

Topic Maps Visualization

Bénédicte Le Grand, Laboratoire d'Informatique de Paris 6

Introduction

Topic maps provide a bridge between the domains of knowledge representation and information management. Topics and associations build a semantic network above information resources, which allows users to navigate at a higher level of abstraction.

However, this semantic network may contain millions of topics and associations, and users may still have problems to find relevant information; therefore, the issue of topic maps visualization and navigation is essential.

A topic map defines a multidimensional topic space: a topic has one or more names within a scope; it can also have occurrences and may play a role as a member of zero or more associations. Topics, associations and occurrence may have a type which is a topic, etc. A topic map is actually a high-dimensional knowledge base.

This chapter is organized as follows: first we will discuss topic maps visualization requirements. Then, we will review existing high-dimensional visualization techniques and study how/if they could be applied to topic maps representation. We will conclude by giving a few directions that could lead to the “ultimate” topic map visualization tool.

1. Topic maps visualization requirements

Topic maps are very powerful as they allow to organize information, but they may be very large. Intuitive visual user interfaces may significantly reduce the cognitive load of the user when working with these complex structures. Visualization is a promising technique for both enhancing users' perception of structure in large information spaces and providing navigation facilities. It also enables people to use a natural tool of observation and processing – their eyes as well as their brain – to extract knowledge more efficiently and find insights [3].

Users needs must be clearly identified in order to design useful visualizations. Before presenting visualization requirements, we need to study how topic maps are used.

1.1. Different uses of topic maps

We distinguish two kinds of uses of a topic map :

- if the user has a specific question, query languages (such as TMQL [5]) are well suited. They take the relationships between objects into account, which allows them to find more accurate answers than traditional search engines – which only consider the occurrences of words.

- if the user has no precise goal of interest and wants to explore the topic map, she needs to have an overview of the topic map, so as to know where to start her navigation.

The first type of information retrieval does not require specific visualizations; textual interfaces may suffice. The second type of information retrieval is more complex as the subject of interest is not clearly defined. In this case, users may be compared to tourists in a city they visit for the first time. They need to know what they should see and how to get to these different places quickly. They may want either to follow a guide or to explore by themselves.

We can see from this example that there are two kinds of requirements: representation and navigation. A good representation helps users identify interesting spots whereas an efficient navigation is essential to access information rapidly.

1.2. Visualization requirements

As stated before, both representation and navigation are essential in a good visualization. According to Schneiderman, "the visual information -seeking mantra is: overview first, zoom and filter, then details on-demand".

1.2.1. Representation requirements

First of all, we need to provide the user with an overview of the topic map. This overview must show the main features of the structure in order to allow us to deduce the topic map's main characteristics at a glance. Visual representations are particularly fitted to these needs, as they exploit human abilities to detect patterns.

The first thing we need to know about a topic map is what it deals with, i.e. what its main concepts are. Once they are identified, we need more structural information, such as the generality or specificity of the topic map. This kind of information should appear clearly on the representation so as to help users compare different topic maps quickly and only explore the most relevant ones in detail. The position of topics on the visual display should reflect their semantic proximity. These properties can be deduced from the computation of metrics on the topic map [7].

Topic maps are multidimensional knowledge bases. We need to represent topics and the relationships between these topics: associations. Topics have many characteristics, such as names, types, occurrences, roles as members of associations, all of which depend on a context – called scope. All these attributes should appear in the visualization.

The requirements we stated before are not compatible, as it is not possible – neither relevant – to display simultaneously general information and details. We can compare this with a geographic map; a map of the world cannot – and should not – be precise. If a user is interested in details, she must precise her interest, for example choose a specific country. As in geographical maps, we need to provide different scales in topic maps representations.

Moreover, visualizations should be dynamic to adapt to users' needs in real time; combinations of time and space can help to ground visual images in one's experience of the real world and so tap into the users' knowledge base and inherent structures.

To sum up the visual display requirements, we need to represent the whole topic map in order to help users understand it globally. This overview should reflect the main properties of the structure. However, users should be able to focus on any part of the topic map and see all the dimensions they need. Providing these several scales requires the use of different levels of detail. Finally, the representation should be updated in real time to enable user interaction.

1.2.2. Navigation requirements

A good navigation allows users to explore the topic map and to access information quickly. Free navigation should be kept for small structures or expert users as the probability of getting lost is very high. For beginners, predefined navigation paths are preferable until topics of interest are identified.

Navigation should be intuitive so that it is easy to get from one place to another. Several metaphors are possible: users may travel by car, by plane, by metro or may simply be “teleported” – as on the Web - to their destination. The differences lie in what they see during their journey. From a car, they see details, from a plane they have an overview of the city... Navigation is essential because it helps users build their own cognitive map – a map-like cognitive representation of an environment - and increase the rate at which they can assimilate and understand information.

We have shown that navigation needed to be intuitive and that there should be constraints for beginner users whereas expert users may explore the structure freely. In the following section, we will study which visualization techniques meet our requirements and may be used to represent topic maps.

2. Visualization techniques

First we will review current topic maps visualizations; then we will present visualization techniques used for more general complex structures.

2.1. Current topic maps visualizations

Several topic maps engines provide visualizations of topic maps. Most of them display lists or indexes from which it is possible to select a topic and see related information. This representation is very convenient when users' needs are clearly identified. The navigation is usually the same as on Web sites, i.e. users click on a link to open a new topic or association. An example of such a visualization is provided by the Ontopia Navigator, as shown on Figure 1.

Apart from these browser-based navigations, other types of visualizations are currently available. Empolis K42 displays topic maps as hyperbolic trees (Figure 2); Mondeca's Topic Navigator [9] builds graph representations in real-time, according to what users are allowed or need to see (Figure 3). Figure 4 is an example of TM4J's [14] topic map dynamic visualization using TheBrain software [13]. A 3D interactive topic map visualization tool – UNIVIT (Universal Interactive Visualization Tool) - is proposed in [6]. The example shown on Figure 5 was drawn with UNIVIT, which uses virtual reality techniques such as 3D, interaction and different levels of detail.

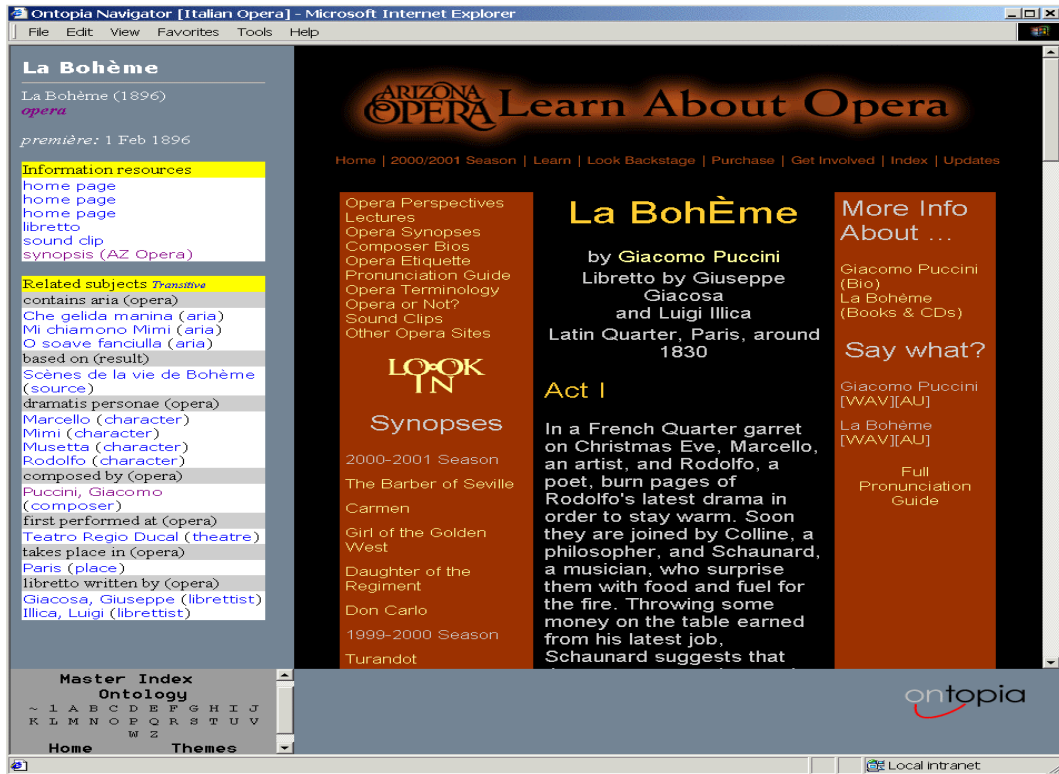


Figure 1. Ontopia Navigator

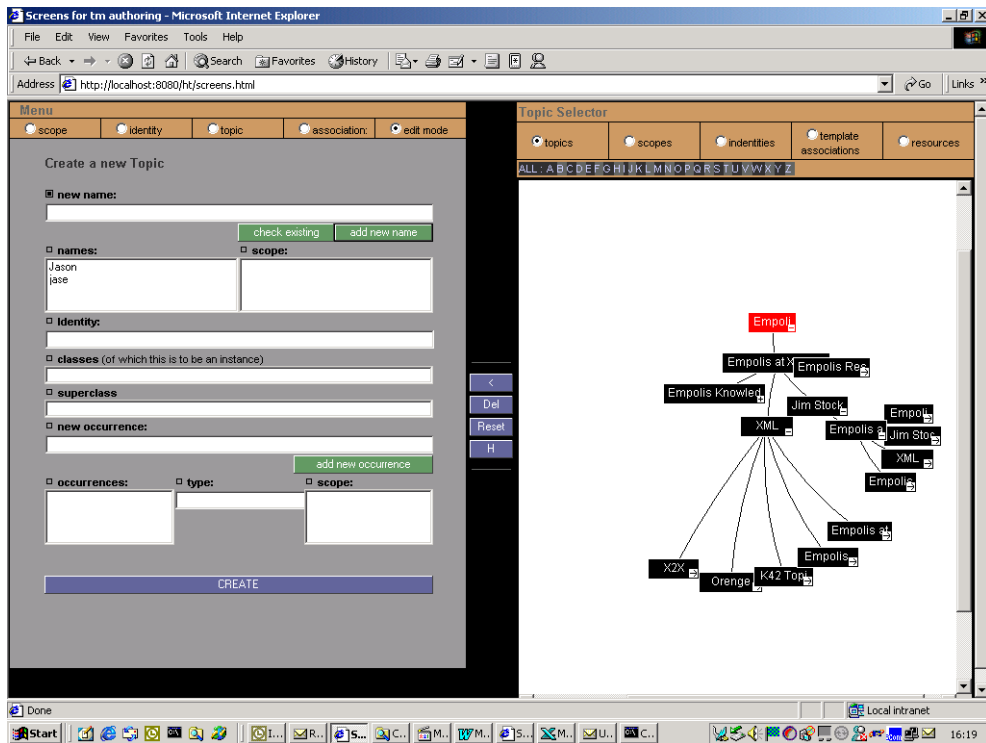


Figure 2. Empolis K42 TMV Application

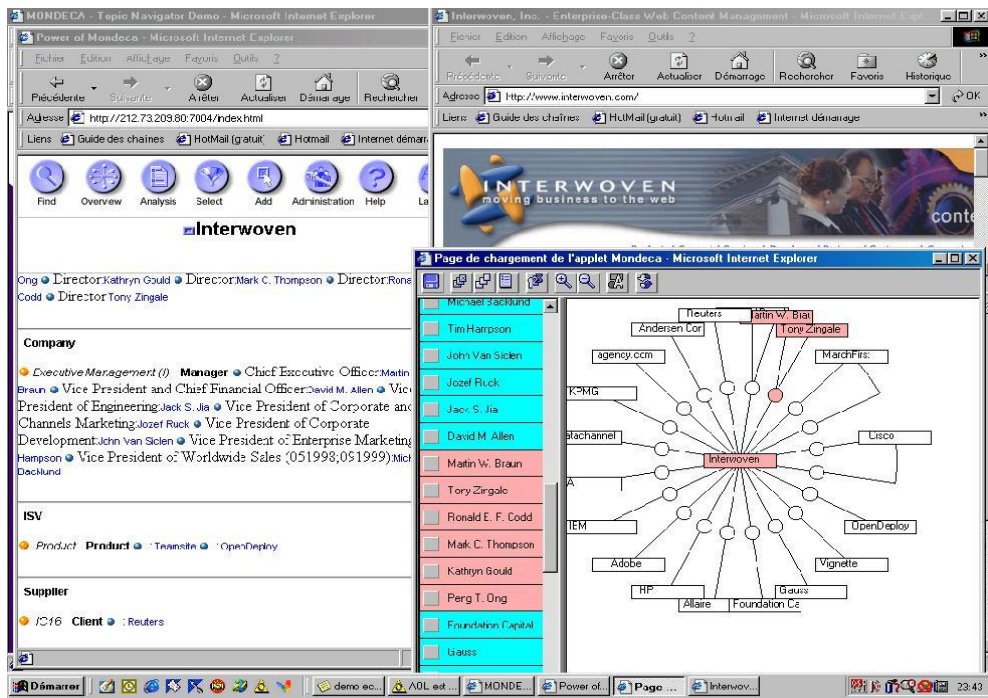


Figure 3. Mondeca's Topic Navigator

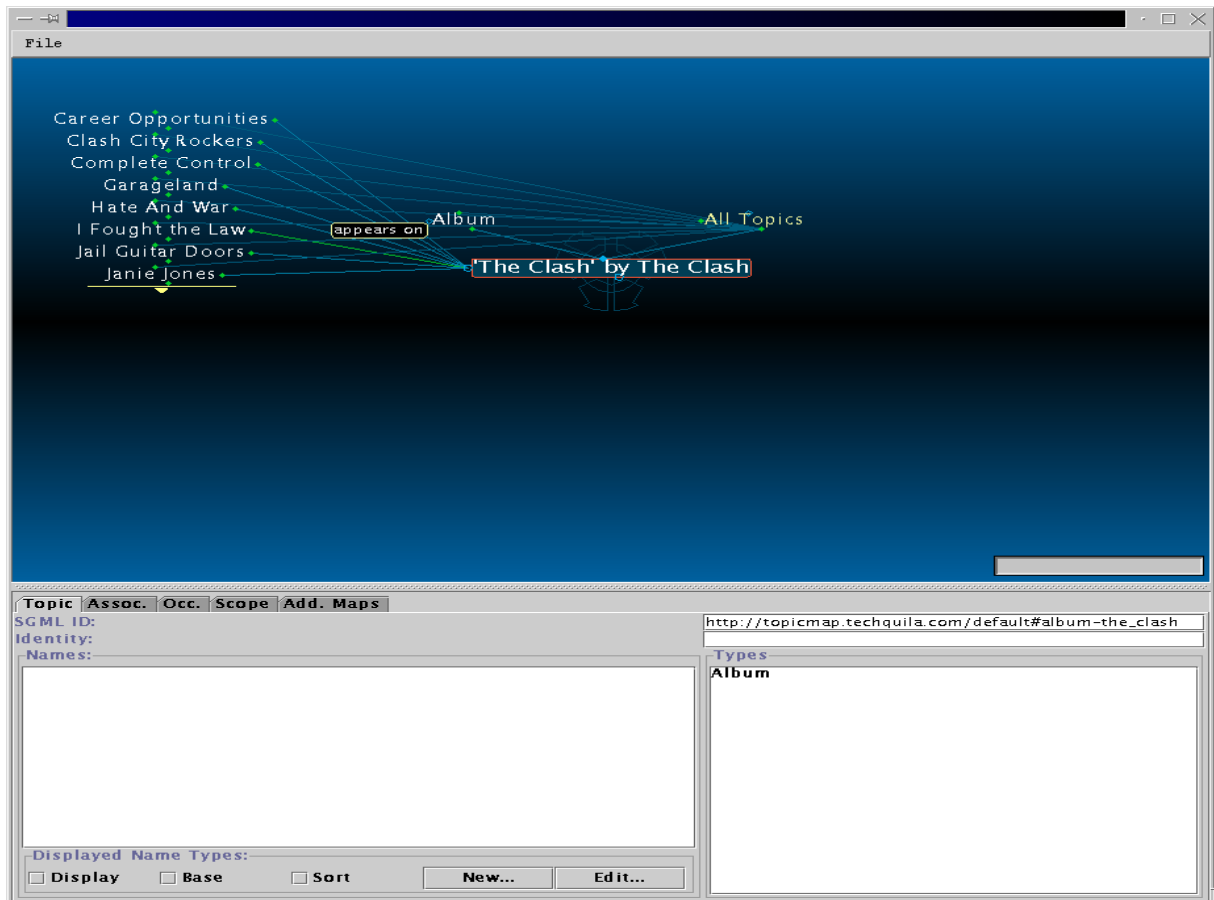


Figure 4. TM4J's topic map visualization with TheBrain software

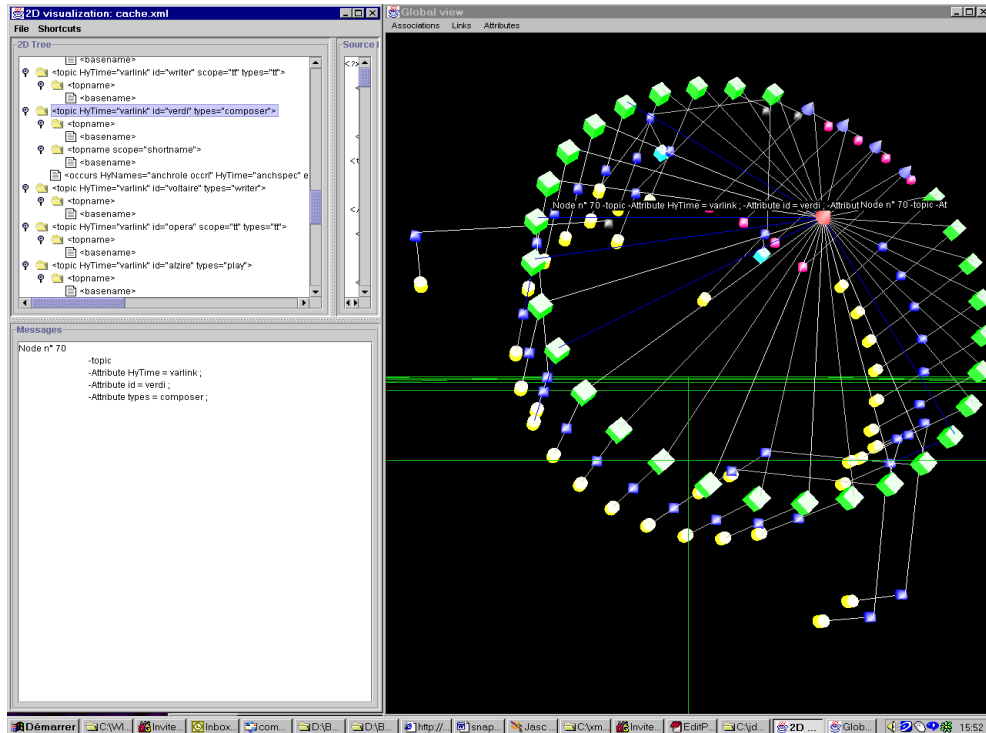


Figure 5. 3D interactive topic map visualization with UNIVIT

2.2. General visualization techniques

Most of the graphical representations described before are based on graphs or trees. Advantages and drawbacks of this type of representation will be studied in this section, as well as other visualization techniques – currently used in other contexts – which may be adapted to topic maps.

2.2.1. Graphs and trees

Topic maps can be seen as a network of topics, so networks/graphs visualizations techniques may be interesting.

Graphs and trees are suited for representing the global structure of the topic map. However, trees are better understood by human beings since they are hierarchical. Trees are easy to interpret. Topic maps are not hierarchies and thus may not be – directly – represented as trees. However, it may be interesting to transform small parts of the topic map into trees. By doing so on a little part of the topic map (to avoid clutter), we may benefit from the advantages of trees.

Techniques such as hyperbolic geometry [10] (Figure 6) allows to display a very large number of nodes.

Figure 6. Example of graph in a 3D hyperbolic space

Efficient node positioning makes it possible to intuitively derive information from the distance between nodes. For instance:

- topics linked together by an association may be represented close to each other in the graph.
- topics of the same type or pointing to the same occurrences may be clustered.

Graphs and trees meet our first requirement since they may represent the whole topic map. However, the representation may become cluttered rapidly as the number of topics and associations increases.

Our second requirement, which consists in representing all the different parameters of a topic map (name, type, scope, etc.), may be really challenging. Figure 7 is a graph obtained with GraphVisualizer3D (now NV3D) [11]. Different shapes and colors are used to symbolize various dimensions of nodes and arcs of the graph. This kind of graph may be used to visualize a topic map; topics would be nodes and associations arcs. However, the number of different shapes, colors, icons and textures is limited. This representation is not suited for a topic map containing millions of topics and associations.

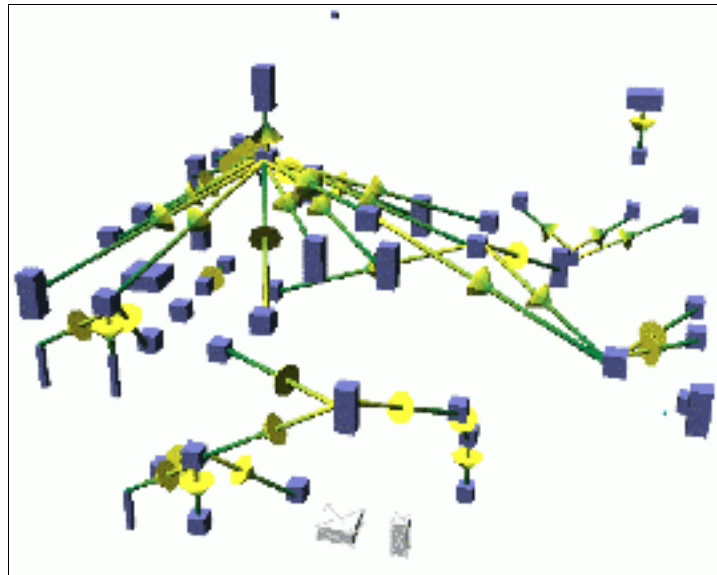


Figure 7. GraphVisualizer 3D



Figure 8. MineSet Tree Visualizer

The MineSet Tree Visualizer [12] displays hierarchical data structures in a 3D landscape, revealing quantitative and multidimensional characteristics of data. Utilizing a fly-through technique, users view data as visual representations of hierarchical nodes and associations. Users explore data with any level of detail or summary, from a bird's-eye perspective down to detailed displays of source data.

2.2.2. Maps

Topic maps are designed to enhance navigation in complex information systems, therefore representing them as maps seems natural.

NicheWorks [15] provides a schematic representation, which is illustrated on Figure 9.

Finding structures in vast multidimensional data sets, be they measurement data, statistics, or textual documents, is difficult and time-consuming. Interesting, novel relations between the data items may be hidden in the data. The self-organizing map (SOM) [4] algorithm of Kohonen can be used to aid the exploration: the structures in the data sets can be illustrated on special map displays. When applied to the mapping of documents, this algorithm automatically organizes the documents onto a two-dimensional grid so that related documents appear close to each other, as shown on Figure 11. This representation is suited to reflect the structure through an efficient node positioning but fails at displaying associations and topic maps multiple dimensions.

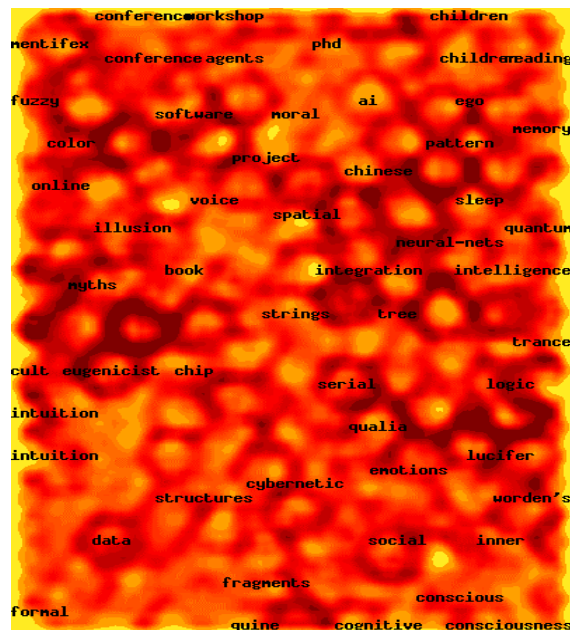


Figure 11. Self-organizing map

ThemeScape [1] provides different types of maps. They look like topographical maps with mountains and valleys, as shown on Figure 12. The concept of the layout is simple: documents with similar content are placed closer together, and peaks appear where there is a concentration of closely related documents.

Peaks represent concentrations of documents about a similar topic. Higher numbers of documents create higher peaks. The valleys between peaks can be interesting because they contain fewer documents and more unique content.

Topic Labels reflect the major two or three topics represented in a given area of the map, providing a quick indication of what the documents are about. Additional labels often appear when we zoom into the map for greater detail. We can zoom to different levels of magnification to declutter the map and reveal additional documents and labels.

Figure 12. ThemeScape map

This visualization is very interesting since it combines different representations in several windows. Users may choose one of them according to the type of information selected.

2.2.3. Virtual worlds and multi-dimensional representations

Visual data mining tools depict original data or resulting models using 3D visualizations, enabling users to interactively explore data and quickly discover meaningful new patterns, trends, and relationships.

Visual tools may utilize animated 3D landscapes that take advantage of a human's ability to navigate in three-dimensional space, recognize patterns, track movement, and compare objects of different sizes and colors. Users may have complete control over the data's appearance.

The concept of Populated Information Terrains (PITS) [8] aims at extending database technology with key ideas from the new fields of Virtual Reality (VR) and Computer Supported Cooperative Work (CSCW [8]). A PIT is defined as a virtual data space that may be inhabited by multiple users. The underlying philosophy of PITS is that they should support people in *working together within data* as opposed to merely *with data*. PITS may be seen both as a means of improving the way in which users browse and interact with data and also as a means of actively supporting cooperative sharing. Users may appear on the visualization; their representations are called avatars, as shown on Figure 13.



Figure 13. Virtual world featuring several avatars

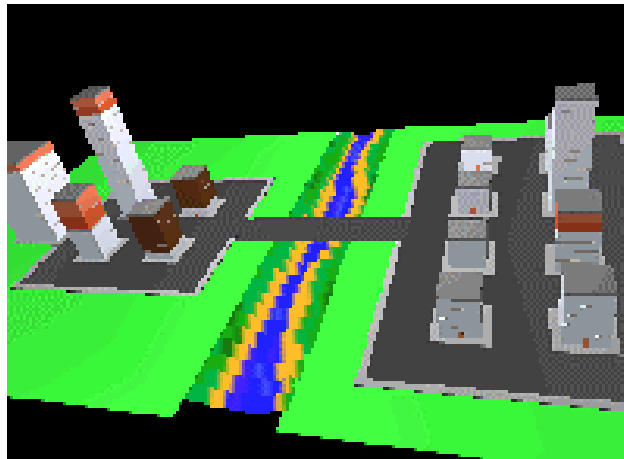


Figure 14. Software City

Figure 14 [16] illustrates the use of a spatial metaphor. Topic maps may be represented as cities in which topics would be buildings and associations streets, bridges etc. Topics and associations related to the same scope could belong to the same district. The multiple dimensions of a topic map can be represented with this metaphor.

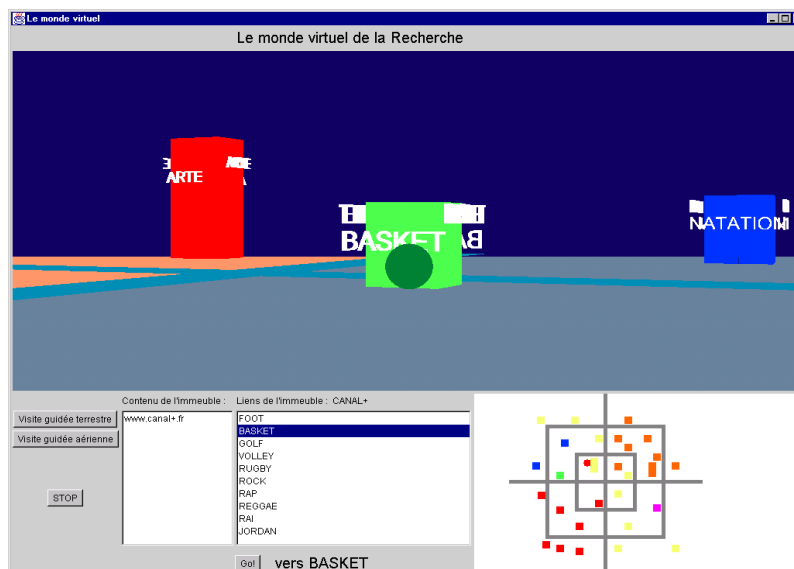


Figure 15. Virtual city and 2D map

One representation of topic maps as virtual cities [7] is shown on Figure 15. Topics are represented as buildings which coordinates are computed from a matrix of similarities between topics. Users may navigate freely or follow a guided tour through the city ; they may also choose to walk or fly. The properties of topics are symbolized by the characteristics of the corresponding buildings, such as name, color, height, width, depth, etc. Occurrences and associated topics are displayed in two windows at the bottom of the screen. As human beings are used to 2D, a traditional 2D map is also provided and the two views – the map and the virtual city – are always consistent.

Representing topic maps as populated virtual worlds may help users work collectively. Virtual reality techniques include interactivity and the use of different levels of detail (LOD). Immersion in virtual worlds makes users feel more involved in the visualization. They may explore the world and interact with data. However, they may get lost in the virtual world. In order to avoid these problems, predefined navigation paths should be proposed. The LOD makes it possible to display many scales: details appear only when the user is close to the subject of interest.

Conclusion

In this chapter, we reviewed several types of information visualization. Some may represent efficiently the global structure while others are better at displaying details or providing interaction with data. It seems that the ultimate topic map visualization tool should benefit from all these advantages by combining several techniques. This can be done by displaying several windows or by selecting the most appropriate representation at a given level of detail.

One visualization tool is usually adapted to display a certain amount of data. If only one tool is to be used for visualization, the topic map may be clustered to reach the scale at which the tool is useful.

References

- [1] Cartia Inc., ThemeScape Product Suite, <http://www.cartia.com/products/index.html>
- [2] Chen, H., Schuffels, C., Orwig, R., *Internet Categorization and Search: A Self-Organizing Approach*, Journal of Visual Communication and Image Representation, Special Issue on Digital Libraries, Volume 7, Number 1, pp 88-102, 1996.
- [3] Gershon, N., Eick, S.G., *Visualization's New Tack: Making Sense of Information*, IEEE Spectrum, Nov. 1995, pp. 38-56.
- [4] Kaski, S., Honkela, T., Lagus, K., Kohonen, T., *WEBSOM – self_organizing maps of document collections*, Neurocomputing, volume 21, pp 101-117, 1998.
- [5] Ksiezyk, R., *Answer is just a question [of matching Topic Maps]*, XML Europe 2000, Paris, France, June 2000.
- [6] Le Grand, B., Soto, M., *Information management – Topic maps v isualization*, XML Europe 2000, Paris, France, June 2000.
- [7] Le Grand, B., Soto, M., *Topic Maps Metrics*, Knowledge Technologies 2001, Austin, Texas, March 2001.
- [8] Mariani, J., Benford, S., *Populated Information Ter rains: Virtual Environments for Data Sharing and Visualization*, CSCW, 1994.
- [9] Mondeca, Topic Navigator, <http://www.mondeca.com/site/products/products.html>

- [10] Munzner, T., *H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space*, IEEE Symposium on Information Visualization, 1997.
- [11] Nvision Software Systems Inc., *NV3D Technical Capabilities Overview*, <http://www.nv3d.com/html/tco.pdf>
- [12] Silicon Graphics Inc., *Mineset 3.0 Datasheet*, <http://www.sgi.com/software/mineset/Mineset3.0.pdf>, 1999.
- [13] TheBrain Technologies Corp., <http://www.thebrain.com>
- [14] TM4J, A Topic Map Engine For Java, <http://www.techquila.com>
- [15] Wills, G. J., NicheWorks – Interactive Visualization of Very Large Graphs, GD'97, Rome, Italy, September 1997.
- [16] Young, P., 3D Software Visualization – Visualizations and Representations, Internal Report, Visualization Research Group, University of Durham, 1997.