



Interactive Visual User Interfaces: A Survey

F. MURTAGH¹, T. TASKAYA², P. CONTRERAS¹, J. MOTHE³ &
K. ENGLMEIER⁴

¹*School of Computer Science, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland (E-mail: f.murtagh@qub.ac.uk);* ²*Department of Electronics and Computer Engineering, University of Surrey, England;* ³*IRIT, 118 route de Narbonne, 31062 Toulouse, France;* ⁴*DIW, Königin-Luise-Str. 5, 14195 Berlin, Germany*

Abstract. Following a short survey of input data types on which to construct interactive visual user interfaces, we report on a new and recent implementation taking concept hierarchies as input data. The visual user interfaces express domain ontologies which are based on these concept hierarchies. We detail a web-based implementation, and show examples of usage. An appendix surveys related systems, many of them commercial.

Keywords: cartographic representation, display, graphical user interfaces, human-machine interfaces, information retrieval, Kohonen map, visualization

1. Visualization-based User Interfaces: Introduction

1.1. Motivation

Visualizing information and data stored in databases or in unstructured or semi-structured repositories is important for the following reasons:

1. It allows the user to have some idea before submitting a query as to what type of outcome is possible. Hence visualization is used to summarize the contents of the database or data collection (i.e., information space).
2. The user's information requirements are often fuzzily defined at the outset of the information search. Hence visualization is used to help the user in their information navigation, by signalling related items, by showing relative density of information, and by inducing a (possibly fuzzy) categorization on the information space.
3. Visualization can therefore help the user before the user interacts with the information space, and during this interaction. It is a natural enough progression that the visualization becomes the user interface.

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.

1.2. *Kohonen map*

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). Kohonen maps lack precision and algorithmic clarity relative to well-defined and well-understood low-dimensional mapping methods such as principal components analysis (PCA), correspondence analysis, or multidimensional scaling.

A brief characterization of these alternative methods follows. Murtagh and Heck (1987) can be referred to for examples and references. PCA projects multidimensional (or vector-valued) inputs into a low-dimensional space, which in practice is usually the best-fitting plane. PCA is implemented using linear algebra, where the eigenvectors of the covariances or correlations among input observation vectors serve to define a new coordinate system. The Kohonen map achieves a similar result through iterative optimization. Importantly the Kohonen map output is highly constrained: rather than a continuous plane, it is instead (nearly always) a regular grid. Correspondence analysis is very similar to PCA. While it is particularly suitable for frequency or categorical data (latter: e.g., frequencies of occurrence in a set of discrete categories), PCA is very suitable for real-valued or continuously-valued input data. Like PCA and correspondence analysis, multidimensional scaling also targets a best-fitting low-dimensional space (e.g., best planar fit, or best coordinate fit). Multidimensional scaling takes all possible ranks as input, and owes its origins to application domains where ranks are easier to define compared to more precise measurement. Examples of where ranks are more easily obtained include perceptual studies in psychology, or aptitude studies in education.

Included in the causes of what we characterize informally as lack of precision in the Kohonen map are the following: the specific criterion being iteratively optimized in training, the existence of local optima, dependence on starting configuration, and lack of interpretation characteristics such as meaningfulness of output representational space axes. However, as stressed in Murtagh and Hernández-Pajares (1995), a regular grid output representation space offers an important advantage in that it easily provides a visual user interface. In a web context, it can be made interactive and responsive.

Illustrations of processing one million newsgroup messages, and discussion of processing 7 million patent abstracts, can be found on the WebSOM

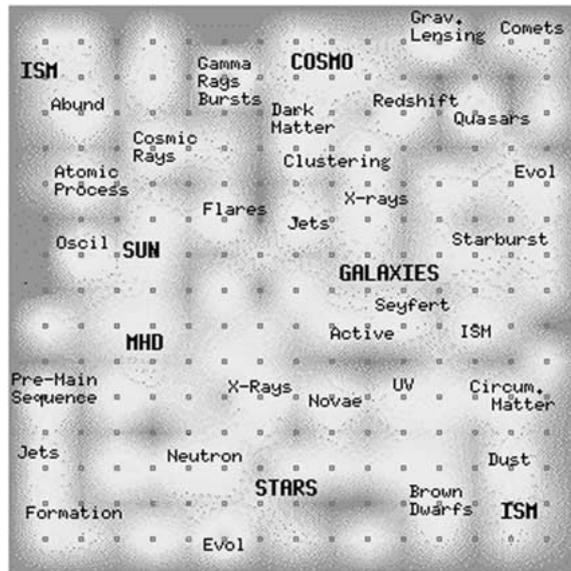


Figure 1. Visual interactive user interface to the journal *Astronomy and Astrophysics* based on 8000 published articles. Relative intensity represents relative document density.

server, <http://websom.hut.fi>. Further results are available in Oja and Kaski (1999) and Kohonen (2001).

1.3. Kohonen-type maps with coordinate input and with graph input

In this section, we will review two visual user interface frameworks, based on two different types of input data. Firstly, we are dealing with document/index term dependencies. From such data, a cartographic representation is periodically (i.e., in batch mode) updated, and made available as a clickable imagemap. The image file itself is created and links made active in it using a batch script on the server side, and provided to the user as an inlined image. The second input data framework is based on linkage data. A multigraph of links connecting three or more types of object is first created. Such objects include author name, publication title, and content-related keyword (e.g., astronomical object name). In this case, a Java application is used to construct the interactive cartographic representation in real time.

Figure 1 shows a visual and interactive user interface map, using a Kohonen self-organizing feature map. The original of this map is in color, and it is shown here in monochrome. Relative color intensity – brightness – 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 deployed as a clickable imagemap, allowing activation on user command of

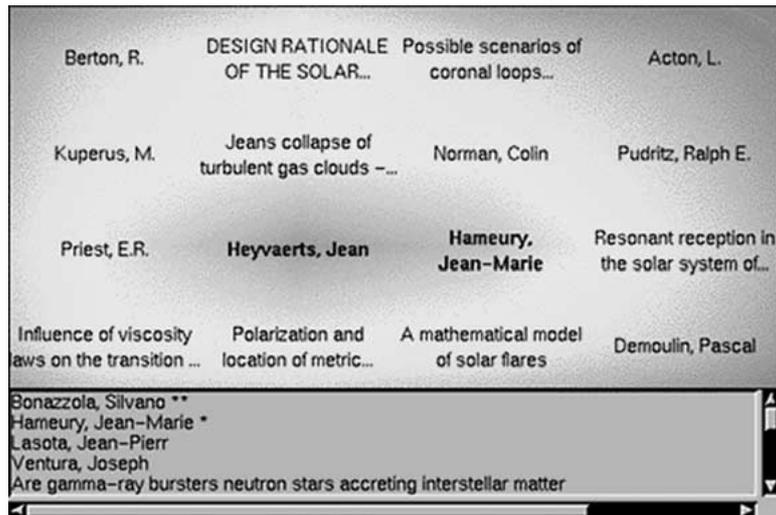


Figure 2. Visual interactive user interface, based on graph edges. Vertices are author names, article titles and (not shown here) astronomical object names.

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 (viz. relational tables). 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:

- The journal *Astronomy and Astrophysics*: <http://simbad.u-strasbg.fr/A+A/map.pl>
- *The Astrophysical Journal*: <http://simbad.u-strasbg.fr/ApJ/map.pl>
- The VizieR catalog server: <http://vizier.u-strasbg.fr/viz-bin/VizieR>

In Figure 1, strongly represented concepts are shown in bold: ISM = interstellar matter, COSMO = cosmology, MHD = magneto-hydrodynamics, SUN, STARS, GALAXIES. MHD processes exist in the Sun, which in turn is a star. Cosmology research is usually galaxy-oriented. This divide between solar system research, and cosmology research outside the solar system, is fairly well represented in this map. Near cosmology, we see terms such as gravitational lensing, redshift, dark matter, gamma ray bursts, and so on. This annotation of the map was informed by the keywords used to construct the map, and was carried out manually. Consideration was given to automated annotation, but left for future work.

In Guillaume (2000) and Guillaume and Murtagh (2000) a Java-based visualization tool was developed for hyperlink-rich data in XML, consisting of astronomers, astronomical object names and article titles. It was open to

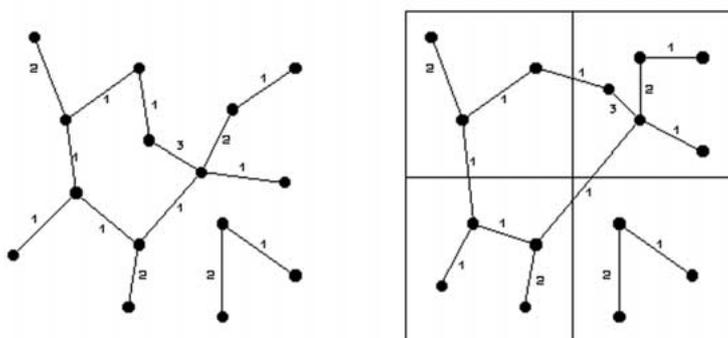
PARTITIONING THE GRAPH AND CREATING A DENSITY MAP

Figure 3. Graph clustering: illustration of the principle used to produce the map shown in Figure 2. Left: graph of objects (author names, articles titles, astronomical objects) with association weights. Right: mapping into a grid of clusters using a homogeneity criterion.

the possibility of handling other objects (images, summary tabulations, etc.). Through weighting, the various types of links could be given priorities. An iterative refinement algorithm was developed to map the nodes (objects) to a regular grid of cells, which as for the Kohonen self-organizing map, are clickable and provide access to the data represented by the cluster. Figure 2 shows an example for an astronomer (Jean Heyvaerts). Figure 3 illustrates the graph algorithm used as a basis for the type of map shown in Figure 2. Given the increasingly central role of XML in access to web information and data, the importance of such clustering for data organization and as a basis for knowledge discovery can be underscored.

In Figure 2, a set of 16 documents was used. The units clustered were authors, titles and astronomical objects. The map was arranged to give a central position to a selected unit (in this case, author Prof. Heyvaerts). The annotations shown in the regular grid are representative ones. Clicking on a location (in this case, Prof. Hameury) provides a list of titles in the lower panel.

We can therefore 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 is the case for the example discussed in Figure 2. This is likely to be the case whenever XML XLink functionality is used as the basis for associations between our objects.



Figure 4. Interactive visual user interface on a Palm V.

If a dense interdependency graph case is available, a convenient way to process the data is to project the objects, using these interdependencies, into a geometric space. This is done using principal coordinates analysis, which is also referred to as classical multidimensional scaling and metric scaling (Torgerson 1958; see also the short description in Murtagh and Heck 1987).

1.4. *Kohonen maps on PDA platforms*

Portable and mobile computing platforms allow for a new area of application of visual user interfaces, given the normal mode of interaction using a stylus. Figure 4 shows the graphical user interface to the catalogs database. Not all web browsers on PDA platforms support imagemaps. In this case the Fire-Viewer application (www.firepad.com) running on a Palm OS platform was used.

1.5. *Kohonen and related maps as visualization-based user interfaces: summary*

Let us call the first system described in Section 1.3 and exemplified in Figure 1, which uses coordinate input, and is in operational use at Strasbourg Observatory, the “Strasbourg” system.

Table 1. Characterization of systems discussed in Section 1 (Strasbourg, Graph and Palm), and those to be discussed in Section 2 (Thesaurus, Hierarchies) (Note: in column 4, O = Concept Hierarchy; in columns 5 and 7, Y = yes, N = no)

	Static/ dynamic	Client/ server	Coord. space/ sparse graph/ concept hier.	XML input	SOM/ other optimization	Web browser for client
Strasbourg	S	S	C	N	S	Y
Graph	D	C	S	Y	O	N
Palm	S	(C)	C	N	S	Y
Thesaurus	S	(S)	O	N	S	N
Hierarchies	D	S	O	N	S	Y

The second system, also described in section 1.3 and exemplified in Figure 2, is based on client-side processing of a sparse multigraph, and will be termed “Graph”.

The third system, described in section 1.4 and illustrated in Figure 4, is based on a Palm OS platform: “Palm”.

In the next section, section 2.1, we will first review older work of ours based on the IAU “Thesaurus”.

This will be followed in subsequent sections with an in-depth look at the IRAIA information visualization system based on concept “Hierarchies” (IRAI, “Getting orientation in complex information spaces as an emergent behavior of autonomous information agents”, was a European Fifth Framework project from 1999 to 2002).

The following table summarizes the functionality of these different systems. Note that “Palm” and “Thesaurus” were not full client-server systems.

2. Visual User Interfaces from Ontologies

2.1. *Defining an ontology*

In Murtagh (1996) we constructed and evaluated cartographic documentation database representations based on a thesaurus.

We used approximately 1500 astronomy observing proposal abstracts. There were short texts comprising type of observation (e.g., “general” or “guaranteed time”, for example, the latter being part of a larger sequence of observations), theme (from a set list, e.g., “galaxies and clusters – gas and dust”), title, proposal authors (from a few names to a few dozen), and proposal

abstract in plain language (from a few lines to a few dozen). These proposals had been accepted in one of the first four Hubble Space Telescope observing “Call for Proposal” cycles, from the period 1989 to 1993.

We took the IAU (International Astronomical Union) Thesaurus (Shobbrook and Shobbrook 1993) to provide an astronomically-aware domain of discourse. A list of astronomical phrases was derived from this, to provide index terms. Examples of index terms used include: “IRAS galax*”, “spectral energy distribution”, “elliptical galax*”, “spectral type”, “absorption line”. Here, “*” is simply notation for a wild-card character, and represents both “galaxy” and “galaxies”. Problems remained however. For example, an “A star” is an A-type star, in a categorization scheme related to physical characteristics like temperature and mass, which includes types labeled O, B, A, F, G, and so on. But clearly the term “A star” could also quite legitimately be the indefinite article followed by a noun. So while “A star” is quite important from the point of view of astronomy semantics, we ignored such syntactically troublesome terms. The consequences of ignoring a number of such cases, as verified by the final results obtained, was not large.

A set of 286 phrases – more accurately characterized as combined terms – was retained, each of which was present in one or more proposal abstracts. To evaluate the potential for such terms to describe the interrelationships between the proposal abstracts, a clustering was carried out on 1434 proposals, crossed (presence/absence) with the 286 terms. A Kohonen SOM 8×8 output square grid was used, and successfully characterized in terms of major categories of observation, including “stars”, “planetary nebulae”, “solar system”, and so on.

We will now describe a recently developed system using a number of concept hierarchies as input (see Augé et al. 2001; Taskaya et al. 2001; Mothe et al. 2002). These concept hierarchies describe the domain: macro-economics and production statistics.

A concept hierarchy defines a sequence of mappings from a set of low-level concepts to higher-level, more general, concepts. These concepts may be defined within two structures. A hierarchical structure is such that a child node cannot have two parents, but a parent can have more than one child. A lattice structure is such that a child node can have two parents. A concept hierarchy can be explicitly generated by expert users before the data is queried, and will be static, or it can be generated automatically, and the user may reform the hierarchy when needed.

The concept hierarchy used in the context of our work was based on hierarchical structure and generated by economists. We will discuss as a case-study the use of the Eurostat database (Eurostat, Statistical Office of the European Union, Luxembourg), and in particular documents from the

- Total Industry
 - Total industry (excluding construction)
 - Mining, quarrying and manufacturing
 - Intermediate goods industry
 - Energy
 - Intermediate goods industry, excluding industry
 - Capital goods industry
 - Consumer goods industry
 - Durable consumer goods industry
 - Non-durable consumer goods industry
 - Mining and quarrying
 - Mining and quarrying of energy producing materials
 - Mining and quarrying except energy producing materials
 - Manufacturing
 - ...
 - Electricity, gas and water supply
 - Construction

Figure 5. A small part taken from Branch concept hierarchy.

Eurostat Economic Bulletins. The Eurostat database supports three concept hierarchies: Branches, Themes, Countries. The total number of Branch concepts is 423, the number of Theme concepts is 30, and the number of Country concepts is 23. An example of the Branch concept hierarchy is shown in Figure 5.

2.2. Document annotation and derivation of document-term weights

The mapping of concept hierarchies onto documents is achieved by a vector voting method. We term this document annotation. We refer to a *term* in a document and an *entry* from a concept hierarchy. The weight of a term derived from a concept hierarchy is proportional to the number of matching occurrences in the documents. Two additional factors are also taken into account in this document-term weight.

- Document term frequency: the weight is inversely proportional to the number of terms associated with a document.
- Entry frequency in the concept hierarchy: the greater the frequency of an entry, the less the importance of the term, and the lower the weight.

Following selection of terms from the concept hierarchies (through removal of stop words and the application of stemming algorithms), the following formula is applied to define a document-term weight:

$E(H)$	an entry E from the concept hierarchy H
D	a document
$F(T,D)$	number of occurrences of term T in document D
$S(D)$	total number of terms in the document D
$S(H)$	total number of terms in the concept hierarchy H
$F(T,H)$	number of occurrences of term T in hierarchy H
$NT(E,D)$	number of terms from E that occur in D
$NT(E)$	number of terms that compose the entry E

$$\text{Weight}(E(H), D) = \sum_T (F(T,D)/S(D) \cdot S(H)/F(T,H)) \cdot \exp(NT(E,D)/NT(E))$$

Before clustering the document-entry dependencies with the Kohonen self-organizing map, the annotation values are taken according to this rule:

*If total number of occurrences of a concept (i.e. entry) within the returned documents is greater than the number of returned documents * α then retain that concept.
Else reject.*

α is a small constant which is taken as 0.5.

2.3. Kohonen map: Implementation

The Kohonen self-organizing map is an unsupervised neural network that provides cartographic visualisation of the results. The algorithm is as follows:

- Assignment step: Let x be the input vector and m the model vector. Find the Euclidean distance between them and determine the minimum model vector.
- Iterative update step: $m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)]$

t is time (iteration count) and $h_{ci}(t)$ is the neighborhood function. This means that the best matching model vectors tend to fire when the winner model vector fires.

3. Implementation

HTML pages are inherently static and do not allow interaction between the programs and the user.

Javascript and CGI (Common Gateway Interface) provide for such interaction. Both CGI and Javascript can be used to create animations, getting input from the user, error checking from input, carrying out some processing of information, providing a query interface to databases, and so on. An important difference is that CGI is a compiled program on the server side and Javascript is a client side script language.

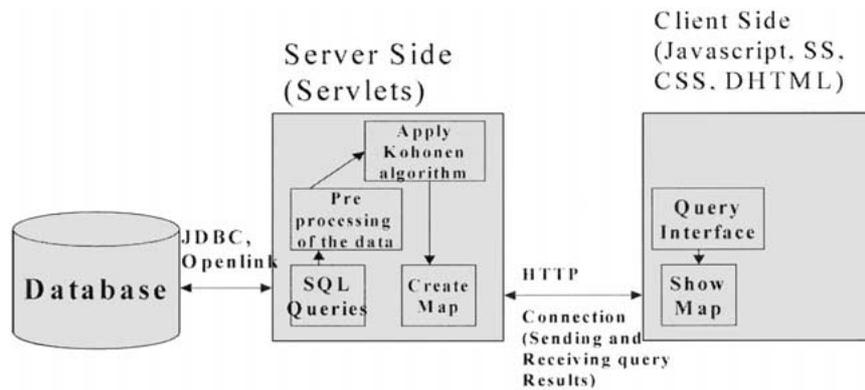


Figure 6. Client-server system design.

Style sheets are for displaying text on web pages. They allow for varied presentation of text. CSS (Cascade Style Sheets) are a widely implemented form of style sheets, especially used in DHTML. DHTML (Dynamic HTML) is a combination of Javascript, CSS and DOM (Document Object Model) which provides a comprehensive support framework for creating dynamic, flexible web pages for the user.

The combination of Javascript, CSS and DOM is supported in Internet Explorer 4.0, and Netscape 4.0, and later versions of both browsers. However DHTML implementation is rather different between Internet Explorer and Netscape. As a result, full cross-browser DHTML implementation is difficult.

These technologies were used in our visualization implementation because they are very appropriate for creating an effective, interactive, dynamic environment with Kohonen maps.

Figure 6 shows the architecture implemented. Objects used were as follows:

- Databases available, user-selected databases, associated concept hierarchy categories.
- Concept hierarchies, user-selected categories.
- Input data for Kohonen map: document-term sets.
- Map images resulting from the Kohonen algorithm.
- Display objects: web page background image, dynamic window information.

Server-side classes used are:

- Communication with the database.
- Organization of selected database information including concept classes.
- Organization of concepts from concept hierarchies.

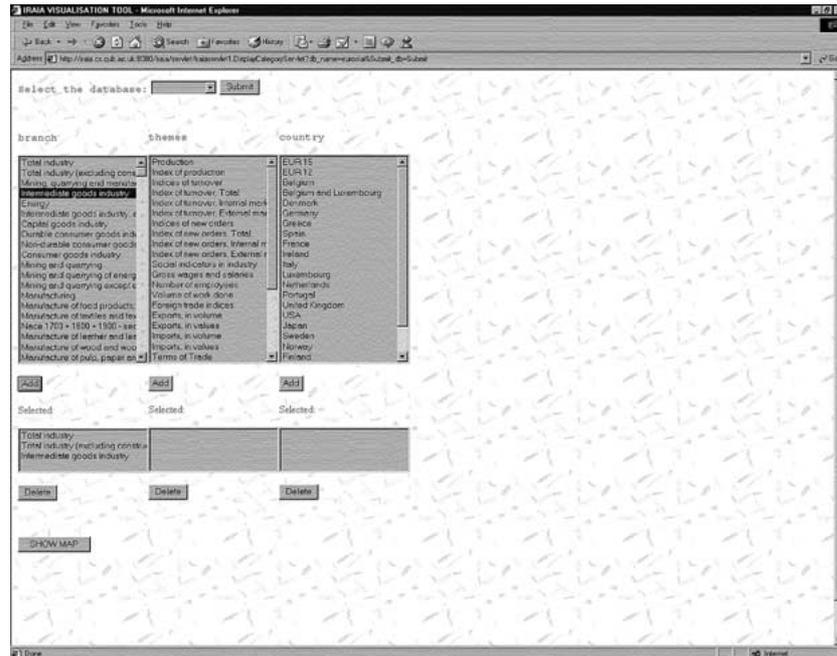


Figure 7. Visualization query interface, using 3 concept hierarchies.

- Display of the concept hierarchies, allowing selection of concepts (entries) to comprise the user query.
- Organization of inputs to the Kohonen algorithm.
- Organization of weights and parameter information for the Kohonen algorithm.
- Building of the Kohonen map.

On the server side, requirements were: JDK, Apache web server, and Jakarta Tomcat server for servlets.

4. Example of Usage

A query is formed from the cross product of the entries of the concept hierarchy categories: see Figure 7.

Example:

Branch: Total Industry, Energy

Theme: Production

Country: Germany, UK

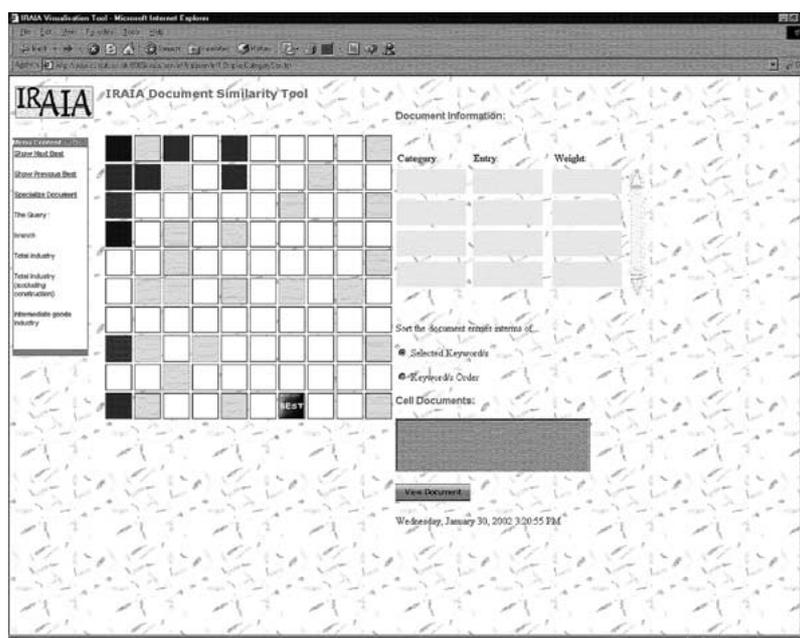


Figure 8. The first display after the query was submitted. The 10×10 grid is the Kohonen map output.

In this case the result query will be:

Total Industry & Production & Germany
 Total Industry & Production & UK
 Energy & Production & Germany
 Energy & Production & UK

Documents will be sought having at least one of the above combinations.

Following query submission, the user will have a display as shown in Figure 8. The explanation of each component on the interface is as follows:

“Best” indicates the cell containing document(s) having the highest selected entry values.

“Show Next Best” and “Show Previous Best” selection options are supported.

Relevance feedback is provided by “Specialize Document”. If there are entries belonging to unselected categories, the user may select such entries and add them to the query. In Figure 9, the user had first selected entries from the Branch category and now in this map has a query box containing Theme and Country categories. Figure 10 results from this.

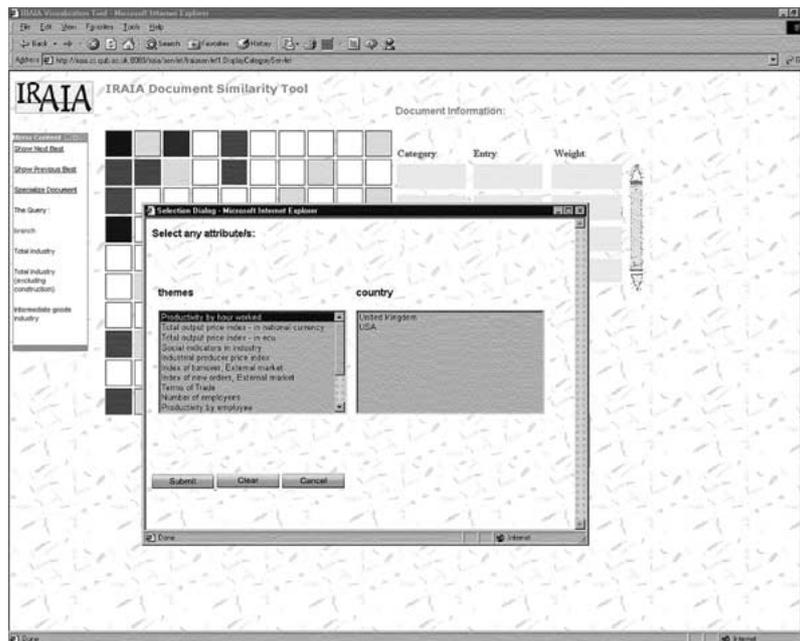


Figure 9. When the user presses “Specialize Document”, the dialog box above appears.

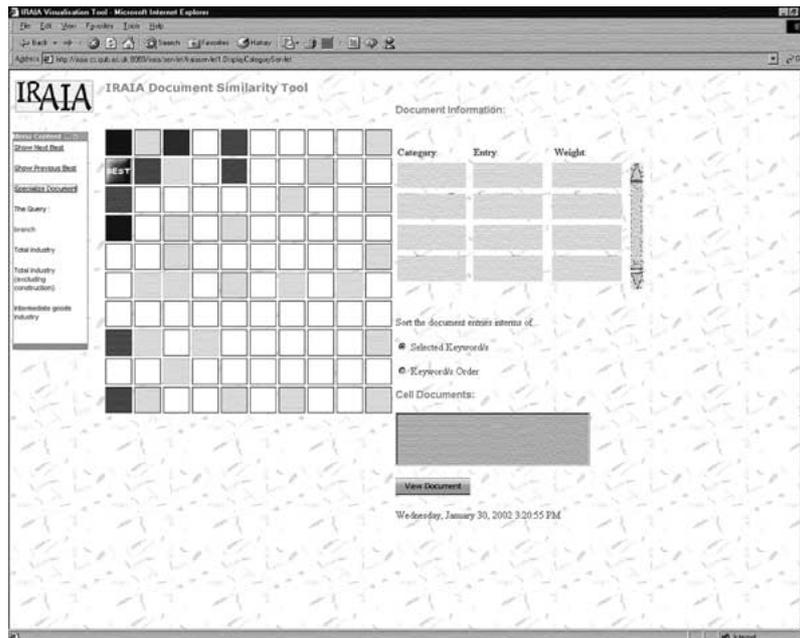


Figure 10. After the user selected the “UK” entry from the “Specialize Document” window, the “best” cell was changed.

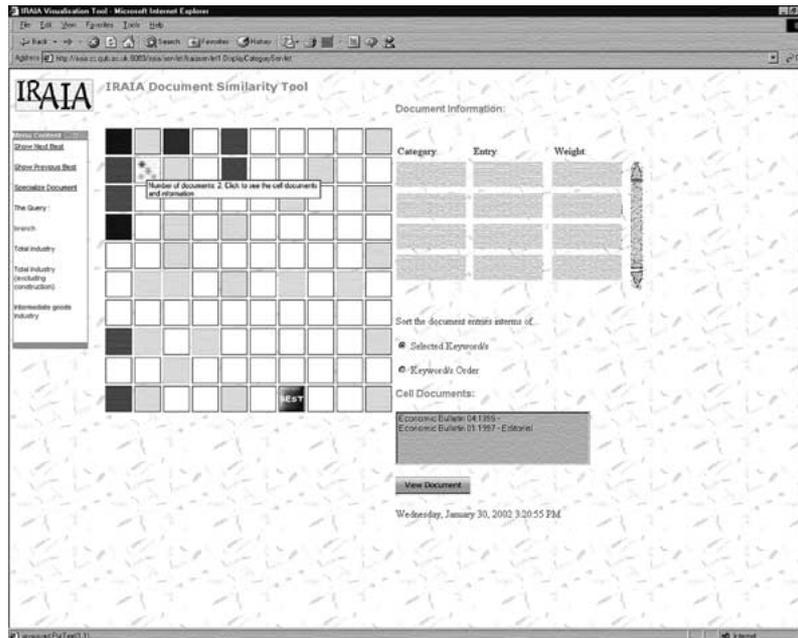


Figure 11. Clicking a cell lists the document names.

When the user clicks a grid cell in the Kohonen map, the document names existing in that cell are displayed in a “Cell Documents” list box as shown in Figure 11. The user can display the document contents by selecting a document and using the “View” button. Tool tip texts are used on the map to show the number of documents associated with a given cell.

Selecting a document on the “Cell Documents” list box causes the display of that selected document’s entries with the weight information to the right side of the display area (see Figure 12).

5. Short Evaluation

In Figure 13, a “Total Industry” and “United Kingdom” query is submitted and displayed on the map. In Figure 14, the document titles belonging to the query used in Figure 13 are displayed. To quantitatively assess this distribution of titles, we have to consider all the entry values assigned to each document, together with their weights. From the titles, we see that Euclidean distances between the documents are well preserved in the Kohonen map.

Figure 15 shows relative locations of topic areas, and Figure 16 shows gradation of document publication dates.

Some general limitations of this approach are:

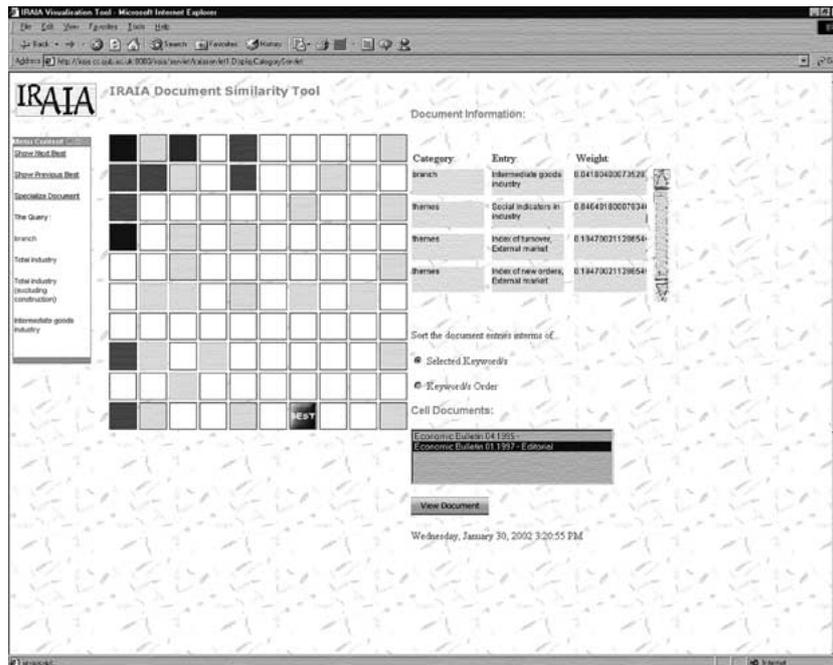


Figure 12. Clicking a document name causes that document’s information to be displayed.

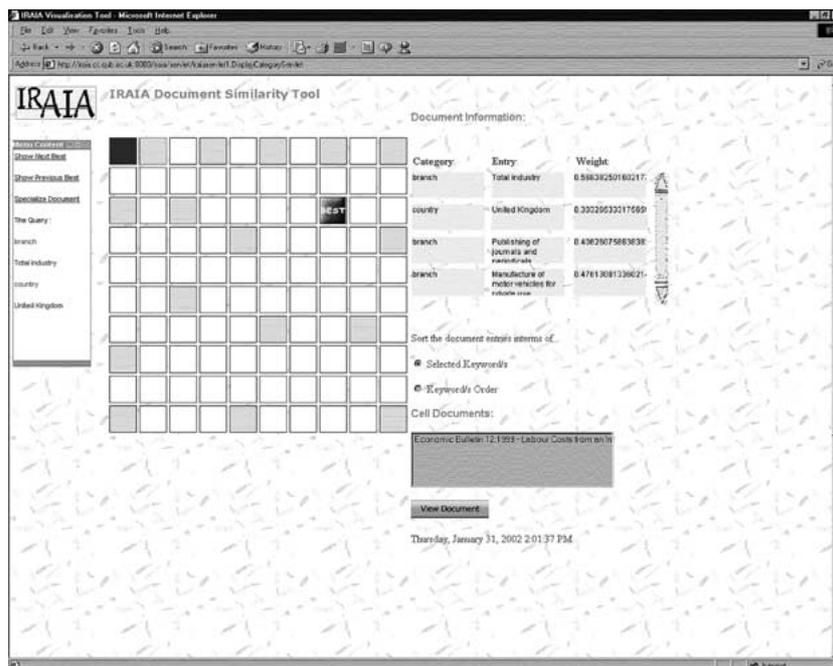


Figure 13. “Total Industry” and “United Kingdom” query result.

08.98. Key National Accounts Data Forecast for 1998 and 1999 Federal Republic of Germany 09.98 Editorial 01.99 Key National Accounts Data Forecast for 1998 and 1999 Federal Republic of Germany	12.99 Labour Costs from an International Comparative Perspective: A Critical Look at Existing Concepts		03.00 Germany Must Intensify its R&D Efforts		04.95 Global Rise in Fuel Consumption by the Transport Sector Undermines Aim of Reducing CO2 Emissions	07.95 Climate Change Policies in the Wake of the First UN Conference of the Parties in Berlin		05.95 German Industry's Voluntary Commitment to Reduce CO2 Emissions - No Substitute for an Active Policy Against Climate Change
06.95 The Investment Requirements of CO		11.97 The German Economy in the Autumn of 1997				11.97 Editorial *		
				07.98 Financial Crisis in Asia:				03.97 Unit Labour Costs Still Higher in East than in West Germany
		05.99 The German Economy in the Spring of 1999						
					02.00 Employment Policy Strategies and Successes - Great Britain and France Compared		09.97 Micro- and Macroeconomic Adjustment Processes in East Germany - Sixteenth Report	
05.00 The German Economy in the Spring of 2000								
11.99 The German Economy in the Autumn of 1999				06.99 The Economic Situation in Russia				02.98 Micro- and Macroeconomic Adjustment Processes in East Germany - Seventeenth Report

Figure 14. Document titles resulting from the query used in Figure 13. The marked cell shows the “best” match. Note that the narrow rows and columns are used simply for display convenience.

- The documents which are available, and processed, in the database will of necessity determine the results obtained.
- There are qualitative differences between documents like “Editorial” versus longer technical reports which are not taken into account directly.
- Everything relies on indexing, which in turn depends on the concept hierarchy. The documents are described within the existing entries. E.g.,

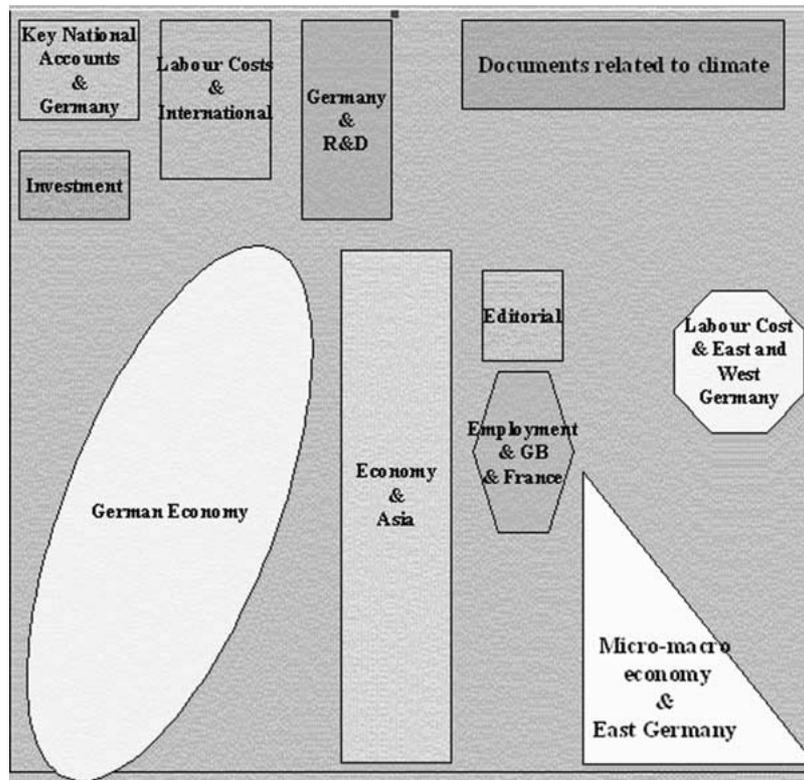


Figure 15. The distribution of topics associated with the query of Figures 13 and 14.

98	99		00		95		95		95
95		97					97		
				98					97
		99							
					00			97	
00									
99				99					98

Figure 16. Document distribution according to publication year.

“United Kingdom” is a valid entry in the concept hierarchy but we do not have any “CO₂” entry (carbon dioxide, assuming we wish to look at environmental issues) in the concept hierarchy.

- We do not consider all the entry weights because it is computationally very demanding to use very large sparse weight vectors in clustering. We consider the frequent items occurring in the documents.

6. Conclusions

Summarizing information is necessary. A visual summary is often a natural way to do this. When we add the possibilities for human-machine interactivity, then visual user interfaces become a toolset of significance. In our survey, we have described the evolution of web technologies that provides further support for this aspect of human-machine interaction. In our operational system, the information resulting from a user query is built on the fly into a graphical interactive user interface.

The ergonomics of interaction based on visual user interfaces is still under investigation. Although human understanding is greatly aided by maps and drawings of all sorts, we have yet to find the most appropriate visual displays for use in visual user interfaces.

The present work is a contribution towards increased use and greater reliance on visual user interfaces in the future.

Acknowledgements

This work was carried out in the context of the IRAIA project, “Getting orientation in complex information spaces as an emergent behavior of autonomous information agents”, Fifth Framework project IST-1999-10602, 1999–2002. We are grateful for extensive referee comments on an earlier version of this paper.

Appendix: Interactive Visualization Systems

Systems Related to the Kohonen SOM

The self-organizing feature map was developed by Kohonen in the early 1980s and is one important paradigm in the cartographic representation of information. As discussed in Murtagh and Hernández-Pajares (1995), it can be considered as a contiguity-constrained two-dimensional planar projection of a multidimensional feature space, and in addition lends itself very well to pixelated display. A range of directly and indirectly related visual

display tools will be reviewed in this Appendix. Most are commercial, and most consist of a platform-specific – most often Windows – client application.

1. Viscosity SOMine, from Eudaptics Software, www.eudaptics.com, supports Kohonen SOM and Ward's hierarchical clustering. Supported are interactive display and file input/export.
2. Kensington Discovery Edition, from Inforsense Software, www.inforsense.com. It is Java client based. Supported are: decision-tree based classification, Bayesian classification – AutoClass, SOM, backpropagation neural network, and association rule discovery. 2D and 3D interactive visualization are supported, running on different databases, spreadsheets, free text documents, web documents, XML documents, and others. It has information retrieval functionality, context knowledge management and dynamic data warehousing. Reporting capability is available.
3. Clementine, from SPSS, www.spss.com/clementine. Supported are Kohonen maps, regression, backpropagation neural networks, and ID3 and C5.0 machine learning. Import/export data capability is available, together with reporting, data preprocessing, modeling, and visual programming.
4. IBM Intelligent Miner, from IBM, www-4.ibm.com/software/data/iminer. This works with the DB2 database management system, and also supports ascii input. Kohonen maps are supported, and also decision trees, classification and regression methods. Preprocessing and data warehouse tools are available separately.
5. DataEngine, from MIT (Management Intelligenter Technologien) GmbH, www.dataengine.de. Supports Kohonen maps, statistical methods, and fuzzy methods. It provides for data acquisition, preprocessing and visualization.
6. NGO NeuroGenetic Optimizer, from BioComp Systems, www.bio-comp.com. Supports Kohonen maps and temporal maps, backpropagation, other neural network algorithms, and genetic algorithms. Multiple models can be fit simultaneously.
7. Visipoint, from Visipoint, www.visipoint.fi. Kohonen map supported, with input from any text or spreadsheet format. Various visualizations are available.
8. GeneSight, from BioDiscovery, www.biodiscovery.com. Supported are Kohonen maps, hierarchical clustering, time series analysis, and principal component analysis. Varied visualizations are possible, and also file import/export, and data preprocessing.

Commercial Tools Using Other Visual Methods in Information Retrieval

Because so many information visualization tools are available (Hofman and Grinstein 2002), we will consider a number of web-based document visualization tools here.

1. WebMap, from WebMap Technologies, www.webmap.com. This is a built-in component for browsers. It displays fixed-sized domains, which can be browsed as an interactive map. A search mechanism is also provided.
2. WebTheme, from Pacific Northwest National Laboratory, www.pnl.gov/infoviz/spire/spire.html. Clustering and information retrieval methods are used.
3. Zoef, from MediaLab and Headline Communication, www.zoef.nl. An interactive map can be focused, relative to a specific concept.
4. Umap Web, from Trivium, www.umap.com/highuk/index.htm. Keywords are extracted from web pages retrieved, and an interactive map is created.
5. Thinkmap, from Thinkmap, www.thinkmap.com. A representation of documents is provided, to support searching and browsing.

6. map.net, from Antartica Systems, <http://map.net>. A free service for searching and visualizing web documents, based on a commercially available tool, VisualNet. It is built around open interfaces including SQL, HTTP and XML.
7. Semio Map, from Semio, www.semio.com. Used for semantic indexing, it displays the information as an interactive concept map in 3D. Relational databases and PDF files are supported as input.

A sample of iconized displays from all of these systems, and further discussion, is available at the following address: strule.cs.qub.ac.uk/~fmurtagh/iraia/infovis-market.pdf

References

- Augé, J., Englmeier, K., Hubert, K. & Mothe, J. (2001). Classification automatique de textes basée sur des hiérarchies de concepts. *Proceedings of VSST – Veille Stratégique, Scientifique et Technologique*, 291–300. Barcelona, October 2001.
- Doyle L. B. (1961). Semantic Road Maps for Literature Searchers. *Journal of the ACM* **8**: 553–578.
- Guillaume, D. & Murtagh, F. (2000). Clustering of XML Documents. *Computer Physics Communications* **127**: 215–227.
- Guillaume, D. (2000). Distributed Information Retrieval, Search and Processing in Astronomy. PhD thesis, Université Louis Pasteur – Strasbourg I.
- Hoffman, P. E. & Grinstein, G. G. (2002). A Survey of Visualizations for High-Dimensional Data Mining. In Fayyad, U., Grinstein, G. G. and Wierse, A. (eds.) *Information Visualization in Data Mining and Knowledge Discovery*, 47–82. Morgan Kaufmann.
- Kohonen, T. (2001). *Self Organizing Maps*, 3rd edition. Springer-Verlag.
- Mothe, J., Chrismont, C. & Alaux, J. (2002). Visualisation globale de collections de documents sous forme d'hypercube – Le système DocCube. *Journées francophones d'Extraction et de Gestion des Connaissances*, EGC.
- Murtagh, F. & Heck, A. (1987). *Multivariate Data Analysis*. Kluwer.
- Murtagh, F. & Hernández-Pajares, M. (1995). The Kohonen Self-Organizing Map Method: An Assesment. *Journal of Classification* **12**: 165–190.
- Murtagh, F. (1996). Content-Based Information Retrieval: New Tools for Textual Data, New Problems for Image Data. In Dubois, J.-E. & Gershon, N. (eds.) *The Information Revolution: Impact on Science and Technology*, 13–20. Springer-Verlag.
- Oja, E. & Kaski, S. (1999). *Kohonen Maps*. Elsevier.
- Poinçot, P., Murtagh, F. & Lesteven, S. (1998). A Spatial User Interface to the Astronomical Literature. *Astronomy and Astrophysics Supplement Series* **130**: 183–191.
- Poinçot, P. (1999). Classification et recherche d'information bibliographique par l'utilisation des cartes auto-organisatrices, applications en astronomie. PhD thesis, Université Louis Pasteur – Strasbourg I.
- Poinçot, P., Murtagh, F. & Lesteven, S. (2000). Maps of Information Spaces: Assessments from Astronomy. *Journal of the American Society for Information Science* **51**: 1081–1089.
- Shobbrook, R. M. & Shobbrook, R.R. (1993). *The Astronomy Thesaurus*, Version 1.1. International Astronomical Union.
- Taskaya, T., Contreras, P., Tao, F. & Murtagh, F. (2001). Interactive Visual User Interfaces to Databases. In Smith, M. J., Salvendy, G., Harris, D. & Koybek, R. J. (eds) *Usability Evaluation and Interface Design*, Vol. 1, 913–917. Lawrence Erlbaum.
- Torgerson, W. S. (1958). *Theory and Methods of Scaling*. Wiley.

