Interactive Visual User Interfaces to Databases

Tugba Taskaya, Pedro Contreras, Tao Feng and Fionn Murtagh
School of Computer Science, Queen's University Belfast
Belfast BT7 1NN, Northern Ireland
f.murtagh@qub.ac.uk

Abstract

We review past achievements, and describe current work, on interactive and responsive visual user interfaces to databases. Such visual user interfaces organize data and information, and also provide interaction with the user. They can be considered as a particular type of agent, helping in the human tasks of information navigating, filtering, seeking, and accessing.

1 Information Clustering and User Interfaces

Information retrieval by means of "semantic road maps" was first detailed in Doyle (1961). The spatial metaphor is a powerful one in human information processing and lends itself well to modern distributed computing environments such as the web. The Kohonen self-organizing feature map (SOM) method is an effective means towards this end of a visual information retrieval user interface.

The Kohonen map is, at heart, k-means clustering with the additional constraint that cluster centers be located on a regular grid (or some other topographic structure) and furthermore their location on the grid be monotonically related to pairwise proximity (Murtagh and Hernández-Pajares, 1995). The nice thing about a regular grid output representation space is that it easily provides a visual user interface. In a web context, it can easily be made interactive and responsive.

Figure 1 shows a visual and interactive user interface map, using a Kohonen self-organizing feature map. Color is related to density of document clusters located at regularly-spaced nodes of the map, and some of these nodes/clusters are annotated. The map is installed as a clickable imagemap, with CGI programs accessing lists of documents and – through further links – in many cases, the full documents. Such maps are maintained for 13000 articles from the *Astrophysical Journal*, 8000 from *Astronomy and Astrophysics*, and over 2000 astronomical catalogs. More information on the design of this visual interface and user assessment can be found in Poinçot et al. (1998, 1999, 2000). For maps in operational use, see:

- A&A: http://simbad.u-strasbg.fr/A+A/map.pl,
- ApJ: http://simbad.u-strasbg.fr/ApJ/map.pl,
- VizieR: http://vizier.u-strasbg.fr/viz-bin/VizieR.

In Guillaume (2000a, 2000b) we developed a Java-based visualization tool for hyperlink-rich data in XML, consisting of astronomers, astronomical object names and article titles. It was open to the possibility of handling other objects (images, tables, etc.). Through weighting, the various types of links could be prioritized. An iterative refinement algorithm was developed to map the nodes (objects) to a regular grid of cells, which as for the Kohonen SOM map, are clickable and provide access to the data represented by the cluster. Figure 2 shows an example for an astronomer (Prof. Jean Heyvaerts, Strasbourg Astronomical Observatory). Given the increasingly central role of XML, the importance of such clustering for data organization and as a basis for knowledge discovery cannot be underestimated. This map is Java-based and client-side.

These new cluster-based visual user interfaces are not unduly computationally demanding, if we assume that they can be set up in advance of use and, if required, periodically updated. Illustrations of processing one million newsgroup messages, and discussion of processing 7 million patent abstracts, can be found on the WebSOM server, http://websom.hut.fi. Further results are available in Oja and Kaski (1999).

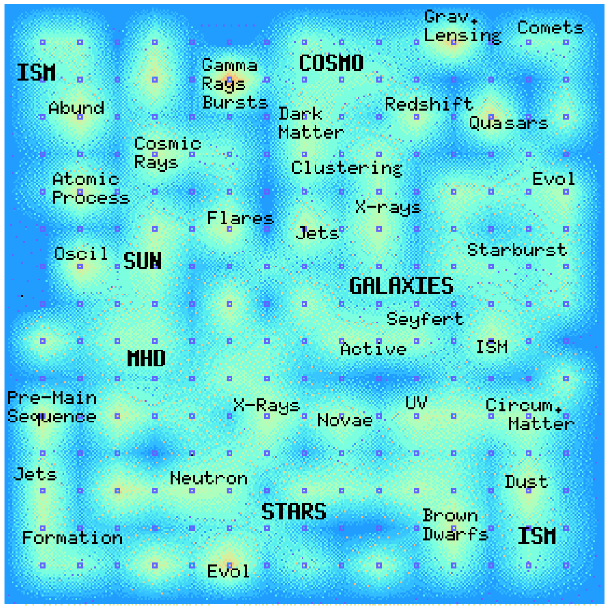

Figure 1: Visual interactive user interface to the journal *Astronomy and Astrophysics* based on 3000 published articles.

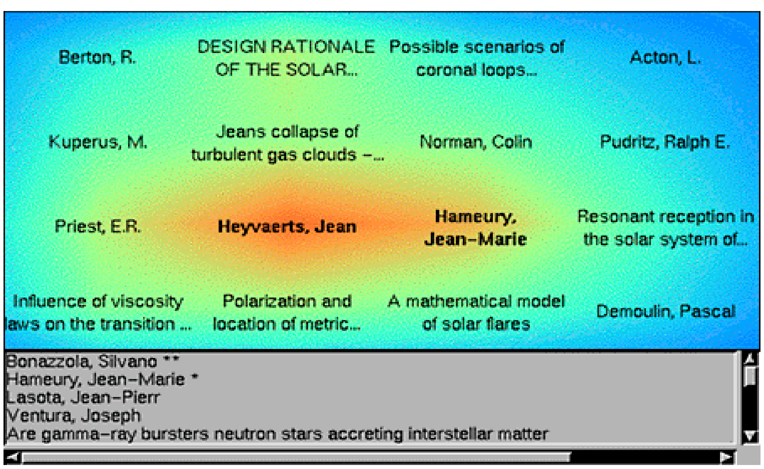

Figure 2: Visual interactive user interface, based on graph edges. Vertices are author names, article titles and (not shown here) astronomical object names. Map for astronomer Jean Heyvaerts.

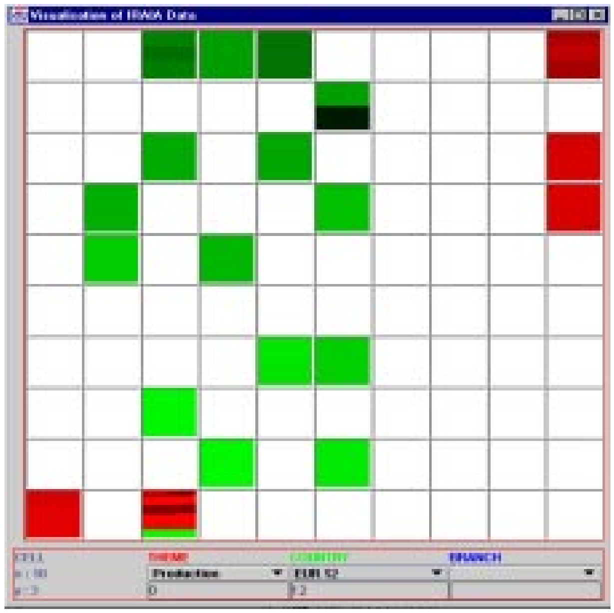

Figure 3: Visual interface to more than 150,000 economic time series. Categories of “countries” and “themes” used.

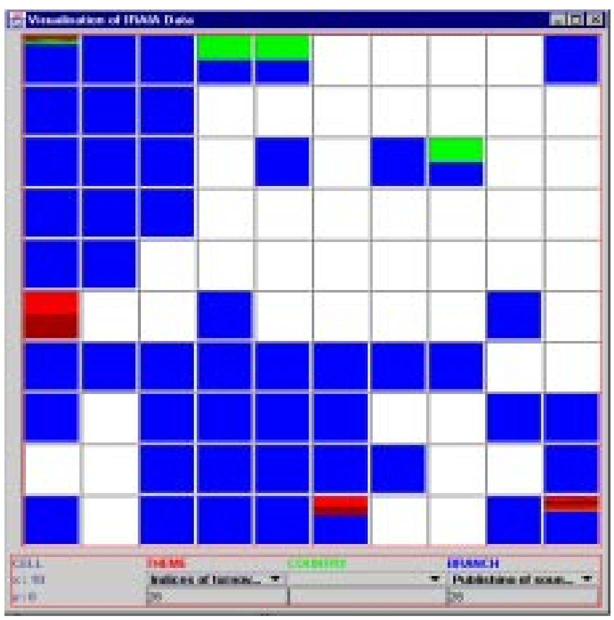

Figure 4: Visual interface to more than 150,000 economic time series. Categories of “countries”, “themes” and “branches” used.

2 Input Data for Maps of Information Spaces

We can distinguish between the following types of input for maps of information spaces.

- Keyword-based: the bibliographic maps exemplified in Figure 1 are of this type. The keywords or index terms provide the dimensions of a geometric space in which our objects are located.
- Sparse graph: this was the case for the example discussed in Figure 2. This is highly likely to be the case whenever XML XLink functionality is used as the basis for associations between our objects.
- Dense graph: this is the case for database occupancy, visual user interfaces for which are illustrated in Figures 3 and 4.

In the dense grap case, a convenient way to process the data is to take the dense graph of interdependencies and map or project the objects, using these interdependencies, into a geometric space. We will briefly look at how this may be done.

Principal coordinates analysis is very similar to principal components analysis (PCA). Rather than the usual objects $\times$ variables array (e.g., documents crossed by index terms), we are given an objects $\times$ objects distance or dependency matrix. A minimal amount of alteration to the approach adopted in PCA allows this type of input data to be handled. Principal coordinates analysis has also been referred to as *classical multidimensional scaling* and *metric scaling* (Torgerson, 1958; see also the short description in Murtagh and Heck, 1987).

We investigated both direct use of principal components analysis, and principal coordinates analysis, on database occupancy data. The principle was the same in both cases: project similarity data into a coordinate space, and use between 2 and 8 best-fitting coordinates (associated with the highest eigenvalues) to characterize the objects. These objects were "countries", economic "themes" and economic "branches". A Kohonen map was then constructed from this data.

Figure 3 considers "theme" and "country" database occupancy frequencies. A color coding is used: themes in red, country in green, and branch in blue. In one grid cell, there can be more than one object. The color intensity varies with density. Also one grid cell is divided into several segments to reflect the different types of object associated with it. In Figure 4, "theme", "country" and "branch" database occupancy frequencies are shown. Clicking on the grid cell give the associated information – a set of economic time series from an OECD and Eurostat (Statistical Office of European Union) database of over 150,000 time series.

3 Conclusion

We have looked at new, closely-related technologies in the area of human-computer interaction. Visual and responsive user interfaces have much promise. The work on visualizing database occupancy which we have described is now being extended to input data derived from relevant documentation. This work is being pursued in the context of the European Fifth Framework IRAIA project, "Getting orientation in complex information spaces as an emergent behavior of autonomous information agents" (http://iraia.diw.de).

References

1. L.B. Doyle, "Semantic road maps for literature searchers", *Journal of the ACM*, 8, 553–578, 1961.
2. D. Guillaume and F. Murtagh, "Clustering of XML documents", *Computer Physics Communications*, 127, 215–227, 2000.
3. D. Guillaume, "Distributed Information Retrieval, Search and Processing in Astronomy", PhD thesis, Université Louis Pasteur, 2000b.
4. F. Murtagh and A. Heck, *Multivariate Data Analysis*, Kluwer, 1987.

5. F. Murtagh and M. Hernández-Pajares, "The Kohonen self-organizing map method: an assessment", *Journal of Classification*, 12, 165–190, 1995.

6. E. Oja and S. Kaski, *Kohonen Maps*, Elsevier, 1999.

7. P. Poinçot, F Murtagh and S Lesteven, "A spatial user interface to the astronomical literature", *Astronomy and Astrophysics Supplement Series*, 130, 183–191, 1998.

8. P. Poinçot, "Classification et recherche d'information bibliographique par l'utilisation des cartes auto-organisatrices, applications en astronomie", PhD thesis, Université Louis Pasteur, 1999.

9. P. Poinçot, F. Murtagh and S. Lesteven, "Maps of information spaces: assessments from astronomy", *Journal of the American Society for Information Science*, 51, 1081–1089, 2000.

10. W.S. Torgerson, *Theory and Methods of Scaling*, Wiley, New York, 1958.