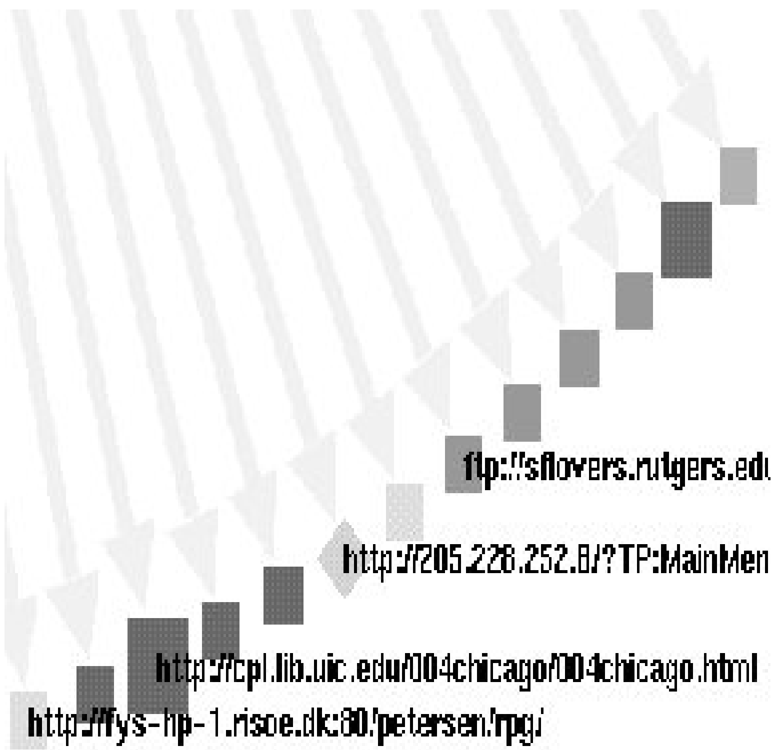

Figure 3. Node appearance details

MAGE software uses color, three dimensional representations and animation to help users see their data in different ways and develop new insight about their data.

In the Laboratoire d'Informatique de Paris 6, we developed a 3D visualization tool for large information hierarchies which are very difficult to use and represent. This tool is fitted to any type of hierarchical data; as explained in , we used XML and the DOM to design a "universal visualization tool". However these structures may not be completely hierarchical: there can also be non-hierarchical links, called "cross connections". We decided to investigate its use for topic maps visualization.

Our tool speeds access to relevant information and helps users find their way within the hierarchy. Fundamental factors for a good visualization interface are ( ):

An overview of the structure for a global understanding of the structure and of the relationships within the hierarchy,

The ability to zoom and to select some nodes,

Dynamic requests in order to filter data in real time.

Two visualizations are provided: a traditional 2D view (Figure 4) so as not to confuse users and a 3D view of the whole structure (Figure 5). 3D allows a more efficient use of screen space. In particular, links between nodes do not intersect. We used cone trees as 3D representation model. Nodes are spheres, cones or cylinders whose color changes according to their level in the hierarchy. Consistency between 2D and 3D views is achieved: when a node is selected in the global view, the corresponding line in the 2D tree is highlighted and node's properties are displayed - name, attributes and so on.

These visualizations are highly interactive; interesting nodes can be put in the foreground with zooms, translations and rotations. Users can delete irrelevant branches of the tree and expand interesting ones. They can also select specific elements and display them in separate 3D windows called detailed views.

However, these representation and navigation methods cannot solve the whole problem: there is still a large amount of information to display. Therefore, the initial hierarchy must be pruned with filtering and aggregating techniques. This tool was used to supervize telecommunication network equipment. Two filters were added to answer network managers' needs. They are particularly interested in elements where an alarm was activated; they need to know the consequences of these problems on connected ports. "Alarmed" nodes are stressed on the visualization and connections starting from these elements can be displayed, as shown on Figure 5. Different colors are used according to alarm types.

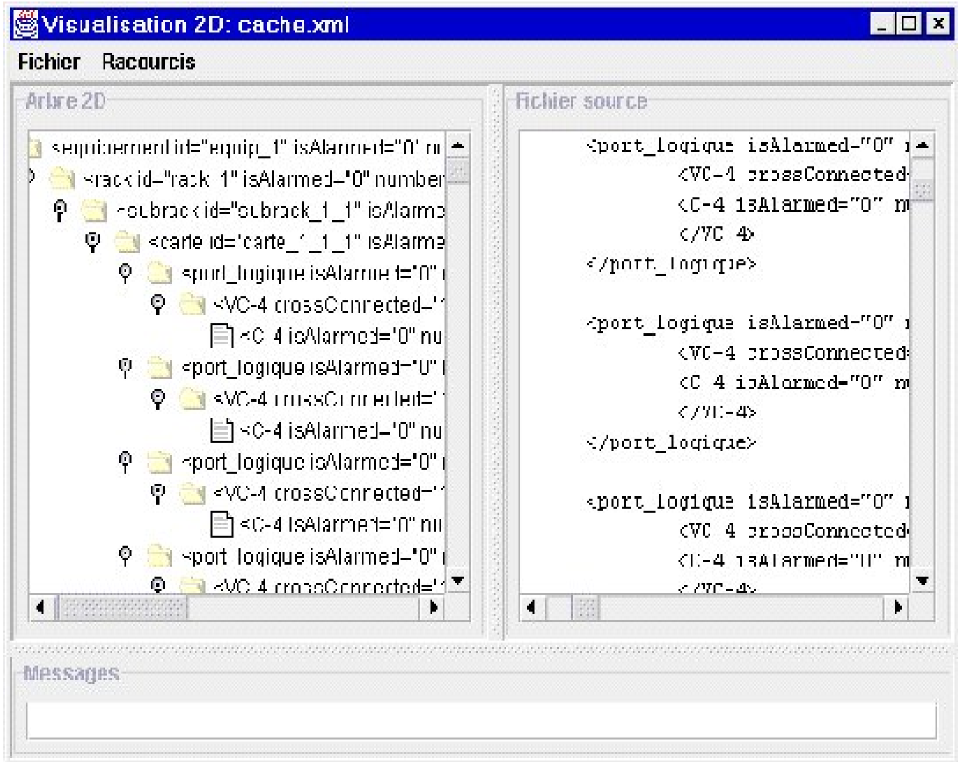

Figure 4. 2D tree and XML source file

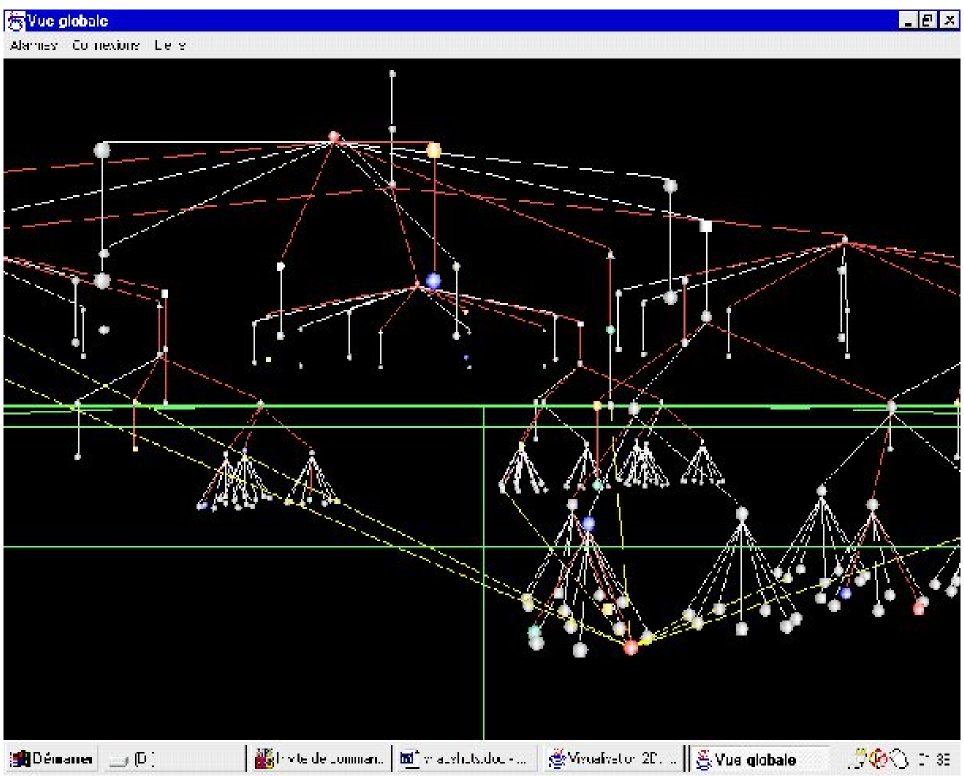

Figure 5. Global view with cross-connections visualization

Application to topic maps

We decided to use the tool we developed at LIP6 to visualize topic maps. In the following we will describe what modifications have to be made so as to adapt our tool to topic maps visualization.

Filtering techniques

Topic maps may contain millions of topics and associations. Therefore it is essential to select relevant information as it is impossible to display the whole data efficiently. Filtering techniques are needed in order to select and display only relevant information. Our tool enables users to filter topics and associations according to their name and/or type, provided that name and type are implemented as XML attributes, as explained in .

In the future, we will enhance our tool so that it can handle scope: users will specify which themes they are interested in and the tool will filter names, associations, etc. on that basis. This can be done with ontologies . Ontologies are widely used instruments for knowledge sharing and reuse; they may be applied to specify domain knowledge in a generic and consensual way. The key ingredients that make up an ontology are a vocabulary of basic terms. Each concept is associated to a term (i.e. a symbol) used to designate the concept, a description in natural language and a formal specification in an appropriate language such as KIF (Knowledge Interchange Format) . Terms are stored in dictionary-like structure. The dictionary is open-ended to allow the addition of new concepts.

Let us compare topic maps with geographical maps. You will never find a map of a country with the whole information about the country on it. There will be a topographical map, a political map, an economic map, etc. In the same way, topics and associations can be classified into different ontologies and different topic maps will be provided to the user according to his interest. If the user is interested in theatre, relevant topics are "play", "author", "tragedy", "culture", etc. This is a way of filtering information according to a specific scope.

Topics and associations representation

Once topics and associations are filtered, they need to be represented efficiently.

Topics

Topics are nodes and their type may be symbolized by different colors, shapes and textures. However, the number of different shapes, colors and icons is limited. Class hierarchies can be used to reduce topic types to a small number of "super-types", as stated in . In this case, we only need a specific shape and/or color for each super-type. Consider the following topic types: "artist", "painter" and "poet" ; they will look alike because they all derive from the super-type "person". In the same way, "was created by" and "was composed by" association types derive from "was caused by".

Currently the topic map standard does not define a standardized way of specifying type hierarchies. However this could be done in our tool - at the application level - by using a kind of stylesheet mechanism. This would allow users to specify which association types represent the supertype / subtype relationship in a particular topic map. Nevertheless, for navigation purpose in a topic map, a user may want to visualise hierarchies that have not been specified by the designer of the map.

We suggest another way of reducing the number of types to represent, by aggregating topics and associations with a classification algorithm. Galois lattices can group objects that share common properties automatically.

These groups are called classes. Therefore we only need to distinguish classes of topics in the representation instead of differentiating all topics. If we consider the example of topic map in Figure 1, "Paris", "Ile de France" and "France" will be represented the same way because they belong to the same class "location". In the same way, associations "is in" and "takes place in" are both "information about location". This classification mechanism makes it possible to display topic maps with different levels of details, as shown in Figure 6.

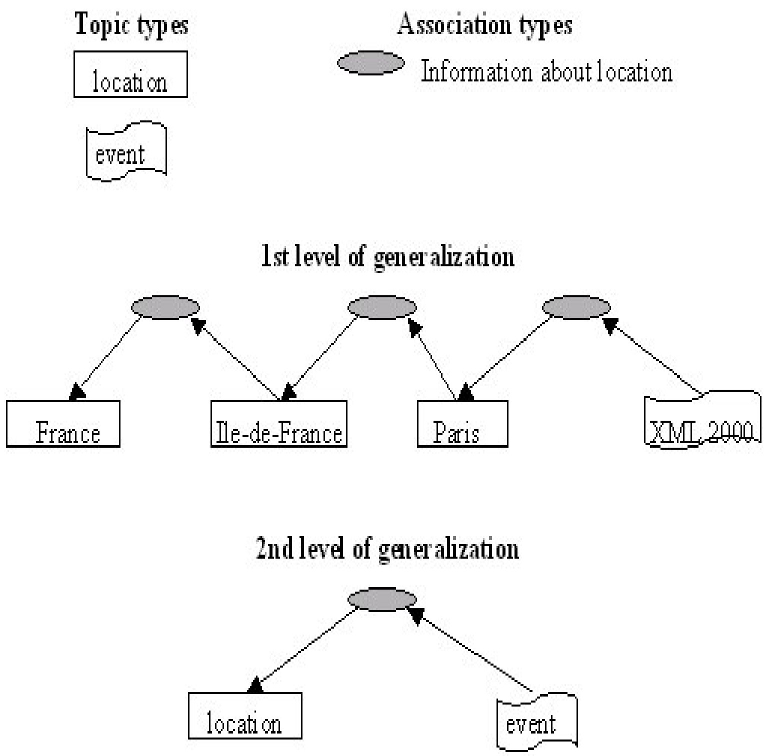

Figure 6. Level of detail in topic maps

Of course information displayed is less precise but this is acceptable for navigation purposes. However it is possible to display more precise information when users focus on specific parts of the topic maps. For example, if the mouse cursor is on "Paris", textual information appears on the screen that states Paris is a city (which is more precise than "location").

Associations

As far as associations are concerned, two solutions are possible: they may be represented as arcs or nodes.

In the first case associations are arcs and their type may be symbolized by the style of lines (full line, dotted line, etc.). However, this representation is limited to binary relationships, which means that n-ary relationships have to be decomposed into several binary relationships. This is possible but might add too much complexity to the resulting visualization.

On the other hand, associations can be represented the same way as topic - as nodes. More complex associations can be represented, in particular associations that involve more than two topics. In this case, associations types are distinguished the same way as topic types, as described in part 3.2.1.

A mouse-over event can display topics and associations names and types when the cursor is positioned over these elements.

Figure 7 is an example of topic maps visualization with our tool, in which associations are represented by nodes. Topics and associations are symbolized by different shapes according to their type. We use three different shapes in the current version of our prototype - cones, cylinders and spheres - but more are possible.

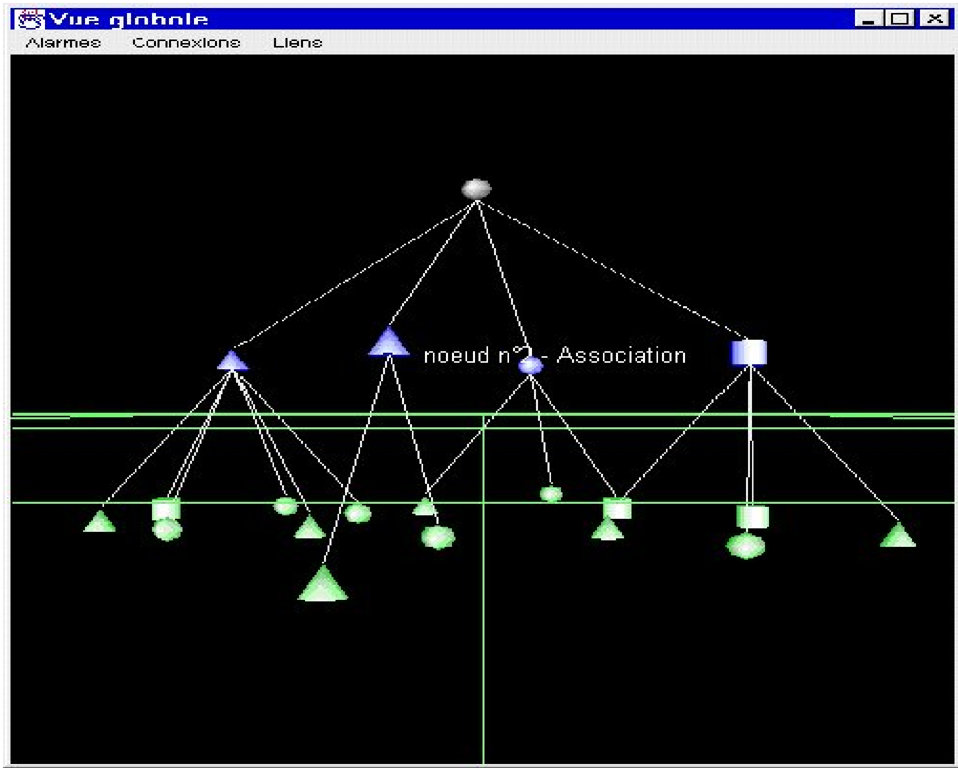

Figure 7. Example of topic map visualization. Associations = nodes

In Figure 8, associations are symbolized by arcs between topics.

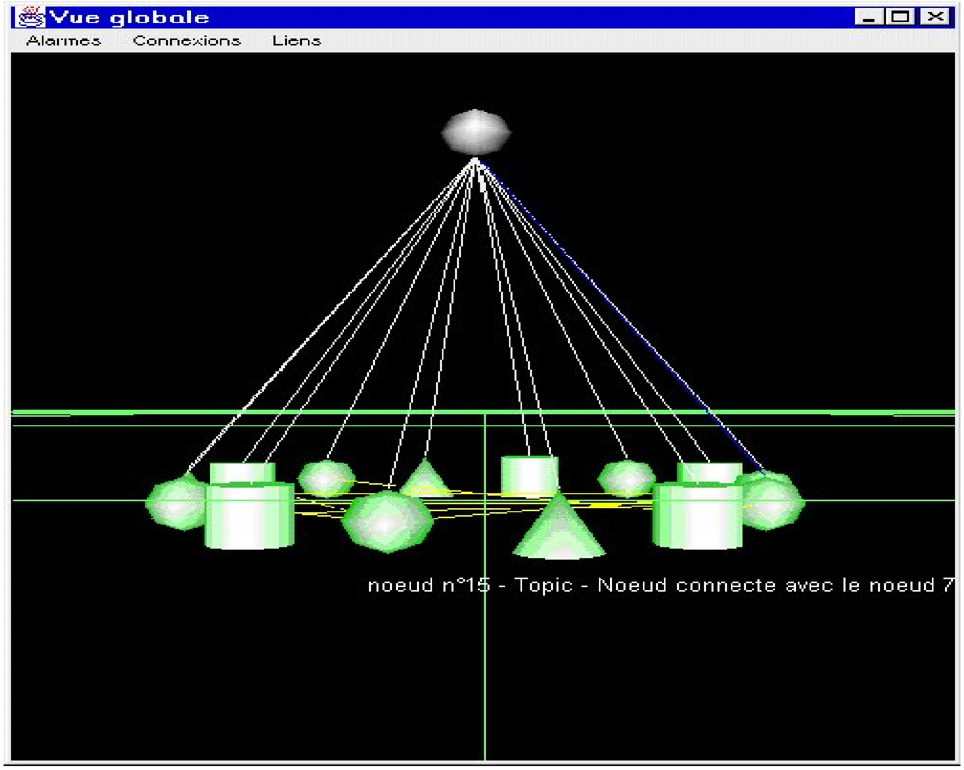

Figure 8. Example of topic map visualization. Associations = arcs

Conclusions and future work

In this paper, we investigated how the visualization tool we developed at LIP6 may be used for topic maps visualization. This prototype provides 3D interactive visualizations. Topic types can be distinguished with different colors and shapes, but we will add class hierarchies functionalities to reduce the number of types to represent. In the future, we will also handle scope, so as to allow users to specify their need for information more precisely.

Acknowledgments

We would like to thank Steve Pepper for his great help. He originated our interest in topic maps and his ideas and comments allowed us to complete this research.

Bibliography

[ISO 99] Information Technology-SGML Applications-Topic Maps, International Organization for Standardization, ISO-IEC 13250, Geneva, ISO, forthcoming.

[Desclefs 99] XML and Information Visualization - Application to Network Management, Desclefs B., Soto M., Markup Technologies 99, Philadelphia, USA, December 1999.

[Freeman 98] Exploring social structure using dynamic three-dimensional color images, Freeman L.C., Webster C. M. and Kirke D. M., Social Networks, 20, 1998, pp. 109-118.

[Genesereth 92] Genesereth M., Fikes R. et al., Knowledge Interchange Format Version 3 Reference Manual, Logic-92-1, Stanford University Logic Group, 1992.

[Godin 95] Incremental concept formation algorithms based on galois (concept) lattice, Godin R., Missaoui R. and Alaoui H., Computational Intelligence, 11(2), 1995.

[Guarino 95] Formal Ontology, Conceptual Analysis and Knowledge Representation, Guarino N., International Journal of Human and Computer Studies, volume 43 number 5-6, 1995.

[Kurnar] Visual Information for Network Configuration, Kurnar H., Plaisant C., Teittinen M., Schneiderman B., University of Maryland CS technical reports, Maryland, USA, June 1994.

[Pepper 99] Topic Maps: Introduction and Allegro, Pepper S., Rath H. H., Markup Technologies 99, Philadelphia, USA, December 1999.

[Robertson 91] Cone Trees: Animated 3D Visualizations of Hierarchical Information, Robertson G. G., Mackinlay J.D. and Card S., Proceedings of the ACM SICHI'91 Conference on Human Factors in Computing Systems, pp. 189-194, New Orleans, LA, USA, April 1991.

[Wills 99] NicheWorks - Interactive Visualization of Very Large Graphs, Wills G., Journal of Computational and Graphical Statistics, 8(2), pp. 190-212, June 1999.

[XML 98] Extensible Markup Language (XML) Specification Version 1.0, World Wide Web Consortium, 10 February 1998.

[DOM 98] Level 1 Document Object Model (DOM) Specification Version 1.0, World Wide Web Consortium, 20 July 1998.

Authors

Benedicte Le Grand

Laboratoire d'Informatique de Paris 6, Universite Pierre et Marie Curie
Postal Address:
8, rue du Capitaine Scott
Paris
France
Telephon: 33 1 44 27 75 12
Fax: 33 1 44 27 74 95
E-mail: Benedicte.Le-Grand@lip6.fr
Web: www-rp.lip6.fr/~blegrand

Benedicte Le Grand - was born in 1975. She received her engineer diploma from the Institut National des Telecommunications in 1997 and is currently a PhD student at LIP6 (Laboratoire d'Informatique de Paris 6) in the Network Department. Her research deals with Virtual Reality and its use for complex systems visualization. She is interested in markup technologies and especially XML; she was a speaker at ACM CIKM'99 (Conference on Information and Knowledge Management) in Kansas City and at Markup Technologies'99 in Philadelphia.

Michel Soto

Laboratoire d'Informatique de Paris 6, Universite Pierre et Marie Curie
Postal Address:
8, rue du Capitaine Scott
Paris
France
Telephon: 33 1 44 27 88 30
Fax: 33 1 44 27 74 95
E-mail: Michel.Soto@lip6.fr
Web: www-rp.lip6.fr/~ms

Michel Soto - is an associate professor at the LIP6 Laboratory (Laboratoire d'Informatique de Paris 6), University Pierre et Marie Curie, Paris, France. His main research interests are currently virtual worlds interoperability and complex systems visualization.