Map Displays for Information Retrieval

Xia Lin
School of Library and Information Science, University of Kentucky, Lexington, KY 40506-0039.
E-mail: xlin@ukcc.uky.edu

The focus of this article is to develop a map display for information retrieval. Through an examination of relationships among visual displays, information retrieval, and browsing, advantages of visual displays for information retrieval are characterized as (1) the ability to convey a large amount of information in a limited space, (2) the potential to reveal semantic relationships of terms and documents, and (3) the facilitation of browsing and perceptual inferences on retrieval interfaces. These advantages are further demonstrated through a map display generated by a neural network's self-organizing algorithm. The map display detects complex relationships among given documents, and reveals the relationships through a spatial arrangement of terms abstracted from the documents. The map display also provides interactive tools to allow the user to interact with the underlying information. Examples of the map displays show that such map displays can be used both as an overview tool and an access or exploration tool, and the map displays will likely increase the amount of information that the user is willing to browse.

Introduction

The world in which we live is a visual world. We look, we see, and we find things. Even for written information, we expect to see it on tangible media stored in visible locations. We have developed a whole set of perception and association techniques to get what we need from the environment. However, when we search information by computer, all the things, information, media, locations, as well as their relationships, become invisible; all the clues for visual perceptions and associations disappear. We heavily rely on our competence to generate queries and the computer's ability to match the queries with information stored in the computer. At best, when interaction is involved, we can modify queries based on the retrieved results of a previous query. This is like searching books in a library without light. We can walk around from stack to stack in the dark, without knowing what stacks that we have walked through. We can get a few books each time and walk out of the library to see if the book we are searching for is among them. If not, we have to walk in again, based on our knowledge and our previous experience, to a location where we hope the book would be. In this situation, success in finding a book greatly depends on whether we can walk to the right place in the dark (to generate a good query), and whether we know how to adjust our locations until we get to the right place (to modify queries interactively).

Can we turn on the light for such a library? Can we develop some visible cues in our retrieval systems so that we can use our perception for information seeking in the digital environment? Answers to these questions may reside in the recent advance of information technology. ''Information processing has been evolving from numerical computation to character handling, and now to visual information processing'' (Kunii, 1989). As a result, visualization will become ''the second computer revolution'' (Friedhoff & Benzon, 1989). Computers are expected to be used to reveal associations and properties of electronic information to allow people to use their visual capabilities for information seeking (Veith, 1988).

This article discusses using some mapping techniques to create visual displays for textual information stored in the computer. The focus of the article is on the development of a map display for information retrieval. The map display attempts to show both contents and semantic structures of a document space by mapping major concepts and documents of a document space to a two-dimensional map. It preserves, as faithfully as possible, document semantic relationships and reveals these relationships through various visual components of the display. By communicating a spatial metaphor to the user and making information in a document space graphically visible, the map display provides a visual guide to help the user perceive contents and structures of the document space.

In the following, relationships among information retrieval, browsing, and visual displays are discussed first,

Received May 26, 1995; revised December 18, 1995; accepted December 18, 1995.

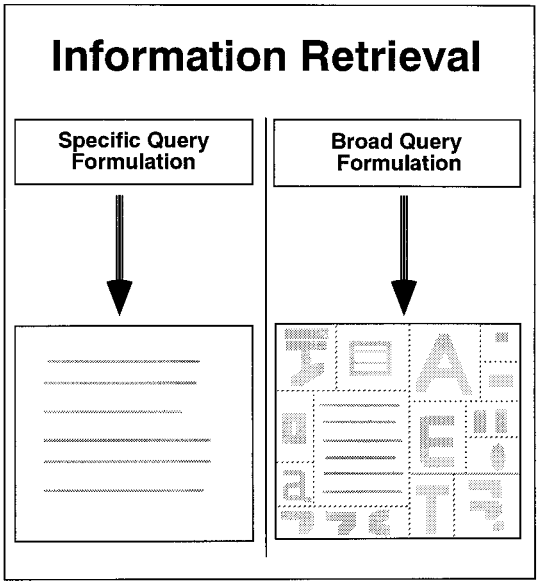


FIG. 1. Frameworks for information retrieval. Two frameworks for information retrieval are presented. The **left** hand side is the traditional framework characterized by retrieval systems that require a specific query to generate a set of documents shown as a list. The **right** hand side is a new framework to characterize systems that will take a broad query, conduct a broad search, and show a large set of documents in a graphical form to support scanning and browsing.

followed by a review of different types of visual displays. Construction of the map displays will then be demonstrated through the application of a neural networks' self-organizing algorithm, Kohonen's feature map, to several document collections. Properties of the map displays, as well as the interaction with the map displays, is explored last.

A Framework of Visual Displays for Information Retrieval

Information retrieval is to search and identify relevant documents that will satisfy the user's information needs. After more than 40 years of research, it is quite clear that traditional query-based retrieval techniques do not always satisfy the needs of prospective users (Salton, 1989). Aids must be provided to these users when they encounter a set of documents retrieved by the system (Korfhage, 1991). Figure 1 illustrates the process of information retrieval, with and without visual displays. In this figure, two frameworks for information retrieval are shown, both represent the information retrieval process in two steps: querying and displaying. The framework on the left characterizes traditional retrieval systems: Given a specific query, usually a Boolean logic expression and/or a proximity matching expression, a retrieval system searches through the whole document space and returns a list of documents that match the query.

While most commercial systems still function in accord with general principles of this framework, problems associated with such systems or the framework are well known and well documented (Belkin & Croft, 1987). The first problem is the ''query formulation''—most users have difficulty specifying their needs by a specific query formulation; even if users are successful in doing so, systems have difficulty retrieving all relevant documents without overwhelming the users with irrelevant documents. The issue of precision/recall has been a bottleneck for retrieval systems: Retrieving more relevant documents (high recall) is often at the price of getting more irrelevant documents (low precision).

The second problem of the framework is the ''linear display.'' In most cases, documents matched to the query are displayed in a linear form, such as a list of authors or a list of document titles. Even when relationships of those matching documents are considered, the displays usually simplify all the relationships to a one-dimensional, linear order, either chronologically, alphabetically, or according to some kind of relevant (weighted) order between the matching documents and the query.

The second framework, represented on the right of the figure, addresses these two problems. In this framework, a broad query is used to find a preliminary large set of documents, and a structured visual display is applied to let the user browse and comprehend the documents and their relationships. This framework emphasizes that information exploration, or browsing, should be implemented not as a necessary but an integral part of the retrieval process. While it is generally desirable to have retrieval mechanisms that would deliver both high recall and high precision, it is likely more practical to provide browsing aids to assist the user in identifying relevant documents in a high recall/low precision situation. Visual displays that show terms and document relationships and reveal underlying structures of the document space will be such browsing aids that will relax demands on the performance of retrieval mechanisms and query generations. Such displays will allow the user to interact and browse a large quantity of search results in a limited display space. The difficult task of specific query formulation will be replaced by the much easier task of recognizing relevant documents in a structured display. The user will be able to apply not only logical and linguistic inferences, but also perceptual inferences, to discover information patterns or relationships and to recognize relevant documents.

Visual Displays and Browsing

Browsing is a direct application of human perception for information seeking, both in the electronic and non-electronic environment (Chang & Rice, 1993). Browsing is explorative; it is an interactive process in which one will scan large amounts of information, perceive or discover information structures or relationships, and select information items through focusing one's visual attention. In relation to information retrieval, browsing is particularly useful when:

- there is a good organizational structure and related information items are often located near each other (Thompson & Croft, 1989),
- users are not familiar with the content of the collection and they need to explore the collection (Motro, 1986),
- users have less understanding of how information is organized in the system and they prefer to take a low cognitive load approach to explore the system (Marchionini, 1987),
- users have difficulties in articulating their information needs (Belkin, Oddy, & Brooks, 1982), and
- users look for information that is easier to recognize than to describe (Bates, 1986).

Because of its interactivity and its dependence on human perception, browsing is much difficult to support by the computer. While it has become more and more clear to most system designers that information systems need to support browsing, it is less clear what functions of browsing may be supported and what techniques should be used to support these functions. Some techniques that researchers have explored to support browsing for information retrieval include:

- displaying a dynamic hierarchical information structure (Frei & Jauslin, 1983),
- providing an overview map of the information space (Halasz, Moran, & Trigg, 1987),
- providing a neighborhood map for each item (Thompson & Croft, 1989),
- showing both a miniature of the entire information space and a detailed local map (Beard and Walker, 1990),
- distorting the display so that the center of focus will be shown in more detail than other areas—the fish-eye views (Furnas, 1986), and
- supporting interactive functions such as zoom in, zoom out functions so that the user can select different level of details to display (Schatz & Caplinger, 1989).

Clearly, most of these techniques are associated with graphical displays of information. It is only natural to let users see ''information'' when they browse. One of the major issues that these techniques addressed is how information should be organized and fitted to the computer screen in a way that people can still browse and find the information they want. To do so, these techniques attempt to define a structure that will reduce complexity of information structures and ''fit'' a large amount of information to a limited display space. These techniques also create and add links to the information in order to facilitate association and navigation. Finally, these techniques also seek to present information only at the most appropriate level of details to the user in order to avoid disorientation and cognitive overload.

We may argue that these purposes may be better addressed in the study of how information should be visually displayed and how the visual displays should be organized and linked to the underlying information. The basic principle that needed to be studied is how information in the computer should be organized to ''reveal itself '' to users to encourage and enhance browsing, as the organizations in nature do (Marchionini, 1995).

Visual Displays for Information Retrieval

A central issue of organizing information for visualization is what formats and features of visual displays will help to organize large amounts of information to reveal information structures and to support effective use of human visual capabilities. In this section, research related to visual displays for information retrieval are reviewed. Four types of display formats are identified and their features are described and compared. These formats are hierarchical displays, network displays, scatter displays, and map displays.

Hierarchical Displays

A graphical display that shows data in a hierarchical form is a hierarchical display. Hierarchical displays simplify complex data structures and separate data into different levels, branches, or clusters. These functions help to represent both global and local views of data, to utilize the display screen effectively, and to direct the viewer's attention to the appropriate level of generality.

Hierarchical displays have been used in several prototype information systems. For example, in the Learning Support Environment (Hammond & Allinson, 1988), hierarchically stored data are collapsed into a global view showed on the screen to allow data comprehension and comparison. In the Electronic Document System (Feiner, 1988), ''pages'' of information are clustered into a hierarchy of ''chapters'' to provide a global view of the link network for easy browsing. In another system by Burgess & Swigger (1986), a hierarchical structure of the database is generated to allow the user to see local items within the context of a global structure. Other hierarchical displays include Treemaps (Turo & Johnson, 1992) that convey hierarchical properties of information spaces through sizes, colors, positions and patterns of individual nodes, and Cone trees (Robertson, Mackinlay, & Card, 1991) that add rotation as a third dimension to increase the hierarchical display space.

Hierarchical displays can be generated automatically for data that have hierarchical structures (such as files in a computer, Turo & Johnson, 1992), or for any other data with a defined measurement, using hierarchical cluster algorithms (Spath, 1980). Such algorithms begin with each entity as a cluster. At each successive step, pairwise distances between clusters are measured and the two closest clusters are merged as a new cluster. The process continues until all data form one cluster or until a small set of clusters is obtained. Willett (1988) provided a comprehensive review on using hierarchical clustering algorithms for document retrieval. Cutting, Karger, Pedersen, & Tukey, (1992) showed that hierarchical clustering could be an effective information access tool, particularly for browsing.

Despite many advantages of hierarchical displays, their disadvantages are also apparent. These disadvantages include (1) oversimplification of structures for certain data, particularly for those that are more appropriate to be represented by structures other than a hierarchy, (2) difficulty in generating and displaying hierarchical displays for large information spaces, and (3) increased cognitive load for users who are forced to make selections among the hierarchical branches, especially when the whole hierarchy is not displayed on the screen.

Network Displays

Network displays refer to the graphical display of links and nodes. Network displays show associative structures on the screen and let the viewer follow the links to browse items represented by the nodes. Network displays are often based on network representational models such as semantic nets (Quillian, 1968) and hypertext (Conklin, 1987; Foss, 1989). They can represent more general and complicated structures than hierarchical displays can. In this respect, network displays seem to be more suitable than hierarchical displays for revealing structures of a document space as the document space is often more complex than a hierarchy. However, if all the relationships in a complex document space are displayed in a network, the network display simply becomes a network maze. The network displays thus often present more information than the user can immediately comprehend (Beard & Walker, 1990).

A major issue of network displays is how to simplify the display without losing its useful structures. Common solutions to this problem include truncating and shrinking. Truncating refers to the display of a portion of the structure while providing tools to move to any other portions to be displayed. Shrinking provides the information at different levels of detail and suppresses any other detailed information from the display. Beard and Walker (1990) applied these methods in combination to allow easy access to various locations in a two-dimensional space.

Another solution to simplify network displays involves dimensional reduction. This can be done globally, as in the latent semantic indexing approach (Furnas et al., 1988), where a small set of ‘‘best’’ dimensions is selected. It can also be done selectively and let the user select the dimensions to be displayed. A problem of dimensional reduction, however, is how to minimize distortions due to the reduction. It has been noted that using latent semantic indexing to reduce the dimension of underlying data to a two-dimensional space would lose too much information (Furnas et al., 1988).

Network displays will be more effective if certain perceptual cues are incorporated, such as distances, sizes, and colors of nodes and links. For example, in the SCALIR system (Rose & Belew, 1991), different types of lines are used to distinguish different types of links (whether they are citation links, taxonomic links, or structural links). In the gIBIS system (Conklin & Begeman, 1989), colors are used to indicate both node types and link types so that type identification can become a rapid and unconscious activity. In the PFNETS display system (Fowler, Fowler, & Wilson, 1991), distances between nodes are used to reflect the degree of association of the nodes, while sizes of the nodes are used to indicate the connectivity of nodes (number of possible links associated with the nodes).

Finally, while most network displays are two-dimensional, there also exist three-dimensional network displays. SemNet (Fairchild, Poltrock, & Furnas, 1988) is a three-dimensional network display that represents data as labeled nodes connected by links. In this system, the links are color coded to represent types of links, the nodes are in positions such that no links intersect (this is only possible in a three-dimensional display). The unique feature of SemNet is its ability to let the user manipulate the view point: The user can select different views to observe different relationships of links and nodes. Another major project of three-dimensional display is the Information Visualizer developed at Xerox PARC (Robertson, Card, & Mackinlay, 1993).

In summary, network displays show both information contents and structures in the form of links and nodes. They can show complex data structures, and they facilitate and encourage visual inference through explicit links on the display. However, complex structures on network displays may confuse and distract the viewer, particularly when the organization of the display is not clear. Various techniques are needed to simplify the display and to create useful structures of the data. Because of the limited display space, network displays often show only a small portion of a network thus making it difficult to show the overall structure and to use spatial information on the displays. Like hierarchical displays, automatic construction of network displays remains a difficult problem. It is even more difficult to automatically identify different types of link and nodes in a network display.

Scatter Displays

Scatter displays refer to the graphical (dotted) image resulting from mapping high-dimensional data to a two-dimensional visual space. The basic elements of scatter displays are dots (or other small individual icons) that represent the mapping data. Most of the scatter displays are generated automatically by mapping algorithms. Because the mapping is usually driven by an error-minimum process or by the principle of finding a display configuration whose overall layout most closely matches the structure of the given data, the mapping creates a spatial orientation that reflects the overall layout of underlying data.

One of the earliest scatter displays for information retrieval was described by Sammon (1969). Sammon proposed a nonlinear mapping algorithm, which was essentially a variation of multi-dimensional scaling (Siedlecki, Siedlecki, & Sklansky, 1988), to map a high-dimensional space to a lower-dimensional space. He tested his algorithm by using it to evaluate the structure of a document space that contained 188 documents indexed by 17 indexing terms. These documents were mapped to a two-dimensional space by the algorithm. Five queries were collected and the relevant documents to these queries were then identified and mapped to a scatter display. The display clearly showed clusters of documents relevant to the five queries, and the inter-cluster relationships were consistent with their respective subject relationships.

Scatter displays are very useful in revealing underlying data structures of statistical data (Tufte, 1983). In particular, scatter displays can also be used to reveal semantic or intellectual structures embedded in statistical data. Using multi-dimensional scaling for document citation data analysis is such an example. White and Griffith (1981) produced an author map of information scientists from a document citation database. The map identified author groups and revealed relationships among the groups. It also visualized proximities of authors within the groups. The map was used as a tool to help the viewer understand the intellectual structure of information science.

Scatter displays often show meaningful structures of underlying data. Among the three display formats reviewed, scatter displays most faithfully reflect underlying data structures. In a scatter display, the viewer is not constrained to follow predetermined links as in the network display or to follow a rigid hierarchical structure as in the hierarchical display. However, this lack of regularity in the scatter display also poses problems for the viewer trying to discover the underlying structure. In this respect, the scatter display particularly needs the help of other context or interactive probes such as verbal labeling or mouse sensitive areas (e.g., when the mouse moves in, some verbal descriptions will pop up).

In summary, the scatter display is a result of a mapping from the data space to a two-dimensional display. The mapping usually minimizes the errors of resemblance between the display and the data, but it may also create distortion due to the dimension reduction. Such mapping has been widely used as a data analysis tool. As a result of rapid development of computer technologies, using mapping techniques for graphical displays will be greatly increased.

Map Displays

Map displays come from the idea of applying the geographical map metaphor to the information space. The geographic map is clearly the best example of using graphical displays to show large amount of information and their relationships. The geographical map creates a spatial analog for the physical space based on the survey data of the space. It represents any size of the physical space in a limited display framework and provides various levels of details to support different needs of the user. It keeps a dedicated balance among scale, content, and depiction for an effective representation of the physical space. It uses a set of conventional and elaborate signs and symbols to reveal properties of geographic data such as shapes, locations, and distances.

Adapting such a metaphor to information retrieval is much desired by researchers. Doyle (1961) was one of the earliest researchers who articulated the need of ‘‘semantic road maps’’ for retrieval systems. He envisioned a map that could provide ‘‘a view of the entire library at a distance’’ and help the searcher to ‘‘narrow his focus by recognition’’ (Doyle, 1962). To construct such a map, he emphasized that ‘‘natural characterization and organization of information can come from the analysis of frequencies and distributions of words.’’ He believed that such semantic road maps could ‘‘increase the mental contact between the searcher and the information store.’’

Much can be learned from the geographical map for the purpose of visual representation of conceptual space, since there are many similarities between mapping a conceptual space and mapping the physical space (Fisher, 1982). However, compared to the physical space, the document space is much less clearly defined in terms of its measurement, its dimensionality, and its semantic relationships, all of which largely depend on the selected indexing process. Furthermore, there is no apparent spatial organization of the document space, a perceived spatial organization can only be created through mapping the document space to a two-dimensional map. Thus, it would be difficult to have a map that is a ‘‘true’’ representation of the document space like the geographical map is for the physical space. Instead, the map for the document space, if it can be created, should emphasize the ‘‘display’’: The map should be a display (not a representation) of the document space; it should only be a snapshot of the document space with its angles and views defined by the selected indexing and mapping algorithms. It should be dynamically created to reflect the dynamic nature of the document space, and it should have structures that ‘‘emerge’’ from the underlying data, rather than be ‘‘imposed’’ by external programs or other artificial means (pre-defined hierarchies, for example). The map displays should also provide rich visual information, and be able to present dynamic displays at different detail levels to allow the user to interact with the underlying information.

To test these ideas, this article proposes a map display to show both contents and structures of a document space. Like the geographical map, the map display is based on a ‘‘survey’’ of all possible inputs. It ‘‘defines’’ a framework for the space it surveys, and displays selectively only the major concepts of the space. It shows contents and semantic relationships of documents by various spatial and visual cues such as distances, links, clusters, areas, and neighborhoods. The map display was designed to provide the advantages of mapping, linking, and clustering as in the scatter displays, network displays, and hierarchical displays reviewed earlier. The mapping algorithm selected will keep the display structure as similar to the underlying data structure as possible. Meanwhile, links and clusters will also be displayed to support associations and visual inference, and to simplify the display format while revealing the underlying structure. Finally, the map displays also endure some naturally occurring ambiguities to allow individual viewers to impose their interpretations on the display depending on their interests at the time of observation.

Construction of Map Displays

To generate the map display, various mapping techniques and graphical tools need to be explored and tested. The map displays given in this article were the result of a study on the application of Kohonen's feature map algorithm to document visualization. With an appropriate indexing, the algorithm can be used to ''survey'' contents of a document space, to ''detect'' semantic relationships of terms and documents, and to generate map displays that will show both contents and semantic relationships of documents. Detail properties and features of such map displays are presented through several examples in this section after a brief discussion of the feature map algorithm and the document indexing process.

Kohonen's Feature Map Algorithm

Kohonen's feature map algorithm is a major unsupervised learning method in the family of artificial neural networks (Kohonen, 1989). The algorithm takes a set of input objects, each represented by an N-dimensional vector, and maps them onto nodes of a two-dimensional grid. The mapping procedure is a recursive learning process of the following:

- Select an input vector randomly from the set of all input vectors,
- find the node (which is also represented by an N-dimensional vector called weights) closest to the input vector in the N-dimensional space,
- adjust weights of the node (called the winning node), so that it will more likely be selected again if this input is presented later,
- adjust the weights of those nodes within a neighborhood of the winning node, so that nodes within this neighborhood will have similar weight patterns.

This process goes through many iterations until it converges, i.e., the adjustments all approach zero. To ensure its convergence, two control mechanisms are imposed. The first is the updating parameter. It approaches to zero as the number of iterations increases. The second is the neighborhood structure that shrinks gradually during the process. A large neighborhood will achieve ordering and a small neighborhood will help to achieve a stable convergence of the map (Kohonen, 1989). By beginning with a large neighborhood and then gradually reducing it to a very small neighborhood, the feature map achieves both ordering and convergence properties. In practice, the process is usually divided into two phrases. The first phrase takes relatively fewer iterations and starts with a large neighborhood; this phrase aims at achieving the ordering. The second phrase is for fine adjustments of the feature map. It starts with both small neighborhoods and small updating parameters, and requires many more iterations to complete.

As the result of the recursive learning process, a grid of nodes with randomly assigned weights will be transferred into an orderly feature map that reflects the patterns or structures of the input. The mapping process not only categorizes the input data but also preserves local neighborhood relationships among the input data. It creates a ''continuous elastic surface'' and distributes the input data on the surface based on their statistical patterns (Ritter & Kohonen, 1989).

The algorithm has been applied to various practical problems in areas such as speech recognition, robot arms control, optimization problems, and analysis of semantic information (Kohonen, 1990). Early applications of the algorithm mostly demonstrated that the feature map could preserve metric relationships and the topology of input patterns. The application of the feature map to cognitive information processing (Ritter & Kohonen, 1989) particularly offered the possibility ''to create in an unsupervised process topographical representations of semantic, nonmetric relationships implicit in linguistic data'' (p. 243).

Document Indexing

A pre-process to convert documents to vectors is needed before Kohonen's algorithm can be applied to documents. A general procedure for such a process, based on the vector space model (Salton, 1989), includes following steps:

1. Identify a list of words from a document collection; it may include every word from the document titles, from the titles and abstracts, or from the full text of the documents.
2. Compare the list to a stop list to delete common words such as ''and,'' ''of,'' ''or,'' and ''the.''
3. Use a word-stem procedure to reduce the list to a stem form and remove duplicates.
4. Remove some of the most and least frequently occurring terms from the list;
5. Index the collection based on the remaining list. A vector is created for each document where each component of the vector can be:
   i) A binary digit—each component will be either 1 or 0 depending on whether the corresponding term appeared in the document or not;

ii) a weight based on the term frequency (the term frequency is the number of times a term appears in a document);
iii) a weight based on both the term frequency and the inverse document frequency (the inverse document frequency is the inverse of the number of documents in which a term occurs).

Choosing a different type of indexing may represent a document space differently. The map displays, based on different types of indexing, may also show different characteristics of the document space. In the following three subsections, three examples of map displays are presented, each is based on a different type of indexing.

A Map Display for a Retrieved Set of Documents

This example used a set of documents retrieved by a search done on INSPEC database in DIALOG for the topic of multilingual information retrieval. The set contains 311 documents. The indexing for this document set was based on titles only. After the stopword-removing and stemming procedures, the top three most frequently occurring words (multilingual, information, and system) were eliminated. So were those that occurred no more than 4 times. As the result, 85 terms were retained to index the document set. A vector of 85 dimensions was created for each document, where a component was a ''1'' if the corresponding term occurred in the document title and a ''0'' otherwise.

The document vectors were used as input to train a feature map of 85 input features and 10 by 14 output nodes arranged in a grid. Following the Kohonen's algorithm:

- each feature corresponds to an indexing term;
- each document vector is used as an input;
- each output node is represented by a vector of weights which are assigned small random values at the beginning of the training;
- during the training process, a document vector is randomly selected, the node whose weights are closest to the document vector is chosen as the winning node; and the weight of the winning node, as well as weights of those nodes within a neighborhood of this winning node, are adjusted to make them closer to the selected document vector.
- the training process proceeds iteratively for a certain number of training cycles. When the training process is completed, weights of every term are compared to weights of each output node and a best-match term is selected for every node. Connecting those nodes that have the same best-match term generates areas on the map displays (see Fig. 2).
- terms can also be mapped to the grid by comparing weights of each term to weights of every output node. The best-match node of a term determines where the term will be located on the map display.

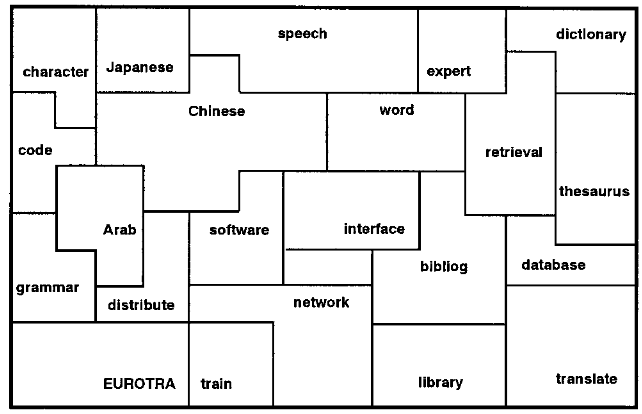


FIG. 2. The map display for documents on multilingual information retrieval. The map shows the content of the collection roughly in three parts, the languages on the left, the technologies in the middle, and the tools on the right.

Figure 2 shows the map display generated from the document collection. The map display contains very rich information:

- The areas on the map can be seen as concept areas (more precisely, word areas). These areas, determined by mapping of unit vectors after the training is completed, visualize contents of the collection. As shown on the display, multilingual information retrieval is represented by the map in roughly three parts, the languages such Chinese, Japanese, Arabic and EUROTRA (the left part), the technologies such as software, interface, networks, speech (understanding), and expert (systems), (the middle part), and the tools such as (multilingual) dictionaries, thesaurus, bibliographies and libraries (the right part).
- The size of the areas corresponds to the word occurrence frequencies. Thus, that the area ''Chinese'' is larger than the area ''Japanese'' indicates that there are more documents on Chinese text retrieval than on Japanese text retrieval in this collection. Similarly, there are likely more discussions on ''distributed networks'' than on ''interface.''
- The neighboring relationships of areas indicate frequencies of the co-occurrence of words represented by the areas. Areas of ''character coding'' is next to areas of ''Japanese'' and ''Chinese'' because they often co-occur in the documents. Thus, it may be perceived that a major issue for Asian languages' retrieval is character coding. On the other hand, ''EUROTRA'' is next to ''grammar,'' ''distributed networks,'' and ''training,'' so EUROTRA (a major project on cross-language retrieval for European languages) will likely deal with more on grammars and distributed networks than on character coding. This kind of perception will help the viewer clarify some terms without even reading any document.

A Map Display for a Personal Collection

The second example is a map display for a personal document collection. The collection contained 660 docu-

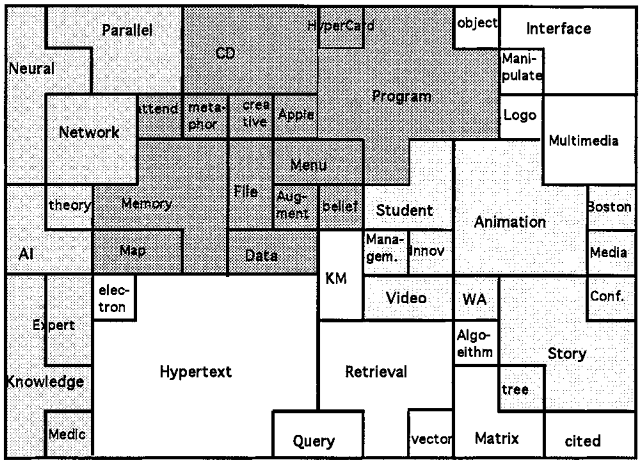


FIG. 3. A map display for a personal collection. The display shows the person's major research areas and their relationships. The size of areas indicates relative importance of the areas to the searcher. The neighboring relationships reflect associations of the areas resulting from research activities of the person.

ments, which were accumulated over many years as a by-product of a researcher's research activities. Each document in this collection contained a citation (author, title, source, etc.), an abstract, and sometimes the first page of the article. The indexing for this collection was fulltext-based—every word in the titles, keywords, and abstracts was used. After the stopword-removing and stemming procedures, and the elimination of the most-frequently and the least-frequently occurring terms, 1,472 terms remained in the indexing list. To create the indexing vectors, weights of each term were computed based on both the term frequency and the inverse document frequency. The 660 document vectors of 1,472 dimensions were then used as input to train a 10 by 14 Kohonen's feature map of 1,472 input features.

The map display generated (Fig. 3) shows the researcher's major research areas and the relationships of these areas. The size of areas, corresponding to the frequencies of the words, indicates relative importance of the areas to the researcher (the more often a word appears in the personal collection, the more likely the word will correspond to a large area in the space). The neighboring relationships, corresponding to the frequencies of co-occurring words, reflect degrees of word associations as derived from the researcher's collection. Some of these associations, such as the association of ''query'' with ''retrieval,'' are clear to anyone in the field of research, other associations may need the interpretation of the researcher. For example, without knowing that the researcher attended several conferences in Boston on animation and multimedia (and thus collected many papers from these conferences), one may puzzle why ''Boston'' appears nearby the areas marked ''animation'' and ''Multimedia.''

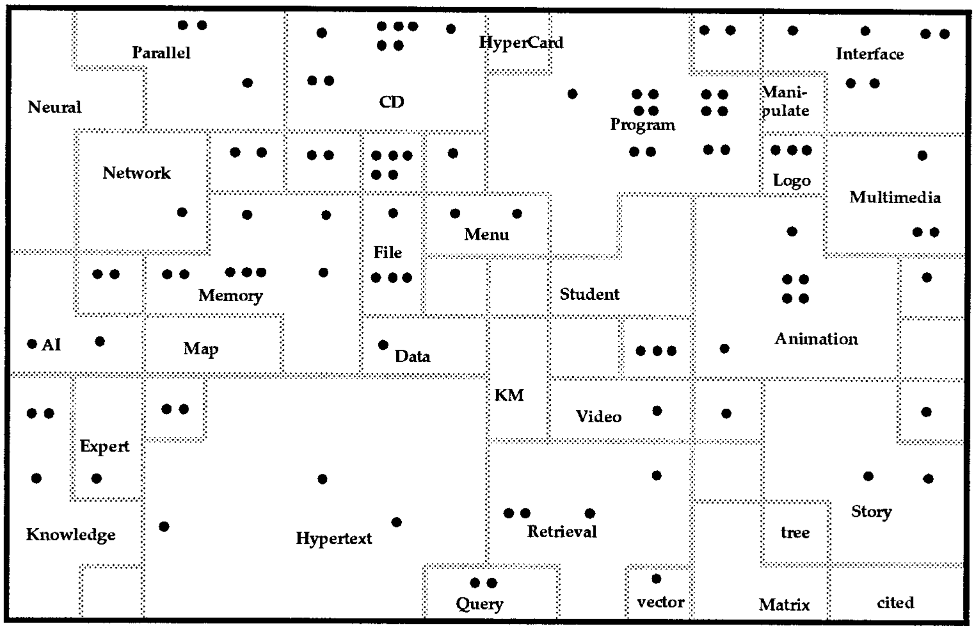
The map display appears to have a hierarchical form. Four top levels in the hierarchy can be visualized (as indicated by the four different gray areas). Within each, terms are organized by their inter-relationships. For example, within the top branch of the hierarchy on the left-hand side, ''expert (systems)'' and ''neural networks'' are located on different sides of the area ''AI,'' indicating two different but associated approaches to artificial intelligence; while ''parallel'' is close to ''neural networks,'' ''expert (systems)'' is close to ''Knowledge (base),'' and so on. Looking further into the ''neural networks'' area (through analysis of mapping results not shown on the map), we found that words mapped to this area were: ''connectionist,'' ''model,'' ''learning,'' ''process,'' ''unsupervised,'' ''net,'' ''error,'' and ''principal component analysis.'' (Evidently, the researcher is interested in comparing neural networks to principal component analysis.) Similarly, words mapped into the ''hypertext'' area included ''link,'' ''text,'' ''color,'' ''structure,'' ''browsing,'' and ''University of North Carolina at Chapel Hill'' (where the first hypertext conference was held).

When each document in the collection was mapped to a position on the display, the document distribution over the map display can also be shown. This kind of distribution may remind the researcher of the areas on which he has more documents and of the areas on which he may need to have more documents. It can also reveal migration of the researcher's interest over time. Figure 4 shows two displays that give the distributions of the first 100 documents and the last 100 documents in the collection. Since the documents were collected sequentially over time, a clear migration of the researcher's interest can be seen from an early broad distribution over several areas such as ''CD-ROMs,'' ''programming languages,'' and ''interfaces,'' to a more current concentration on areas of ''neural networks,'' ''matrix operation'' and ''retrieval.'' This kind of distributions even revealed some interesting relationships that the researcher had not thought about before.

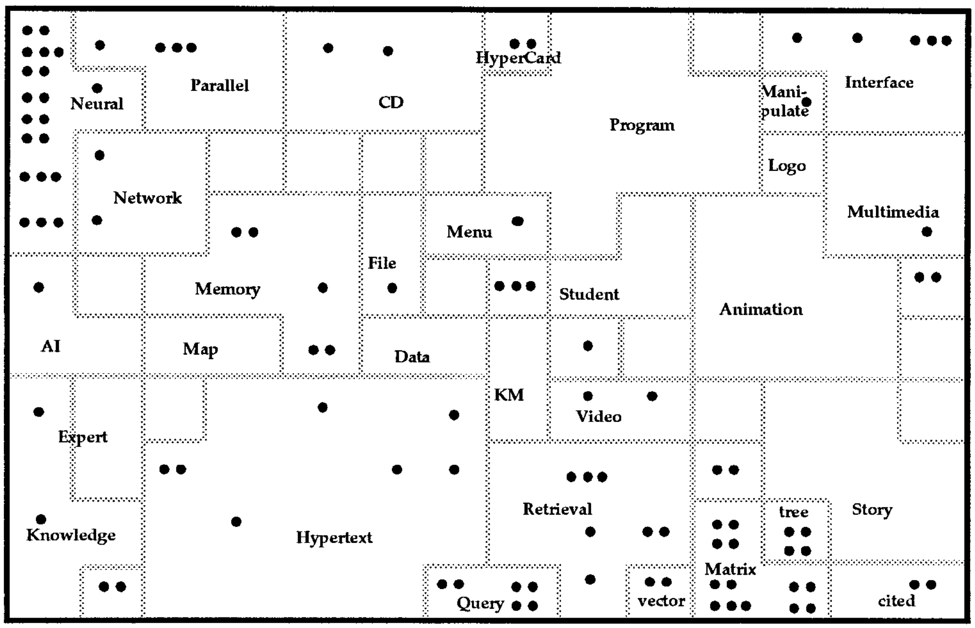
A Map Display for Conference Proceedings

The third example is about documents from 1990–1993 SIGIR conference proceedings. These proceedings contain 143 documents. The indexing terms for this collection were collected from titles only, but the weights of terms were computed based on the term frequency in titles, keywords, and abstracts. Such indexing uses the same low dimensions as the title indexing, but the indexing vectors will reflect how the indexing terms are distributed in titles, keywords, and abstracts. After the same stopword-removing, stemming procedures, and elimination of the most- and the least-frequently occurring terms, 154 terms were used to index the collection, resulting in 143 vectors of 154 dimensions. These vectors were then used to train a 14 by 14 feature map of 154 input features.

In this example, the map display was shown in a map



(a) Distribution of the first 100 documents in the personal collection



(b) Distribution of the latest 100 documents in the personal collection

FIG. 4. Document distribution of the personal collection. The two different time periods of documents showed very different distributions, indicating changes of personal research activities. **Top:** Distribution of the first 100 documents in the personal collection. **Bottom:** Distribution of the latest 100 documents in the personal collection.

interface prototype with two interactive tools, the slider and the pop-up window. The static map displays shown in last two examples provide only one ''snap-shot'' of the trained feature map. Much more information in the feature map can be used for display. The prototype interface uses the ''activation'' level of terms and documents to create the two interactive tools. The slider on the bottom of the map display is used to control the number of terms shown on the display. When the slider is moved to the left, only those terms with high activation levels are displayed. When the slider is moved toward the right, terms with lower activation levels are added to the display. Because the activation level of a term indicates frequencies of term occurrence and co-occurrence, those with high activation levels can be regarded as ''major'' terms in the collection, which are shown first by the inter-

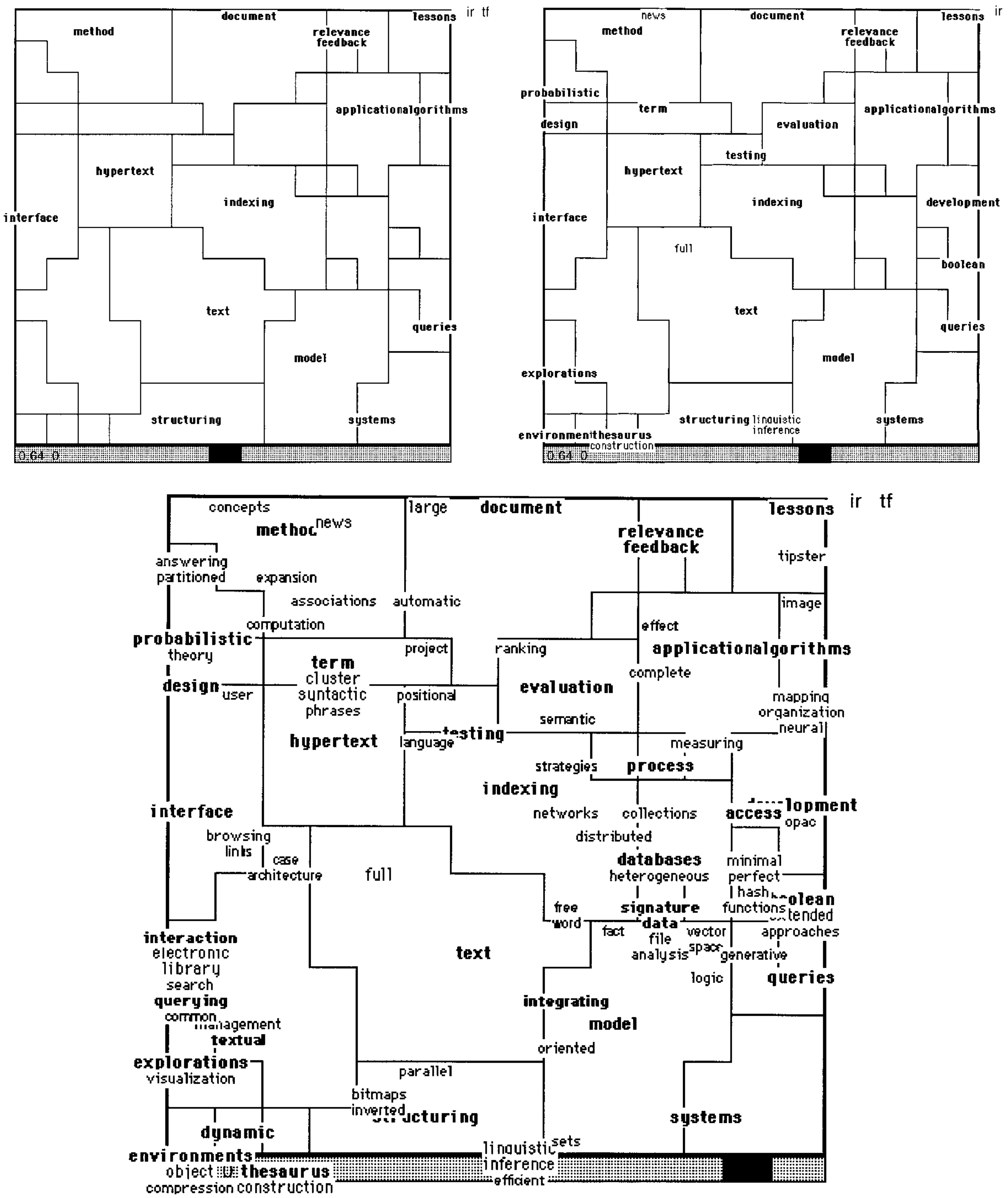


FIG. 5. Map displays for SIGIR conference proceedings. The map displays are shown on a map interface prototype which has two interactive tools, the slider and the pop-up window. When the slider is moved from left to right, more terms are added to the map on the locations determined by their associations (**a–c**). When a term is selected by clicking on it, a window will pop-up to show titles associated with the term. These titles are also arranged by their associations. Two windows were shown in (**d**) after ''document'' and ''text'' were clicked (this is for demonstration only. Normally only one window will show up at a time). They can be compared with those shown in Figure 6.

face. Figure 5(a–c) shows the SIGIR map display in three levels: First, the display shows only a few terms; these terms are considered to be the major ones (such as, in this case, ''document,'' ''text,'' ''indexing,'' ''interface,'' ''relevance feedback,'' etc.); as the slider is moved to the right, more terms are added to the display (''testing,'' ''evaluation,'' ''thesaurus construction,'' ''probabilistic,'' ''(interface) design,'' ''development,'' etc.); these terms are shown on the display at the locations determined by their associations. Finally, when the slider is moved further right, more detailed terms are added to the display; they clarify some terms shown early (such

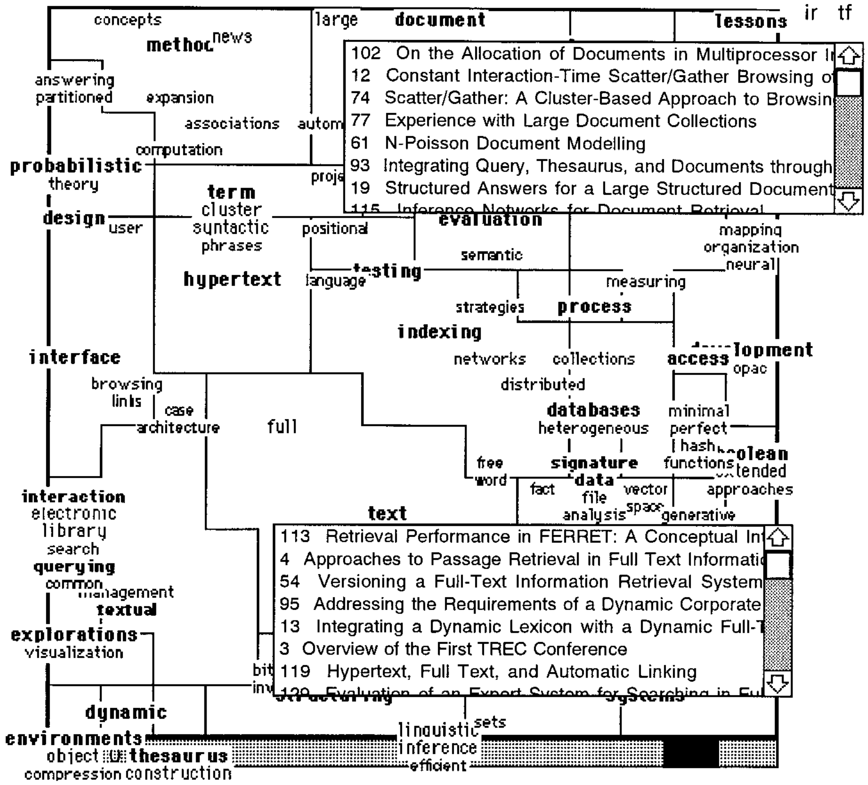


FIG. 5. Continued

as "probabilistic (theory)," "(user) interface design," and "(OPAC) development and access") and give more explanations for other areas (such as, terms related to "structure" are "bitmaps," "inverted," and "parallel," and terms related to "model" are "vector space," "logic," "data," "file," "integrating," and "set-oriented").

Using different levels of displays will show very rich semantic information about the collection. The user of such an interface can decide at what level of details the map display should be shown, knowing that the levels can always be changed as needed. A user will typically start with a few terms on the map to browse their relationships. If a term or an area is perceived to be related to his/her information needs, he/she will naturally focus on that area of the display when more terms are added to the map display. If the map display becomes confusing, the user can always move back the slider to reduce the number of terms on the display. When one or several terms are identified to be associated with the needed information, the user can click on that location and a pop-up window will show the top 10 titles associated with this location (Fig. 5d). These titles are arranged by their associative weights. They are likely semantically related to the terms identified. By browsing through these related documents, the user will understand better what concepts or subjects are associated with this particular location. If these documents are not what the user is looking for, he/she can adjust the clicking location based on the terms shown, or move the slider to get more terms. Since Kohonen's map is assumed to be a "continuous elastic surface," how the titles and terms are associated will become clearer to the user after he/she moves around the surface for a while (A new implementation of the interface can be seen at the web page http://www.uky.edu//~xLin/sitemap.html).

The user can click on any location of the map to pop-up the title window. A major difference of the term display and the title display is that while each term is mapped to a unique location on the map, each title may be shown in multiple locations based on their associations with the terms. Therefore, a title can be found in several places, depending on how many concepts or terms it associates with. This unique feature of multiple categorizations is believed to be desirable in the information retrieval environment.

Discussions

The above three examples demonstrate that map displays can be generated based on the underlying document space. The analysis also showed how the map displays revealed contents and semantic relationships of documents. Nevertheless, much more research is needed to study issues related to the mapping process, map interface design, and use of map interface in retrieval systems. This section discusses some of the issues that need further research.

The Mapping Process

The application of the feature map algorithm to visualizing a term-document space seems to be successful. As a mapping tool, the feature map has the properties of economic representation of underlying data and their interrelationships (in our examples, the feature map self-organizes major terms selected from hundreds or thousands of indexing terms to represent the document spaces). As a visualization tool, the feature map produces rich geographical features that can be used for visual inferences. Adding to these strengths is its output form. The algorithm generates an associative network as the output, rather than the direct mapping of the input. This makes it easy to implement various interactive tools and provide different ''views'' of the underlying information. Finally, that the algorithm allows classification of any input to more than one location is certainly beneficial to information retrieval.

However, the success of the mapping process also depends on the indexing process. How to select different types of indexing for the map display needs to be studied. While it is clear that increasing the input dimension will dramatically increase the processing time needed by the algorithm, it is not clear whether the map display will be improved when the indexing is changed, for example, from title indexing to full-text indexing, which typically increases the number of input dimensions from hundreds to thousands. Because the limited display space can only show a restricted amount of information, the map display may just be able to show as much information from those inputs of thousand-dimensions as it can from those inputs of hundred-dimensions. Studies are needed to compare map displays generated by different types of indexing or by different numbers of input dimensions.

The feature map algorithm also needs to be compared with other self-organizing algorithms or mapping procedures such as Principal Component Analysis (PCA) and multi-dimensional scaling. Figure 6 shows the result of PCA mapping of the SIGIR collection. The display is based on the first two dimensions of the PCA output, and each document is represented by its sequential number (the same numbers were used in the pop-up windows in Fig. 5d). Some clusters on this display are easily comparable to the map display of Figure 5(d) (the two dotted areas include almost identical document sets as the two windows in Fig. 5d). However, a comprehensive comparison of the two displays is not so apparent. On this display, each document is mapped to only one location, and it is difficult to map any term labels to this display. On the map display, clusters are dynamically created by clicking on different locations. Similarities and differences between the two displays may need to be studied empirically.

The other way to learn the mapping process is to observe how human subjects arrange a map display based on given documents. An experiment was conducted in which subjects were asked to put 133 document titles on a 10 by 14 grid. In that experiment, how the subjects approached the task was observed. What decisions the subjects needed to make and what difficulties they confronted during the mapping process were analyzed. It was found that the subjects seemed to follow either the group approach (putting a group of titles on the map before organizing them) or the individual approach (adding titles to the map individually), and the maps they generated were either association-based or category-based. Kohonen's mapping is similar to the ''individual approach,'' and its map is ''association-based.'' Details of the experiment have been reported in Lin, Marchionini, & Soergel (1993).

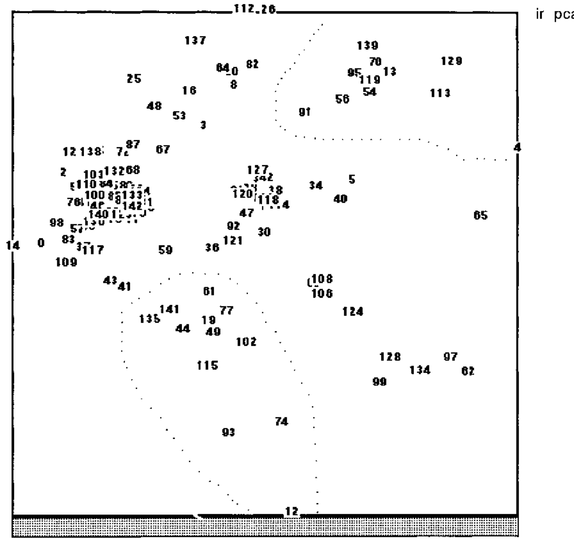


FIG. 6. The PCA mapping of SIGIR conference proceedings. The PCA algorithm was applied to 143 documents in SIGIR conference proceedings and the result was mapped based on the first two dimensions. Documents shown in the two dotted areas are almost the same as those shown in the two windows in Figure 5(d).

Comparison of Map Displays

A major challenge to the success of document mapping is how to evaluate map displays and how to compare different map displays. Because there are many different indexing methods and many different mapping algorithms, there can be many different ''views'' of the same document space. The map displays may all look different but all reflect semantic structures of the document space. It would be difficult to compare different map displays just based on the static, two-dimensional ''views'' of the map displays. Dynamic displays as well as interactions with the map displays must be included in the process of evaluations. Furthermore, functionality of the map displays and their usefulness in the information seeking environment need to be tested in a real retrieval environment. We conducted an experiment on using the map displays (both human-generated and machine-generated) for some simple retrieval tasks (Lin, 1993). Subjects were given one title at a time, and were asked to locate the same title on the map displays. It was found that there was a significant difference between the time spent on these map displays and the time spent on the random map display, but there were no significant differences between the time spent on the machine-generated map display and on the human-generated map display, even though the map displays looked different. This result suggests that comparisons of map displays need to be done on how the map displays help the user locate documents, not just how they look. It is quite possible to have different organizations of map displays that can provide the same level of access to a document space. This experiment was conducted on paper maps of table-size grids. More experiments of this kind are needed for the map interface in a more realistic retrieval environment.

Interactions with the Map Displays

The map displays are essentially a new type of indexing that makes it possible to reorganize documents or terms based on their associative relationships. This new organization can be visually displayed to reveal different aspects of the contents and relationships. The user can interact with the visual display to control or manipulate various views for effective searching and browsing. The slider and the clickable pop-up window discussed in this article are two examples of such interactive tools. They are designed to provide functions that we learned from the two experiments. Based on observations and comments collected from the two experiments, we identified two essential functions that the user needed to interact with the map displays. One is to help the user quickly focus on one area of the map displays for scanning or browsing—when the subjects were able to say ''It has got to be in this area,'' they have much more confidence and success in locating the document. The other is to help the user make judgments on whether or not the right location is selected—a display of related titles is useful for this kind of judgment.

The success of the interaction will rely on the associative organization, or the ''semantics'' of the map displays. When related terms and documents are brought together on the map displays, the interaction will help to develop a ''sense'' or ''orientation'' of the maps, which will allow the user to become familiar with the maps. This is another advantage that the map displays may bring to the user.

Applications of the Map Displays

Implementation of the map interface needs a high computational power, particularly when the input dimension is high. The training time (not including the indexing time) for the three examples ranged from 202 seconds in a Cray super computer for the second example, and the 255 and 353 seconds for the first and the third examples in a convex machine. The training time needs to be reduced if the map displays are to be generated online. One way to do it is to show the map display gradually. A primary map can be shown to the user when major areas and term locations are basically defined. Since the user will need some time to explore the map display, the computer can continue to refine the map display while the user is interacting with the map display. The possibility and the effect of transitions, from the primary map display to a fine-tuned one, need to be studied.

The map display may best be applied to situations where a few hundred documents are involved. Three examples given in this article are, in fact, three types of potential applications of the map displays. First, the map display may be implemented as a front-end interface to online databases. The user can type a broad query (such as ''multilingual information retrieval'') to retrieve a few hundred documents. These documents and terms associated with them will be organized and presented to the user on the map interface. The user then can interact with the map interface to explore semantic relationships of the terms and documents. Such an interface will effectively increase the number of documents that the user is willing to browse (in a high-recall and low-precision situation, for example), and will suggest many terms to the user when he or she needs to modify the query. Second, the map interface may be implemented in personal information systems. Such a map interface will help a researcher to maintain and organize a personal collection, and may even help to monitor new documents over the network—any documents that cause substantive activation on the map will likely be interesting to the researcher since the map display is based on his or her personal collection. The third application is a ''graphical table of contents'' for conference proceedings or other electronic books. A map display such as the SIGIR map will provide not only an overview for contents of several books, but also access points that can be used to locate articles in these proceedings or books.

Conclusions

The main purpose of this article is to demonstrate that, by using appropriate spatial mapping techniques, map displays can be generated to visualize contents and semantic relationships of a document space. Such map displays can be seen as an example of creating visual cues in a retrieval interface to allow visual and perceptual inferences during the search process. Several examples of map displays showed that the map displays could be used both as an overview tool and an access or exploration tool, and interactive tools implemented on the map displays might be particularly useful in assisting the user to see and interact with the rich information revealed by the map displays. This article also identifies many issues and research questions that need to be addressed when we are continuing to search for visual displays that will improve effective use of human visual capabilities for information seeking in the electronic environment.

Acknowledgments

The main part of this article is revised from the author's dissertation. The author thanks members of the dissertation committee for their comments, suggestions, and other input that helped to improve this work.

References

Bates, M. J. (1986). An exploratory paradigm for online information retrieval. In B. C. Brooks (Ed.), *Intelligent Information Systems for the Information Society: Proceedings of the 6th IRFIS Conference* (pp. 91–99). New York: North-Holland.

Beard, D. V., & Walker, J. Q. (1990). Navigational techniques to improve the display of large two-dimensional spaces. *Behaviour & Information Technology, 9(6)*, 451–466.

Belkin, N. J., & Croft, W. B. (1987). Retrieval techniques. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology, 22*, 109–145. Amsterdam: Elsevier.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval. Part I: Background and theory. *Journal of Documentation, 38*(2), 61–71. Part II: Results of a design study. *Journal of Documentation, 38(3)*, 145–167.

Burgess, C., & Swigger, K. (1986). A graphical database interface for casual naive users. *Information Processing & Management, 22(6)*, 511–521.

Chang, S. J., & Rice, R. E. (1993). Browsing: A multidimensional framework. *Annual Review of Information Science and Technology, 28*, 231–276.

Conklin, J. (1987). Hypertext: An introduction and survey. *Computer, 20*, 17–41.

Conklin, J., & Begeman, M. (1989). gIBIS: A tool for all reasons. *Journal of the American Society for Information Science, 40*, 200–213.

Cutting, D. R., Karger, D. R., Pedersen, J. O., & Tukey, J. W. (1992). Scatter/gather: A cluster-based approach to browsing large document collections. *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, (pp. 318–329). New York: ACM.

Doyle, L. B. (1961). Semantic road maps for literature searcher. *Journal of the Association for Computing Machinery, 8(4)*, 553–578.

Doyle, L. B. (1962). Indexing and abstracting by association. *American Documentation, 13*, 378–390.

Fairchild, K. M., Poltrock, S. E., & Furnas, G. W. (1988). SemNet: Three-dimensional graphic representations of large knowledge bases. In R. Guindon (Ed.), *Cognitive science and its applications for human-computer interaction* (pp. 201–233). Hillsdale, NJ: Erlbaum.

Feiner, S. (1988). Seeing the forest for the trees: Hierarchical display of hypertext structure. In *Proceedings of the Conference on Office Information Systems* (pp. 205–212). New York: ACM.

Fisher, H. T. (1982). Mapping information: The graphic display of quantitative information. Cambridge, MA: Abt Associates.

Foss, C. L. (1989). Tools for reading and browsing hypertext. *Information Processing & Management, 25(4)*, 407–418.

Fowler, R. H., Fowler, W. A., & Wilson, B. A. (1991). Integrating query, thesaurus, and documents through a common visual representation. *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval* (pp. 142–151). New York: ACM.

Frei, H. P., & Jauslin, J. F. (1983). Graphical representation of information services: A user-oriented interface. *Information Technology: Research and Development, 2(1)*, 23–42.

Friedhoff, R. M., & Benzon, W. (1989). *The second computer revolution: Visualization.* New York: Abrams.

Furnas, G. W. (1986). Generalized fisheye views. *Proceedings of Human Factors in Computing Systems (CHI'86)* (pp. 16–23). New York: ACM.

Furnas, G. W., Deerwester, S., Dumais, S. T., Landauer, T. K., Harshman, R. A., Streeter, L. A., & Lochbaum, R. (1988). Information retrieval using a singular value decomposition model of latent semantic structure. In Y. Chiaramella (Ed.), *Proceedings of the Eleventh Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval* (pp. 465–480). Grenoble, France: Presses Universitaires DE Grenoble.

Halasz, F. G., Moran, T. P., & Trigg, R. H. (1987). NoteCards in a nutshell. *Proceedings of the 1987 ACM Conference of Human Factors in Computer Systems* (pp. 45–52). New York: ACM.

Hammond, N. & Allinson, L. (1988). Travels around a learning support environment: Rambling, orienteering, or touring? *Proceedings of CHI'88: Human Factors in Computing Systems* (pp. 269–273). New York: ACM.

Kohonen, T. (1989). *Self-organization and associate memory* (3rd ed.). New York: Springer-Verlag.

Kohonen, T. (1990). Some practical aspects of the self-organizing maps. *IJCNN-90-WASH-DC, Vol. II, Application track* (pp. 253–256). Hillsdale, NJ: Erlbaum.

Korfhage, R. R. (1991). To see, or not to see—is that the query? *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval* (pp. 134–141). New York: ACM.

Kunii, T. I. (1989). *Visual database systems.* New York: North-Holland.

Lin, X. (1993). *Self-organizing semantic maps as graphical interfaces for information retrieval.* Unpublished doctoral dissertation, University of Maryland, College Park.

Lin, X., Marchionini, G., & Soergel, D. (1993). Category-based and association-based map displays by human subjects. *Proceedings of the 4th ASIS SIG/CR Classification Research Workshop* (pp. 147–164). Silver Spring, MD: ASIS.

Marchionini, G. (1987). An invitation to browse: Designing full-text systems for novice users. *Canadian Journal of Information Science, 12(3–4)*, 69–79.

Marchionini, G. (1995). *Information seeking in electronic environments.* New York: Cambridge University Press.

Motro, A. (1986). BAROQUE: A browser for relational databases. *ACM Transactions on Office Information Systems, 4*, 164–181.

Quillian, M. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227–270). Cambridge, MA: MIT Press.

Ritter, H., & Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics, 61*, 241–254.

Robertson, G. G., Card, S. K., & Mackinlay, J. D. (1993). Information visualization using 3D interactive animation. *Communications of the ACM, 36(4)*, 57–71.

Robertson, G. G., Mackinlay, J. D., & Card, S. K. (1991). Cone trees: Animated 3D visualizations of hierarchical information. *CHI'91 Conference Proceedings: Reaching through technology* (pp. 189–194). New York: ACM.

Rose, D. E., & Belew, R. K. (1991). A connectionist and symbolic hybrid for improving legal research. *International Journal of Man-Machine Studies, 35(1)*, 1–33.

Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer.* Reading, MA: Addison-Wesley.

Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers, 18(5)*, 401–409.

Schatz, B. R., & Caplinger, M. A. (1989). Searching in a hyperlibrary. *Proceedings of the Fifth International Conference on Data Engi-*

neering (pp. 188–197). Washington, DC: IEEE Computer Society Press.

Siedlecki, W., Siedlecki, K., & Sklansky, J. (1988). An overview of mapping techniques for exploratory pattern analysis. *Pattern Recognition, 21(5)*, 411–430.

Spath, H. (1980). *Cluster analysis algorithms.* Chichester, UK: Ellis Horwood.

Thompson, R. H., & Croft, W. B. (1989). Support for browsing in an intelligent text retrieval system. *International Journal of Man-Machine Studies, 30,* 639–668.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Turo, D., & Johnson, B. (1992). Improving the visualization of hierarchies with treemaps: Design issues and experimentation. *Proceedings of Visualization '92,* Boston, October 21–23, 1992, (pp. 124–131). Los Alamitos, CA: IEEE.

Veith, R. H. (1988). *Visual information systems: The power of graphics and video.* Boston: G. K. Hall & Co.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intelligent structure. *Journal of the American Society for Information Science, 32,* 163–171.

Willett, P. (1988). Recent trends in hierarchical document clustering: A critical review. *Information Processing & Management, 24(5)*, 577–597.