Integrating Query, Thesaurus, and Documents through a Common Visual Representation

Richard H. Fowler, Wendy A. L. Fowler, and Bradley A. Wilson

University of Texas - Pan American
Edinburg, TX 78539-2999
fowler@panam.bitnet

Abstract

Document retrieval is a highly interactive process dealing with large amounts of information. Visual representations can both provide a means for managing the complexity of large information structures and support an interface style well suited to interactive manipulation. The system we have designed utilizes visually displayed graphic structures and a direct manipulation interface style to supply an integrated environment for retrieval. A common visually displayed network structure is used for query, document content, and term relations. A query can be modified through direct manipulation of its visual form by incorporating terms from any other information structure the system displays. An associative thesaurus of terms and an interdocument network provide information about a document collection that can complement other retrieval aids. Visualization of these large data structures makes use of fisheye views and overview diagrams to help overcome some of the difficulties of orientation and navigation in large information structures.

1. Introduction

1.1 Visualization

Visualization is emerging as a central concern for a wide range of tasks. Visualization brings the human visual system with its pattern recognition capabilities to bear in either discovery or portrayal of information [30]. In scientific visualization large complex data sets are displayed in ways allowing the investigator to view the global nature of numerical solutions and visually explore analyses [32]. Visual programming systems express algorithms in ways that might improve clarity and simplicity of expression [39], as well as manage complexity in large systems [25]. Knowledge engineering environments often supply graphic representations of rules and facts [31]. Even hypertext systems, with entirely textual content, typically rely on visual representations for navigation and orientation [9].

One of the common goals of visualization is to help manage and understand large amounts of data or information. From this perspective, information retrieval systems for large text databases are natural candidates for visualization techniques [14]. Indeed, some thirty years ago Doyle [16] explored the use of visual representations for several components of an information retrieval system at the same time that Sutherland [41] pioneered work in graphic display and manipulation. Recent work on visualization in information retrieval has tended to focus on individual elements of document retrieval systems. A number of systems have included graphic thesaurus displays [6] and visual representations of inter-document relations [13]. Work on query formation and modification has explored visual interfaces for Boolean queries [1] and relational database queries [8], as well as queries constructed as term vectors [12] and networks [10].

However, in text retrieval systems the *integration* of visualization techniques supporting search is yet to be fully exploited. This is perhaps surprising. Today's widespread end-user searching of bibliographic databases provides a population of naive users that might particularly benefit from visual representations. The integration of a common visual representation and interaction style for retrieval aids throughout the retrieval process has been one of our principal goals [21].

1.2 Different Information Needs Require Different Tools

Retrieval systems must supply mechanisms for meeting quite different information needs, as well as support the complex behavior of user information seeking. Information needs can be categorized as those arising when 1) direct bibliographic access is required, 2) the domain is well-known to the user, and 3) the domain is not known to the user [27]. Different retrieval techniques and aids are appropriate for meeting different types of information needs, due in part to the need to provide the user information about indexing and the subject domain. One means to overcome the limitations of a single technique is to enhance a system's flexibility by providing the user a variety of tools and retrieval techniques that provide multiple paths of access to information [2]. Another goal of our system is to develop an environment that supplies a range of alternative retrieval aids.

Various methods for retrieval have been investigated to overcome some of the well-known shortcomings of techniques based on a single best match of a user's query and document representations [4]. Automatic query refinement using relevance feedback [29] and the use of expert systems [20] can automate some of tasks a user or search intermediary would perform. More often, indexing aids such as a thesaurus and feedback from search results play the largest roles in allowing the searcher to refine his or her query. Regardless of the system aids available to the user, query refinement is an iterative process. The iterative nature of information retrieval, together with the need for system flexibility, require an environment in which interaction is natural and straightforward.

1.3 Direct Manipulation and Visual Representation

Accounts of human-computer interaction provide a number of insights into the design and use of visual systems [33] - systems incorporating direct manipulation in the interface [38]. The essential characteristic of direct manipulation systems is that changes in the underlying system components are reflected in visual changes in the interface objects representing those objects. The user operates on the visual interface objects to effect change in the system state. In the ideal interface the user feels directly engaged with the underlying task, rather than with an interface which in turn directs the change of state. Among factors contributing to a feeling of direct engagement with system objects are the cognitive effort required to change the system state and evaluate a resulting state. To provide an environment for the visual manipulation of system objects it is necessary that the output of one process serve as the input to another process: inter-referential input/output [17]. In a visual environment for information retrieval, components for query formulation, retrieval aids, document representations, and as many system components as possible should share a common representation. This shared visual representation allows inter-referential input/output to support the direct manipulation of information objects.

2. A System Integrating Visual Representations

The system we have designed provides the user a visual environment for direct manipulation during information retrieval. A common visual representation is used for query, associative thesaurus, conventional thesaurus and document content. The screen image in Figure 1 shows some of the major system components and examples of the visual displays available to the user.

The query process can begin with the user's entry of a natural language request for information. Term analysis (detailed below) is used to transform the text to a manipulable visual representation, as displayed in the upper middle window. Query revision can be accomplished by deleting nodes, entering more text, or dragging any term the system displays into the query window and connecting it to the query graph. Documents can be viewed both as the text of the abstract (lower right window), and as a visual network representation of the abstract text (lower left window).

Search results are displayed in the upper right window as icons of the documents retrieved. Below this window are documents the user has saved from previous searches, or a network of documents. An overview diagram for the associative term thesaurus is present in the upper left window. Subgraphs of the thesaurus may be viewed in the large window currently displaying a document abstract graph. Finally, a network reflecting inter-document relations and a conventional thesaurus (not shown in this display) are also available for browsing and retrieval. The information items are all selectable and linked so that invoking different information displays can be accomplished with minimum effort.

2.1 User Interaction

One of the goals of a direct manipulation interface is to reduce the user's cognitive load in order to enhance a feeling of engagement in the task. In the system we have designed, all interaction is performed either by selecting one of the screen buttons displayed in the top middle window or dragging one of the information items displayed as an icon. For example, a document icon can be dragged to the SAVED DOCUMENT window to place the document in a list of titles, to the CONCEPT MAP window to display the network of terms in the document, or to the DOCUMENT ABSTRACT window to view the text of the abstract. As mentioned above, a term icon is similarly manipulated to form a query and can also be used to define a view of the associative thesaurus or hierarchical thesaurus.

Display of the large networks reflecting term associations and document interrelations is based on fisheye views and the user navigates by selecting term or docu-

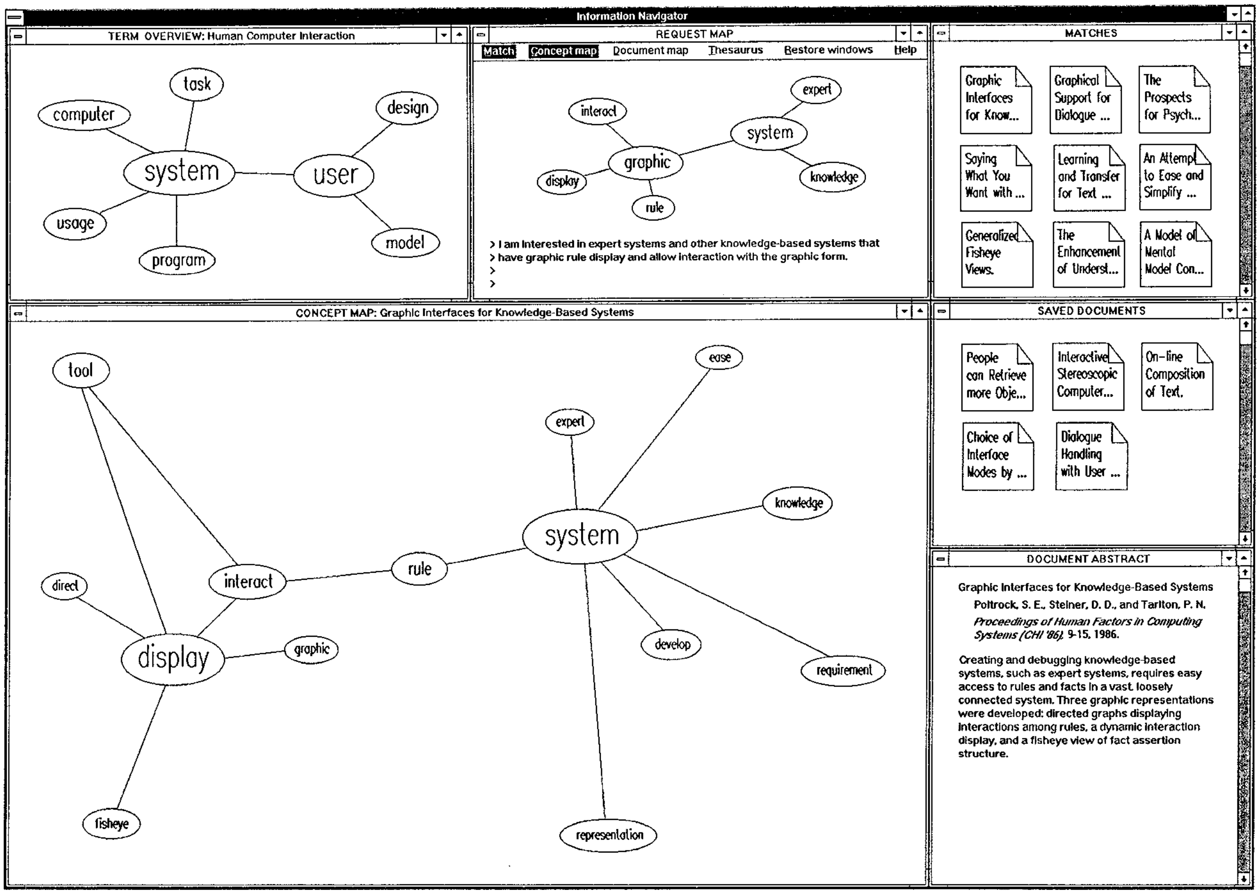

Figure 1: A typical display. The upper windows display 1) an overview diagram of the associative term thesaurus, 2) the text of a natural language query with its visual representation, and 3) an ordered list of documents retrieved in a search. Below the retrieved document icons are 4) documents saved by the user. The lower windows show 5) a document abstract network and 6) the document abstract's text.

ment nodes to define the center of the fisheye. With the display of Figure 1 the user might select the "Concept Map" button and then one of the terms from the query, document, or overview. The document network would then be replaced with a view of the associative thesaurus centered on the selected term.

2.2 The System's Associative Networks

The representations underlying the visual displays are minimum cost networks derived from measures of term and document associations. For queries, document abstracts, and associative term thesauri the associations are derived from natural language text. The network of documents is based on interdocument similarity. The statistical text analyses rely on recovering conceptual information from natural language by considering the frequency and co-occurrence of words. This basic approach has been used in a wide range of contexts and its utility and limitations are well-known [36].

Why associative structures? For the system we have designed there are three reasons for using statistically-based associative structures. One reason follows from the view that information retrieval systems should supply the user with a variety of tools and retrieval techniques. Statistically-based associative information structures provide one class of retrieval tools that can complement other retrieval aids. The associative thesaurus can encourage browsing and exploration, as well as bring the user's own associations into play. For information needs in which the user is not familiar with the domain, and indeed may not even know what his or her information needs are, the associative structures provide one means to explore and gain information to better define the information need.

A second reason for using statistically-based associative structures is the desire to have a representation that can be derived automatically in an interactive system, rather than through knowledge-engineering efforts such as are required for most deep representations. Associative structures can also serve as one component of a hybrid system incorporating both deep and shallow representations [11]. Finally, associative structures can provide a common visual representation for retrieval tools. Networks are naturally represented visually and can serve as a common representation for several information retrieval system components.

Deriving the term associations. For each database the system uses a separate set of terms that includes the most frequently occurring word stems, excluding function words. For some forms of retrieval this simple procedure suffers from the limitation that frequently occurring terms have relatively little value for discriminating among documents [40]. However, one function of the associative thesaurus is to give a picture of all concepts in a document set. The most frequently occurring terms tend to be general terms that provide useful information about the domain of the document collection.

To derive the distances between terms used to construct networks, text is analyzed by first finding the sequence of term stems present in the text. This sequence is used to assign distances between terms based on lexical distance and co-occurrence in syntactic units with a metric similar to that used by Belkin and Kwasnik [5]. Term pair similarity is calculated as the sum of values added when terms are adjacent, or occur in the same sentence, paragraph or document. These similarities provide the associations used in deriving the networks displayed by the system.

3. Pathfinder Networks: A Class of Minimum Cost Networks

The associative networks used in the system are Pathfinder networks (PFNETs). The Pathfinder algorithm was developed to model semantic memory in humans and to provide a paradigm for scaling psychological similarity data. A number of psychological and design studies have compared PFNETs with other scaling techniques and found that they provide a useful tool for revealing conceptual structure [37].

PFNETs are derived by identifying the proximities that provide the most efficient connections between entities. This is accomplished by considering the indirect connections provided by paths through other entities. In this respect it is similar to other minimum-cost methods for deriving network structures [26]. However, it is unique in the generality of the family of networks that can be generated [15]. The edge membership rule assures that edges that are a part of some minimum distance path are preserved between each pair of nodes. To derive a PFNET the direct distances between each pair of nodes are compared with indirect distances. A direct edge between two nodes is included in the PFNET unless the data contain a shorter path having two or more edges.

In constructing a PFNET two parameters are incorporated: r determines how path weight is computed and q specifies the maximum number of edges considered in finding a minimum cost path between entities. Path weight, r, is computed according to the Minkowski r-metric. It is the r^{th} root of the sum of each distance raised to the r^{th} power for all edges in a path between two nodes. The second parameter, q, determines the maximum number of edges which will be included in a path.

Pathfinder allows for systematic variation in the complexity (number of edges) in the resulting networks as the two parameters are varied. Complexity decreases as either r or q increases. As either parameter is manipulated, edges in a less complex network form a subset of the edges in a more complex network. Thus, the algorithm generates two orthogonal families of networks, controlled by r and q. The least complex network is obtained with $r = \infty$ and $q = n-1$, where n is the total number of nodes in the network, and is used in the system.

PFNETS provide an alternative to threshold networks for automatically derived network representations in information retrieval systems. Threshold networks include an edge in the final network if the magnitude of internode association is above some criterion. Edge membership is determined by examining only proximities of nodes adjacent in the original data. PFNETs differ in a fundamental way. An edge is included in a PFNET depending not on a fixed magnitude of association, but on the role it plays in determining minimum cost paths between nodes. Edge membership is determined using a global, path oriented approach that considers other connections, potentially across the entire network. Another consideration is the extent of empirical evaluation of networks used in information retrieval in revealing psychologically salient relations in edge structure. For uses in which psychological salience is important, such as our desire to have a "natural" representation suited to visual display and interaction that minimizes cognitive load, empirical assessment is important. Though statistically based graph representations used in information retrieval have sometimes assessed psychological salience [34], PFNETs have been developed from the outset as a representational scheme for human conceptual structure.

4. Visual Display of Networks

Graph display is an important issue for a number of tasks. Considering the wide application of graph structures in display, there are relatively few algorithms for drawing general undirected networks [42]. This is due in part to the difficulty of precisely specifying the perceptual and aesthetic criteria individuals use in understanding graphs [18]. Nonetheless, when it is possible to specify criteria that can be used to guide the viewer's extraction of information, such as for trees, satisfactory display algorithms can be developed [3].

The system's network displays center on visually conveying information about the networks' edge structures and weights. Several graph theoretic criteria are employed by the display algorithms. *Node degree* is the number of edges incident to a node. This descriptor is used to provide a measure of the density of the network's edge structure. *Path weight* is the weight associated with a series of edges connecting two nodes. *Path distance* is used here to refer to the number of edges connecting two ends of a path.

The system provides displays of relatively small networks for query and document abstract terms, and much larger networks for the associative thesaurus and network of documents. In the smaller networks, the position of term nodes reflects the path weight between terms. The size of a node is an indicator of connectivity with other terms in the network as measured by its degree. Display of the larger networks reflects a user's viewpoint in the network, together with nodes' path weight, path distance and connectivity to give cues for orientation and navigation.

Based on path weight, nodes are positioned using an algorithm that considers all term pairs to be connected by springs in a virtual dynamic system [28]. The strength of the spring between two nodes is determined by the minimum path weight between nodes. Nodes that are close together in the network in terms of path weight have relatively strong springs compared to nodes that are connected by longer paths. The algorithm iteratively arrives at the node positions that minimize overall tension in the system. Though the algorithm does not explicitly consider line crossings, the positioning of nodes distant in terms of network path weight as distant in terms of display distance tends to minimize crossings.

The display size of a term node reflects the number of direct connections it has with other terms. Nodes with high degree are displayed as larger than nodes having fewer connections. High degree nodes are visually emphasized by their size. The query and document term graphs shown in Figure 1 use this display technique.

5. Large Information Structures: Associative Thesaurus and Network of Documents

In addition to the query and document term displays the user can access two other visually displayed network structures: an associative thesaurus of terms, and a network of documents. The associative thesaurus is based on a PFNET of all terms in the database. For the network of documents, distances between documents are calculated using the same matching algorithm used to assess query-document similarity. Network similarity is calculated by combining the number of commons terms with a measure of structural similarity for these common terms [24].

5.1 Orientation and Navigation in Large Information Structures

With the relatively small networks for queries and document abstracts it is possible to display the complete networks. For the much larger networks of all terms and all documents only part of the complete networks can be displayed, and mechanisms allowing the user to browse, or navigate, in the large network are required. It is also useful to relate the small portion of the network being viewed to the complete network to provide the user orientation within the overall structure. Orientation and navigation are challenges shared by hypertext and other large information spaces. The navigation and orientation mechanisms used in this system are based on overview diagrams of the complete network structure and fisheye views of the detailed network view.

5.2 Overview Diagrams

Overview diagrams are a common means of supplying a user with (1) knowledge about the organization of the complete network, (2) a means for navigating the network, and (3) orientation within the complete network. In overview diagrams a small number of nodes, selected to provide information about the organization of the complete network, are displayed to the user. In addition, the nodes typically provide entry points for traversing the network. These nodes provide orientation by serving as landmarks to assist the user in knowing what part of the network is currently being viewed.

In the document collections we have used, PFNETs derived for associative thesauri and networks of documents have a characteristic structure. There tend to be a small number of nodes that have many nodes directly connected and there are relatively short paths between these highly connected nodes. There are relatively few nodes of high degree and the diameter of the network is small. This form of network suggests criteria for selecting nodes to include in overview diagrams. Overview diagrams displayed by the system, as shown in Figure 2, in-

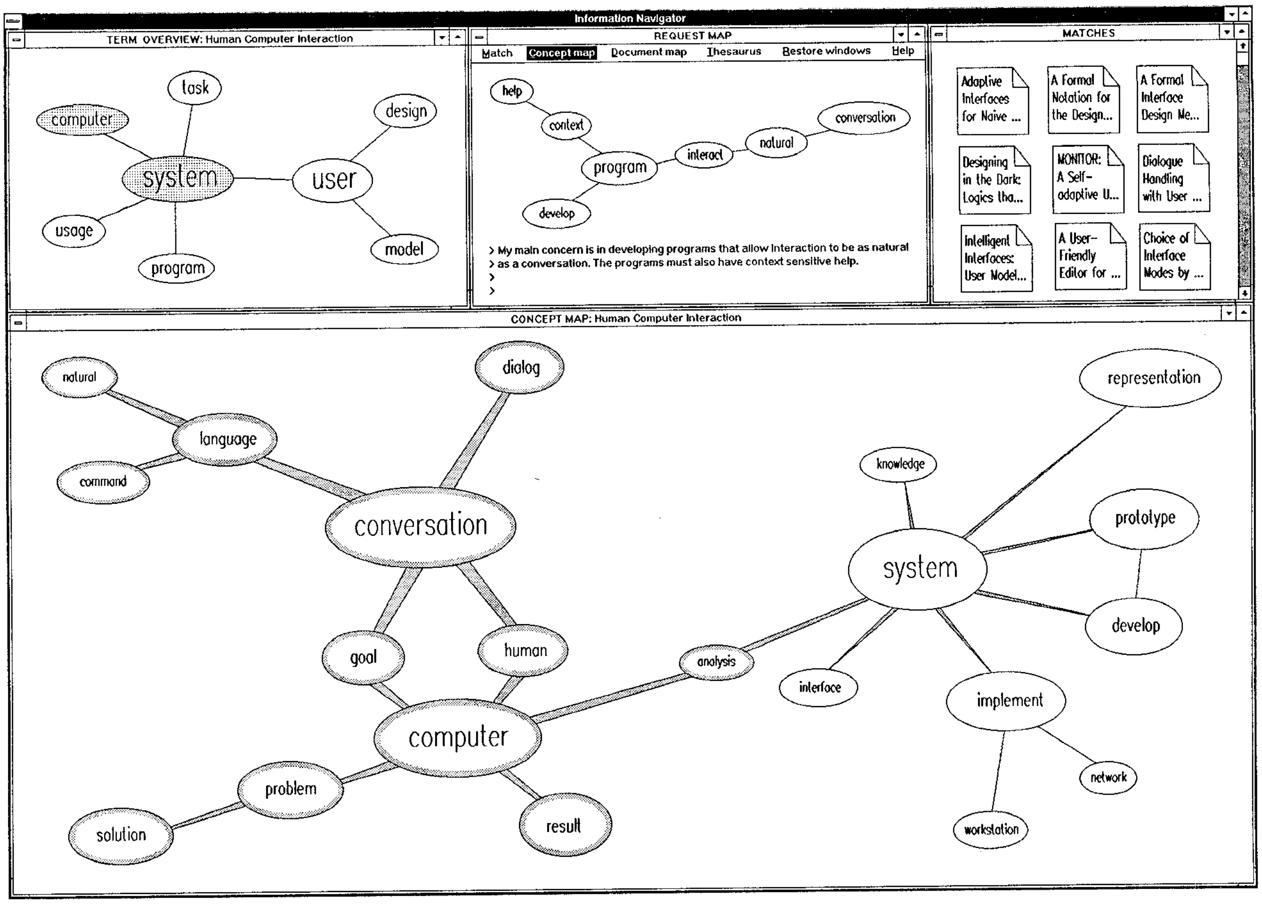

Figure 2: System display after selecting "conversation" node in REQUEST MAP window. A fisheye view of the associative thesaurus is displayed focused on "conversation". The overview diagram orients the fisheye view to the complete network by shading overview nodes present in the fisheye view.

clude those nodes of highest degree in the complete network. The overview is displayed using the same techniques employed for query and document term networks.

The terms selected for overview diagrams of associative thesauri tend to be general terms that provide a guide to the content of the database. They are landmarks in that they supply information about both the content and structure of the database. For example, in Figure 2 "system" is a central term in the overview in that it has the most connections to other terms. Its' high degree reflects frequent co-occurrence with many other terms in the document collection.

The overview diagram can also be used for navigation. By selecting one of the overview terms a fisheye view centered on the selected node can be displayed. The location of the subgraph in relation to the overview diagram is indicated by shading any node in the overview diagram that is displayed in the fisheye view. If no term from the overview diagram is present in the fisheye view, the overview term closest to the center of the fisheye as measured by path distance is shaded.

5.3 Fisheye Views

Furnas' description of fisheye views in display [22] provides a framework that is general enough to account for a number of factors important for navigation and orientation in large information spaces. A fisheye view displays objects close to the current viewing point, or focus, in more detail than things farther away. The display of elements depends on both the viewer's distance from an element and the *a priori* importance assigned element. The *a priori* importance for each element, together with the distance of each element from the viewing point to each element, allows the size (or any factor) of all elements to be related to the view point. View point and *a priori* interest are related by a degree of interest function. For example, the degree of interest function might simply compute the product of *a priori* interest and distance. The degree of interest function supplies a general mechanism for algorithmically providing orientation landmarks within information structures and has been applied effectively in large information spaces [19, 23].

5.4 The User's View of Network Detail

The system's display of network detail within the complete network is based on a fisheye view. To display a fisheye view the user selects a term to define the focus and then drags it to the CONCEPT MAP window or selects the "Concept Map" button. Figure 2 shows the screen when the node labeled "conversation" is selected from the query network. Entry points can also be selected from a document's network.

For the fisheye display a node's *a priori* importance is its degree. Viewing distance is calculated as the path weight of the focus to other nodes. Using these two measures, the system's degree of interest function yields a value for each node in the network. A threshold degree of interest value is used and nodes with values above this criterion are displayed. The value of the threshold reflects the size of the display window, so that an appropriate number of nodes will be shown. Having determined which nodes to display, the node layout algorithm for small networks is used to position nodes.

The visual form of nodes and edges conveys additional information about the relations among nodes in the fisheye view. The degree of a node is reflected in its size. Nodes that have high values are larger, and so are more prominent in the display. The focus node is drawn as large as the largest node in the view. The large, high degree nodes are the nodes the user can explore to find the densest parts of the network.

The shape of the edges also conveys information about the structure of the network. The edges are widest at the focus node and narrow as they connect nodes that are farther away from the focus. However, the measure of distance that this narrowing reflects is not path weight, which was used in computing the degree of interest function. Instead, the narrowing of edges reflects path distance, the number of edges connecting nodes. This is useful information for navigation because not all nodes on the path will be displayed due to the degree of interest threshold criterion for display. As users browse the network by selecting new focus nodes from the fisheye view, they can use these cues about network structure to guide their exploration.

When a node in view is selected and made the new focus, a number of nodes from the earlier view will typically be included in the new fisheye view. The high degree nodes remain displayed and supply orientation information. The high degree nodes of the previous display provide a context that is elaborated by the change of focus and the recalculation of degree of interest values based on the new distances. Additionally, the network overview reflects the new focus position.

5.5 Network of Documents

The final visual structure the user interacts with is a network reflecting interdocument relations, shown in Figure 3. In this display nodes from the network of documents are shown as document icons labeled with abbreviated titles. An overview diagram is constructed and displayed in the same manner as the overview of the associative thesaurus. The network of documents can serve the same browsing function as the associative thesaurus. The same mechanisms for navigation and fisheye display are used. Manipulating the nodes of the document network, as with any document icon displayed in the system, allows the user to save the icon or view the document's abstract text and term network. Functions using the PFNET of documents provide an additional means of gaining domain knowledge and moving among visual structures.

In addition to the retrieval mechanism ordering documents by similarity to a query, a second form of retrieval is available. Documents can be retrieved using the network of documents by traversing the network starting at some entry point document. The entry point can be directly provided by the user by selecting a document icon, or determined by the system as the document that best matches the query. Additional documents are then retrieved by following the edges from the starting point in the order of a breadth first search. The sequence of retrieved documents displayed to the user is ordered by the number of edges from the entry point document.

6. Using the System Components

Users' interactions typically begin with a text description of the information need. Further interaction for query revision and the selection of a final set of documents is with the system's visual forms of terms and documents. For example, Figure 2 shows the text of a query and its visual form in the REQUEST MAP window. Selecting the "Match" button initiates a search and document icons ordered by query-document similarities are displayed in the upper right window. The user can resize any window, or select one of several system-defined arrangements. The user accesses information about a document by dragging its icon to the CONCEPT MAP window to display its term network or the DOCUMENT ABSTRACT window to display the text, as illustrated in Figure 1. The SAVED DOCUMENTS window is available to the user to form a list of selected documents.

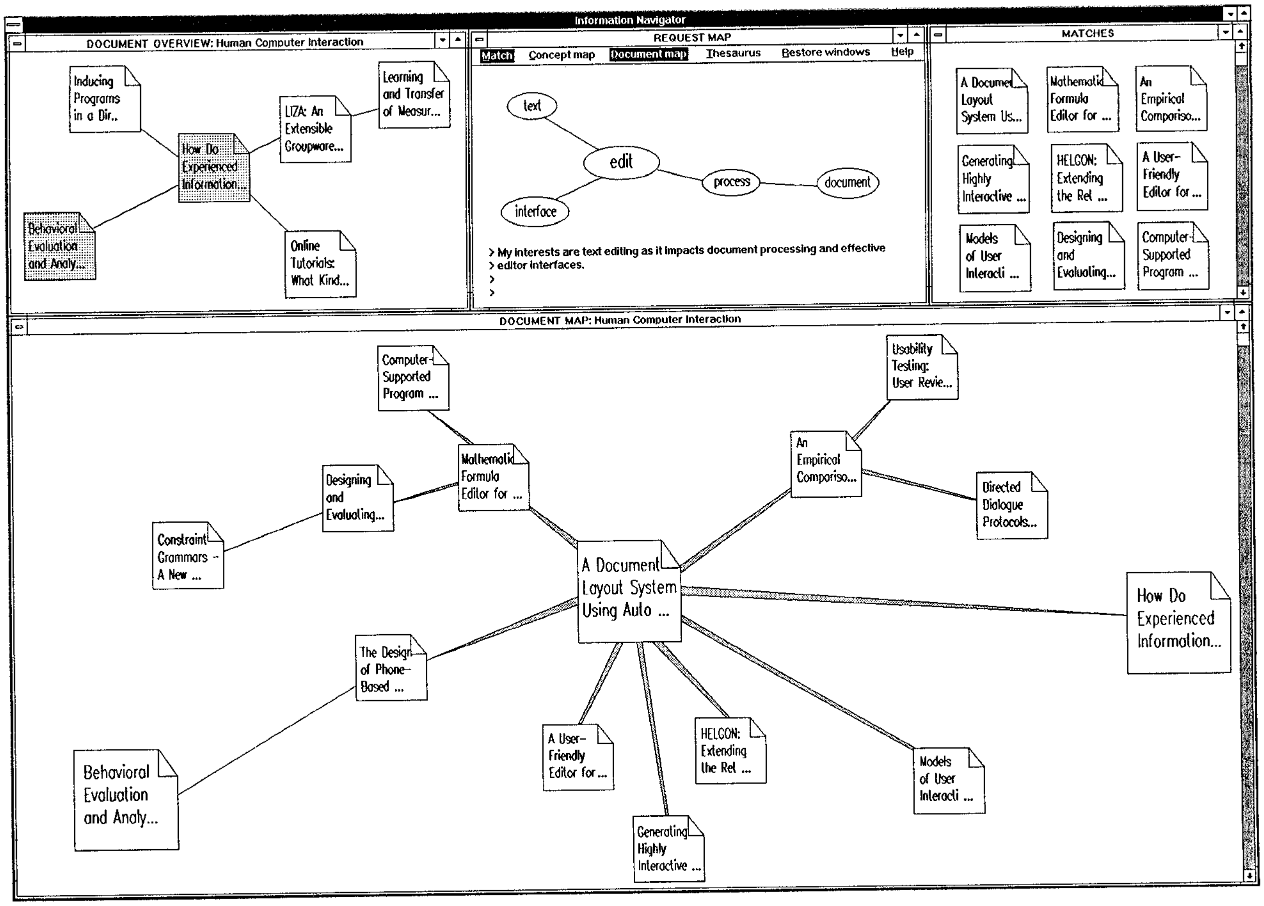

Figure 3: System with fisheye view and overview of the document collection network displayed. Nodes in the network are selectable document icons with abbreviated titles. The results of a search performed by traversing the document collection PFNET are displayed in the upper right window.

The user accesses the associative thesaurus by dragging a term node into the CONCEPT MAP window. A fisheye display of the term network is then presented, focused on that term, as shown in Figure 2. The user can navigate through the term network by selecting any node displayed as part of a fisheye view. The selected node becomes the new focus node and the display is updated. Additional term information is supplied by a conventional thesaurus. When the "Thesaurus" button is selected, a THESAURUS window replaces the CONCEPT MAP window and a hierarchical display of terms is provided. Defining an entry point and navigating within the thesaurus employs the same techniques used with the associative network of terms.

Any term displayed by the system can be used to revise the query by dragging the term node into the query window and connecting it to the query network. For example, in the display of Figure 2 the user could drag a copy of the associative thesaurus node "dialog" into the REQUEST MAP window and revise the query by then connecting the node to the query network.

Finally, the network of documents provides another path to information. For example, the search results shown in Figure 2 can provide access to the network of documents. Selecting the "Document Map" button and dragging a document icon from the MATCHES window into the DOCUMENT MAP window displays the network of documents centered on the document icon placed in the window. Also, having selected the "Document Map" button, any search the user performs will be based on the network of documents, rather than a query-document match for all documents.

One of the goals of the system is to facilitate the manual revision of queries using terms from the associative thesaurus and individual document networks. Retrieval using query expansion based on term co-occurrence alone does not necessarily lead to increased retrieval performance. To increase retrieval performance it is necessary to choose only selected co-occurring terms, such as from relevant documents or infrequently occurring terms [35]. The efficient interactive use of the visual forms in query revision is designed to facilitate selective term expansion.

The evaluations that are planned center on user behavior in utilizing the system's retrieval aids. The effectiveness of each component in manual query expansion can be assessed by allowing access to only selected parts of the system. Performance of automatic query expansion techniques will be compared to interactive query formation. A second aspect of evaluation is to assess the effectiveness of the navigation and orientation mechanisms for large networks. Canter et al. [7] examined users' movements through network-based information structures while seeking different kinds of information. The paths through the networks were characterized as graph traversals of different form. By assessing user movement through the system's network structures at different stages in information seeking, a description can be obtained of the nature of use. An assessment can then be made of the effectiveness of the system's displays in supporting the kind of traversals that users make.

7. Conclusion

Our principal goal has been to provide an environment for information retrieval integrating system components through a visual representation allowing direct manipulation. The system focuses on interaction techniques to facilitate query modification and browsing in large information structures. In deriving visual displays of network structures for the user, a number of issues concerned with graph layout, navigation, and orientation were addressed. Furnas' account of fisheye views supplied a general orientation to display and a technique for managing large information structures.

References

[1] Anick, P. G., Brennan, J. D., Flynn, R. A., Hanssen, D. R., Alvey, B., & Robbins, J. M. (1990). A direct manipulation interface for Boolean information retrieval via natural language query. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval,* 135-150.

[2] Bates, M. J. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science, 37(6),* 357-386.

[3] Batini, C., Nardelli, E., & Tamassia, R. (1986). A layout algorithm for data-flow diagrams. *IEEE Transactions on Software Engineering, 12(4),* 538-546.

[4] Belkin, N. J., & Croft, W. B. (1987). Retrieval techniques. *Annual Review of Information Science and Technology, 22,* 109-146. Amsterdam: Elsevier.

[5] Belkin, N. J., & Kwasnik, B. H. (1986). Using structural representations of anomalous states of knowledge for choosing document retrieval strategies. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval,* 11-22.

[6] Bertrand-Gastaldy, S., & Davidson, C. H. (1986). Improved design of graphic displays in thesauri - through technology and ergonomics. *Journal of Documentation, 42,* 225-251.

[7] Canter, R., Rivers, R., & Storrs, G. (1985). Characterizing user navigation through complex data structures. *Behavior and Information Technology, 4(2),* 93-102.

[8] Chang, S. K. (1989). Visual reasoning for information retrieval from very large databases. *IEEE Workshop on Visual Languages,* 1-6.

[9] Conklin, J. C. (1987). Hypertext: An introduction. *IEEE Computer, 7,* 17-41.

[10] Consens, M. P., & Mendelzon, A. O. (1989). Expressing structural hypertext queries in Graphlog. *Proceedings of Hypertext '89,* 269-292.

[11] Croft, W. B., & Thompson, R. H. (1987). I3R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science, 38,* 236-261.

[12] Crouch, D. (1986). The visual display of information in an information retrieval environment. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval,* 58-67.

[13] Crouch, D. B., Crouch, C. J., & Andreas, G. (1989). The use of cluster hierarchies in hypertext information retrieval. *Proceedings of Hypertext '89,* 225-237.

[14] Crouch, D. B., & Korfhage, R. R. (1990). The use of visual representations in information retrieval applications. In Ichikawa, T., Jungert, E., & Korfhage, R. R. (Eds.), *Visual languages and applications* (pp. 305-326). NY: Pergamon Press.

[15] Dearholt, D. W., & Schvaneveldt, R. W. (1990). Properties of pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 1-30). Norwood, NJ: Ablex.

[16] Doyle, L. (1961). Semantic roadmaps for literature searchers. *Journal of the Association for Computing Machinery, 8(4)*, 553-578.

[17] Draper, S. W. (1986). Display managers as the basis for user-machine communication. In Norman, D. A, & Draper, S. W. (Eds.). *User centered system design* (pp. 339-352). Hillsdale, NJ: Erlbaum.

[18] Eades, P., & Xuemin, L. (1989). How to draw a directed graph. *IEEE Workshop on Visual Languages*, 13-17.

[19] 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 to human-computer interaction* (pp. 201-233). Hillsdale, NJ: Erlbaum.

[20] Fidel, R. (1986). Towards expert systems for the selection of search keys. *Journal of the American Society for Information Science, 37(1)*, 37-44.

[21] Fowler, R. H., & Dearholt, D. W. (1990). Information retrieval using pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 165-178). Norwood, NJ: Ablex.

[22] Furnas, G. W. (1986). Generalized fisheye views. *Proceedings of Human Factors in Computing Systems (CHI '86)*, 16-23.

[23] Godin, R, Gecsei, J., & Pichet, C. (1989). Design of a browsing interface for information retrieval. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, 32-39.

[24] Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 75-88). Norwood, NJ: Ablex.

[25] Graf, M. (1990). Building a visual designer's environment. In S. Chang (Ed.), *Principles of visual programming systems* (pp. 291-325). Englewood Cliffs, NJ: Prentice-Hall.

[26] Hutchinson, J. W. (1989). NETSCAL: A network scaling algorithm for nonsymmetric proximity data. *Psychometrika, 54*, 25-51.

[27] Ingwerson, P., & Wormell, I. (1986). Improved subject access, browsing and scanning mechanisms in modern online ir. *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, 68-76.

[28] Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters, 31*, 7-15.

[29] Maron, M. E., & Kuhn, J. L. (1960). On relevance, probabilistic indexing, and information retrieval. *Journal of the Association for Computing Machinery, 7(3)*, 216-244.

[30] McCormick, B. H., DeFanti, T. A., & Brown, M. D. (Eds). (1987). Visualization in Scientific Computing, *Computer Graphics, 21(6).*

[31] Mettrey, J. (1987). An assessment of tools for building large KB systems. *AI Magazine, 8(4)*, 81-89.

[32] Nielson, G. M. (1989). Visualization in scientific computing. *IEEE Computer, 22(8)*, 10-11.

[33] Norman, D. A, & Draper, S. W. (Eds.). (1986). *User centered system design.* Hillsdale, NJ: Erlbaum.

[34] Palmquist, R. A., & Eisenburg, M. (1984). Testing a text analysis metric using magnitude estimation. *Proceedings of the American Society for Information Science Meeting*, 231-236.

[35] Peat, H. J., & Willett, P. (1991). The limitations of term co-occurrence data for query expansion in document retrieval systems. *Journal of the American Society for Information Science, 42(5)*, 378-383.

[36] Salton, G., & McGill, M. J. (1983). *Modern information retrieval.* New York, McGraw-Hill.

[37] Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1989). Network structures in proximity data. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 249-284). New York: Academic Press.

[38] Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *IEEE Computer, 16(8)*, 57-69.

[39] Shu, N. C. (1988). *Visual programming.* NY: Van Nostrand Reinhold.

[40] Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation, 28(1)*, 11-20.

[41] Sutherland, I. E. (1963). Sketchpad: A man-machine graphical communication system. *Proceedings of the Spring Joint Computer Conference*, 329-346.

[42] Tamassia, R., Di Battista, G, & Battini, C. (1988). Automatic graph drawing and readability of diagrams. *IEEE Transactions on Systems, Man, and Cybernetics, 18(1)*, 61-79.