Chapter 12

Information Retrieval Using Pathfinder Networks*

Richard H. Fowler and Donald W. Dearholt

Collections of written documents are available electronically for a wide range of needs and interests. Commercial vendors now provide access to large collections of scientific, as well as popular documents, and the advent of low-cost optical storage technologies promises an even wider range of accessibility. Yet, widespread use of these collections will require solutions to long-standing problems of document indexing and query formulation that allow the user to express information needs in a style that makes full use of the electronic medium.

Pathtrieve is an information retrieval system that has served as a testbed for exploring the application of Pathfinder Networks (PFNETs) in solutions to these problems. The system interface is based on direct manipulation of graphic network representations of queries, indexing terms, and document texts. PFNETs are used to provide efficient, automatically derived graph representations for natural language text that retain some of the advantages of richer, but less tractable, knowledge representations. In addition, the different associational networks in the system provide the user multiple paths of access to information items.

Document Indexing

At present, document retrieval is the only practical means for accessing knowledge in large databases of natural language text. Information needs are met indirectly by providing documents or references to documents likely to contain relevant information. The basic retrieval strategy matches a query representation to representations of documents. In essence, this is a matching of conceptual representations, and the nature of these representations is central to the performance of a system. Most document retrieval system representations are based on the assignment of indexing terms to documents.

The goal of indexing is to provide a set of terms for each document that reflects the conceptual domain and can be used effectively in retrieving the document. Unfortunately, indexing documents with high reliability and validity is in practice not possible. By several measures, remarkably low consistency both between and within indexers has long been evident (Cooper, 1969). Furnas, Landauer, Gomez, and Dumais (1983) show that human object naming behavior is marked by a lack of agreement among individuals, even in restricted domains with persons of similar experiences. This lack of agreement has important implications for the design of information systems. Bates (1986) suggests that systems should provide a wide range of document descriptors that include associations beyond conventional thesaural relations to allow alternative paths of access to documents. One means

*This work was supported by Grant IST-850676 from the National Science Foundation.

for dealing with problems inherent in labeling is provided by associational structures that reflect not only semantic relations, but other types of relations as well. This is the approach taken in the Pathtrieve system.

User Knowledge

Information retrieval (IR) systems should be capable of assisting users with varying levels of expertise. For a subject area the user knows well, a principal function of the IR system is to facilitate the mapping of the user's knowledge structures and natural language vocabulary to the indexing terms used by the system for retrieval. However, when the user does not have well-developed knowledge structures in a subject domain, retrieval is problematic. This situation seems to present a paradox: The user's information need is to know the subject area, but in order to use the system to meet the information need, the subject area must be known.

One way to resolve this paradox is provided by iterative query techniques such as relevance feedback (Robertson & Sparck Jones, 1976). The user provides feedback to the system with respect to the relevance of documents retrieved by a particular query and the system automatically adjusts the query. Another approach is to develop an expert system to aid in formulating searches. The expert system automates the role of a human search intermediary who is familiar with the indexing terms and search capabilities of a particular IR system. By using information provided by the expert system about the knowledge structures and indexing of the database, the user can be guided to appropriate search strategies (Fidel, 1986).

Other approaches to assisting relatively naive users provide information for increasing the user's knowledge of indexing terms so that a more effective search is possible. The most widely used aid is a thesaurus of index terms that gives subject information by presenting the vocabulary of terms used in classifying documents in a domain. Most thesauri specify a limited number of relations (e.g., synonym, superordinate) among terms and group-related terms together to facilitate browsing and searches in a particular subject area. Conceptual analysis, in terms appropriate for searches using a particular system and database, can be facilitated by the thesaurus.

The Pathtrieve system provides a graphic associative thesaurus to assist the user in learning about terms and relationships of terms in a domain. The relations of the thesaurus are derived from statistical measures of term similarity and provide associations beyond those found in conventional thesauri. This network, based not only on linguistic relations, but also on other types of relations, provides a means for exploration of the term vocabulary to provide cues that are complementary to the user's cognitive structure. Information requests using terms from the thesaurus are formulated through exploration of associations. The associational structure also helps overcome some of the difficulties of indexing indeterminacy by providing multiple paths of access to terms used in query formulation. A broader, less structured approach to information description is possible, with the Pathtrieve system, that is perhaps better suited to the realities of document indexing than conventional thesaural relations alone.

Interface Styles

As online systems and optical storage have become more widely available, direct end-user searching has become the norm. However, elimination of the expert human

intermediary requires an interface that facilitates conceptual interaction throughout the retrieval process. The Pathtrieve system addresses this concern by providing the user visual presentations of all conceptual structures used by the system. Work on the graphic presentation of problems suggests that a spatial, pictorial mode is useful in conveying certain types of knowledge (Billingsley, 1982; Bocker, Fischer, & Nieper, 1986), and graphic network displays are increasingly used to display knowledge structures (Fairchild, Poltrock, & Furnas, 1988; Mettrey, 1987). Additionally, viewing techniques for large data structures based on psychological considerations for display (Furnas, 1986) provide a promising approach for the very large databases often found in information systems. Though some IR systems have used graphic displays (Frei & Jauslin, 1983; Thiel & Hammwohner, 1987), this style of interface has received relatively little attention (Crouch, 1986). In particular, direct manipulation interfaces (Hutchins, Hollan, & Norman, 1986; Shneiderman, 1982) seem to provide an interaction style yet to be explored in IR systems, that is especially appropriate for the iterative, highly interactive processes of information seeking.

Pathfinder Networks in Information Retrieval

Graph-based representations have been a cornerstone of artificial intelligence (AI) research, but most cannot be used in large scale IR systems for two reasons. First, producing knowledge bases such as those used in AI applications is not done automatically. Virtually all AI knowledge bases are produced by extensive human encoding, that is, knowledge engineering. The size of the bodies of natural language text accessed by IR systems and the need to have documents made available in a timely fashion argue against deep manual encoding of documents. Second, the time complexity of procedures for manipulating these representations is not acceptable for the very large databases IR systems access. In IR systems even polynomial time procedures are to be avoided, and in most systems indexing is used extensively to allow efficient access. Though the development of tractable procedures for use with representations characteristic of AI research is a promising line of investigation, the direct transfer of procedural mechanisms from AI techniques to IR is not possible. Nonetheless, graph structure representations are increasingly found in IR systems (Belkin & Croft, 1987; Belkin, Oddy, & Brooks, 1982). Their use follows a general trend that systems designed to work on relatively small bodies of text incorporate more domain knowledge and apply AI knowledge representation techniques more directly than systems that are to be used with large bodies of text. Systems for very large databases rely at most on domain knowledge encoded as part of a term thesaurus and use shallow, statistical processing to automatically form representations of document texts.

Pathfinder provides an alternative procedure for automatically deriving network representations for IR systems. Statistically-based graph construction techniques used in IR and Pathfinder can both utilize proximity measures in deriving link structure, yet there are important differences in the resulting representations. One difference is that statistically-based graph representations used in IR have seldom been concerned with providing a psychologically salient representation, whereas the PFNET representation has been developed from the outset as a representational scheme for human conceptual structure. Only a few IR systems, such as Jones' (1986) Memory Extender system using spreading activation retrieval, have explicitly applied models of human cognition in the retrieval process. Another difference is that most statistically-based graph derivation techniques in IR systems are based on link membership determined by similarity thresholds between node pairs; links are included in a graph if the association between two nodes is above some criterion. In contrast, Pathfinder considers similarity over a wider range of nodes in deriving networks. The two approaches lead to quite different link structures.

The Pathtrieve System

The functions of the Pathtrieve system, shown in Figure 1, allow the user to access several PFNET structures during query formulation and document retrieval. The retrieval process is begun by deriving a PFNET from the natural language of a user's information request and serves as the initial query representation. This representation is displayed and can be graphically modified using terms from a second PFNET structure, the associational term thesaurus. A document collection can be automatically selected through matching of the query and the different database thesauri available to the system. One document retrieval mechanism matches the query representation to PFNETs derived from document abstracts. These document PFNETs are also available for use in graphic modification of the query. The final PFNET structure used in the system is based on the interrelationships of documents in a database. A second form of retrieval uses a selected document as an entry point into this structure and retrieves document clusters by traversing the network. The following sections describe derivation of natural language-based PFNETs, network displays and navigation tools, and document retrieval techniques.

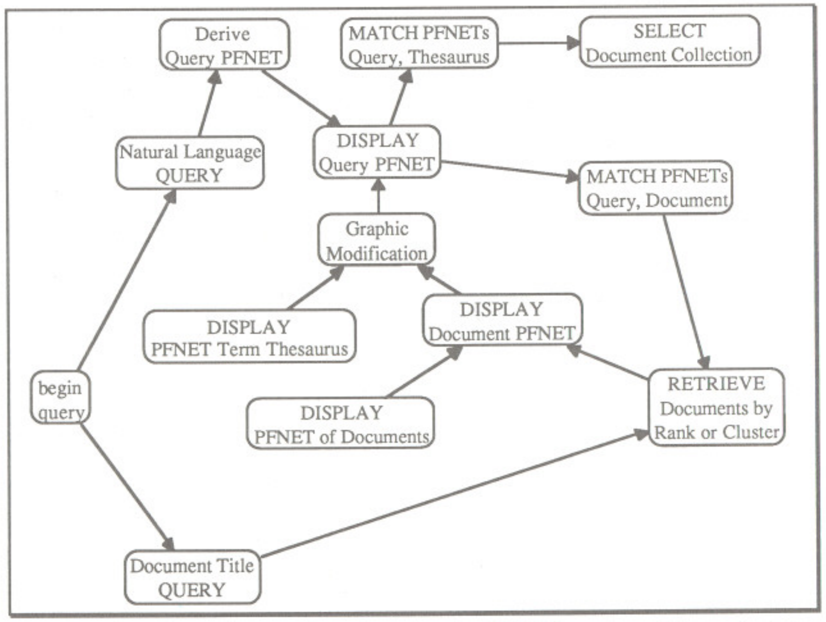

Figure 1. Overview of system functions showing PFNETs and graphic manipulations.

PFNETs from Natural Language

The term sets used in natural language-based PFNETs play a central role in the use and interpretation of the structures. The system allows access to multiple databases of document abstracts and for each database a unique set of terms is used. To construct a database's term set, a stop list is first used to eliminate function words from the texts of abstracts in a database. Stems are then derived from the remaining words and the most frequently occurring stems are used as the term set for that database. For some forms of retrieval this simple procedure suffers from the limitation that frequently occurring terms have relatively little value for discriminating among documents. However, with multiple document collections one of the functions of the term thesaurus is to provide a picture of all concepts in a document set. The most frequently occurring terms tend to be general terms that provide valuable information about the domain of the document collection and relations among concepts.

The statistical text analyses performed during PFNET derivations rely on recovering conceptual information from natural language by considering the form of text. This shallow analysis has advantages in that it can be performed sufficiently quickly for use in interactive query formulation, and term similarity measures reflect relatedness from a wide range of sources. The proximities used in constructing the different natural language PFNETs are derived from term co-occurrence in different textual units. The system uses PFNET representations of the user query, single document abstracts, and the term thesauri. Calculation of proximities for the term thesauri PFNETs uses all abstracts in a database and so provides similarity measures for all term pairs. In analyzing each body of text, stop list words are removed, the remaining words are stemmed, and stems are eliminated that are not in the database's term set. A weighted co-occurrence measure, similar to that used by Belkin and Kwasnik (1986), is computed using the remaining terms. For term pairs, similarity is increased by 5 for terms adjacent in a sentence, 3 for nonadjacent terms occurring in the same sentence, and 1 for terms occurring in the same document. PFNETs use $r = \infty$ and $q = n-1$, where n is the number of terms.

Query PFNET. A query is usually initiated by the user's entry of a natural language request for information. The request is transformed to a PFNET and the Pathfinder Graph (PFG) is displayed in the Query window, as shown in Figure 2. Two means of modifying the query can be used: further input of natural language text or graphic manipulation using other PFGs displayed by the system. When additional text is entered, it is appended to the existing text and reanalyzed to produce a new query PFG. The query graph can also be modified by graphically moving any term node displayed by the system (i.e., in document PFGs and the thesaurus PFG) into the Query window and connecting it to the query graph. Query nodes can also be deleted, and the link structure changed. The query representation can be made more elaborate and revised to better reflect information needs using knowledge gained through evaluation of search results and interaction with the thesaurus.

Document Abstract PFNETs. The representations of documents used in retrieval and graphic display are constructed from the texts of document abstracts. A typical Pathtrieve display is shown in Figure 3 with a document PFG as the central network display and the document abstract text in the lower right window. In all graphic displays, word forms of term stems are presented. In this example the 48-word abstract contains 14 unique term stems that provide the document PFG nodes. The upper windows display an abstracted overview of the term thesaurus, the natural language query PFG, and an ordered list of documents retrieved in a search based on the query PFNET. Any document title in view can be selected to display the abstract PFG or text.

In addition to providing information about term usage in a document collection, the document PFG also allows the user to rapidly preview a document. A document PFG can be scanned for terms or constellations of terms of interest. This ease of interacting with a document collection encourages exploration. In meeting information needs in subject areas not well-known to the user, exploration or browsing can be useful in directly retrieving documents, as well as in becoming acquainted with the domain.

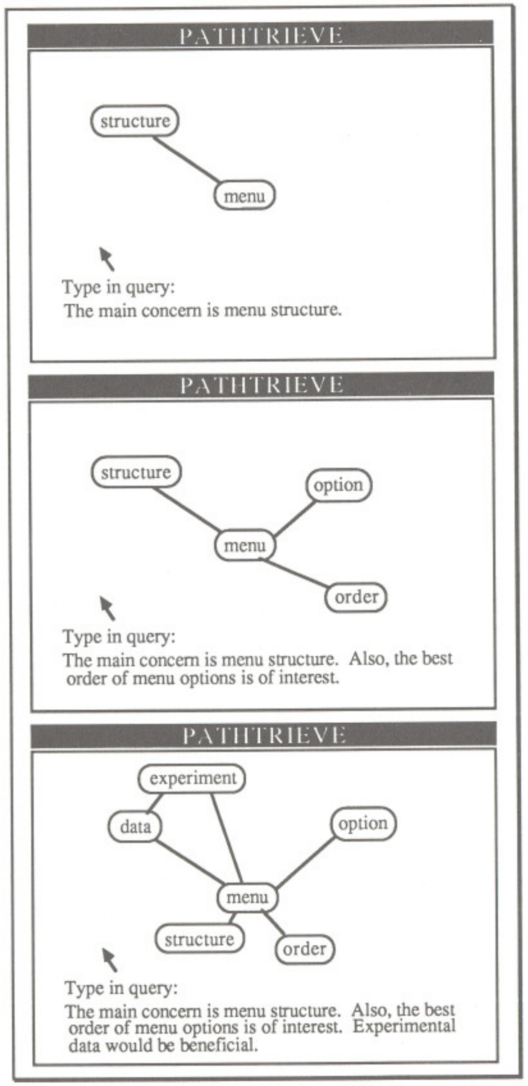

Figure 2. Query window displays of natural-language queries and PFGs.

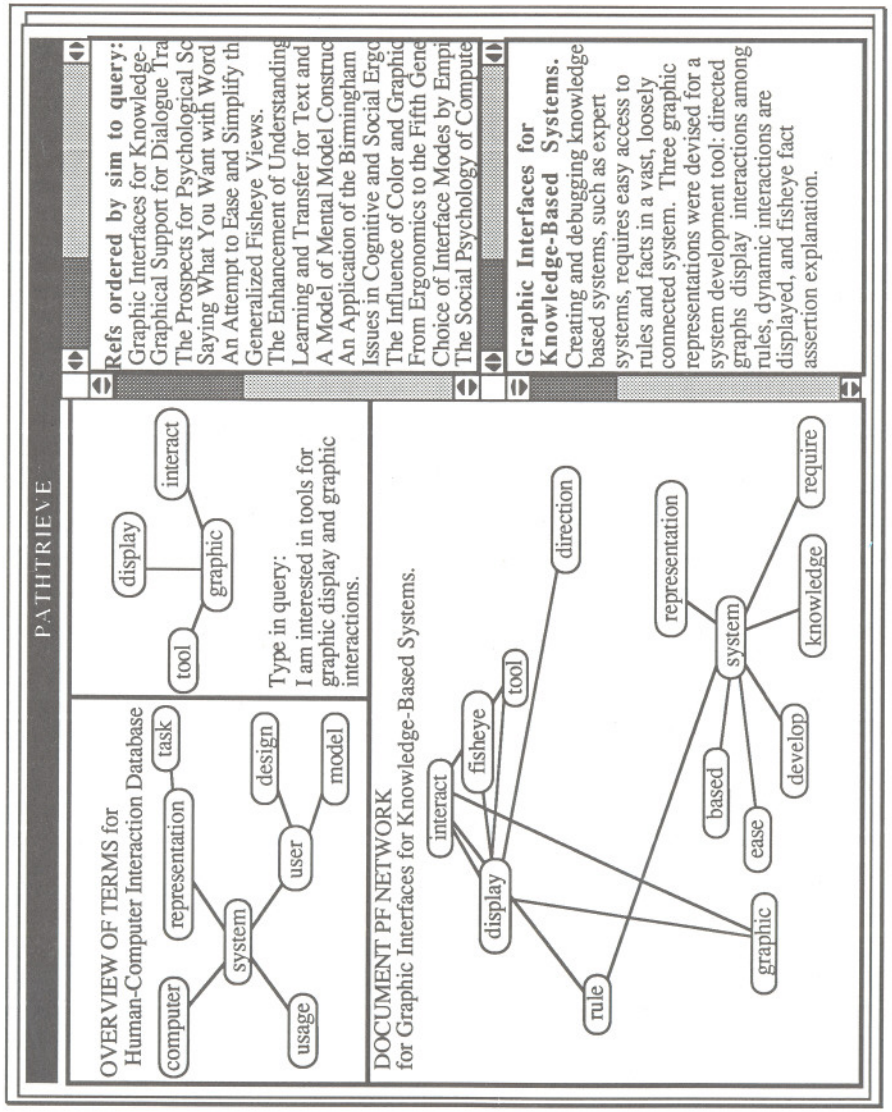

Figure 3. A typical Pathtrieve display. The upper windows display (1) an abstracted view of the term thesaurus, (2) the text of a natural language query with the PFG derived from it, and (3) an ordered list of documents retrieved in a search. The lower windows show (4) a document abstract PFG, and (5) the document abstract text.

Term Thesaurus PFNET. The term thesaurus provides information about indexing terms in a conceptual domain and should be flexible since users will think in different terms and perceive different relations as important. A thesaural organization that permits flexible navigation will be useful for both expert and novice users since the type of relation that makes a term relevant depends on the specific information need. Even in subject areas well-known to the user, an associational thesaurus is likely to provide cues that facilitate exploration. Both the need for an IR system to deal with different types of information needs and indexing indeterminacy suggest that mechanisms be provided to allow multiple paths of access to information objects. In the Pathtrieve system a PFNET constructed from all abstract texts in a database, using shallow text analysis to reflect a wide range of relations, provides an associational network thesaurus as one mechanism to address these concerns.

The PFG displays and navigation tools are designed to exploit the potential of user-controlled visual presentations of information. A graphic thesaurus display has several advantages as compared to textual display of the same material (Bertrand-Gastaldy & Davidson, 1986; Craven, 1984; Doyle, 1961). Network displays can group related terms more compactly and appropriately than the lists used in printed thesauri. Closely tied to this feature of graphic display is the ability to easily and quickly move or navigate through the terms by following thesaural relations. The associational thesaurus is one mechanism for conveying domain knowledge and information about term usage. The user is provided with more detailed term and domain knowledge through viewing the natural language of abstracts together with the term PFG for an abstract. The incorporation of a conventional thesaurus based on linguistic relations would provide an additional means for conveying knowledge about terms, thereby complementing the information conveyed by the associational PFNET thesaurus.

Network Displays and Navigation Tools

The development of procedures for spatial organization in graphic displays is a pervasive problem. Fortunately, some scaling techniques suggest a spatial form, for example, a tree for hierarchical clustering results. In knowledge engineering applications, display is often constrained by properties of the representation, such as temporal ordering for production systems or common slots in a schema-based representation. For PFNET displays, some constraints are needed to determine node layout.

In the Pathtrieve system PFGs are spatially organized using nonmetric multidimensional scaling (MDS). This solution to weighted network display has been applied to a variety of graph structures (Klovdahl & Brindle, 1987) and may provide additional information through the positioning of nodes. In determining the maximum graph size that can effectively be displayed using this technique, line (link) crossings are the limiting factor. As the MDS solution seeks to minimize a global criterion in deriving the spatial coordinates, and Pathfinder tends to emphasize local relationships in the link structure, limitations are readily apparent. However, with small (up to about 15 nodes), relatively sparse structures, MDS node layout is usually satisfactory. This includes most query and document PFNETs and determines a practical limit of the subgraph size that can be displayed effectively with this technique.

In accessing the PFNET thesaurus network, many more terms are available than can be presented on a screen, so only subgraphs of the complete thesaurus can be displayed in the viewing area. Two kinds of tools are available for subgraph viewing: an abstracted aerial view of the complete network providing orientation to the overall structure, and navigational functions allowing selection of subgraphs from the complete thesaurus for detailed viewing.

Aerial View. While a number of abstraction techniques have been explored for PFNETs (McDonald & Schvaneveldt, 1988), the abstracted, aerial view of the thesaurus PFNET used in the Pathtrieve system is simple yet useful in thesaurus navigation and orientation. The aerial view includes the nodes with highest degree in the term thesaurus PFNET. Links are displayed for nodes that are separated by paths not exceeding a criterion based on the size of the thesaurus. This sort of "bifocal display" has proven useful in providing orientation within large data structures when only parts of the complete structure can be displayed (Englebart & English, 1968; Ramsey & Grimes, 1983).

Subgraph Selection. The user can define a thesaurus entry point and initiate display of a thesaurus subgraph by selecting a focus node from one of several displays: aerial view, query PFG, document PFG, or the current thesaurus subgraph. A thesaurus subgraph is formed by following links from the focus node. If the subgraph size, defined by a default number of links (usually 3), exceeds a criterion set to reflect the number of nodes that can be satisfactorily displayed, shorter traversals are tried until the subgraph size meets the criterion. In the subgraph displayed in the system's network window, the focus node is inverted, and the size of nodes decreases with the number of links away from the focus node in order to convey a visual perspective centered on the focus node.

Orientation and Conceptual Elaboration within the Thesaurus. The different sources for defining an entry point into the thesaurus PFNET not only facilitate traversal of the thesaurus PFNET, but also provide different contexts and paths of access for a term. A goal of the system is to provide the user at least one structure that includes terms useful in formulating the information request. The PFG derived from the users natural language query supplies information about terms by showing, as the nodes of the PFG, which words of the query are indexing terms for a particular database. In addition, documents retrieved after a search can be inspected individually and provide another context and means of access to index terms. For a document, both PFG and abstract text can provide contexts for a term. Often other terms of interest will be found in documents retrieved in a search. Both of these structures can provide an initial context for terms before more information about a term's relation is learned through navigation within the thesaurus.

The abstracted, aerial view of the complete network is the principal means of orientation to the complete thesaurus. By incorporating the most frequently occurring terms, the most general terms tend to be selected. These terms are the most likely to be familiar to the user not acquainted with a domain and can provide a starting point for elaborating conceptual structures already available. For example, Figure 4 shows the system display produced by selecting the node labeled *user* in the Aerial View window. By viewing the thesaurus subgraph displayed in the central window, the user learns more about the range and relations of terms. Here, the user finds both general terms, such as *knowledge* and *software*, and less common terms, such as *UNIX* and *dialog*. Additionally, the information that all of these terms are associated with "user" is conveyed.

Selecting nodes from the aerial view to initiate subgraph display allows the user to quickly move around the thesaurus network and can provide an overview and orientation to the indexing terms used in a database. When changing the focus node within a subgraph display, the aerial view also provides orientation by tracking aerial nodes that come into view as part of a subgraph. Any aerial view node that is displayed as part of a thesaurus subgraph is inverted in the Aerial View window to show the position of the subgraph in the complete thesaurus.

In practice, thesaurus subgraphs usually contain an aerial view node due to the form that thesaurus PFNETs tend to have. Using term sets of the most frequently occurring terms, common, high frequency terms tend to co-occur with a relatively large number of

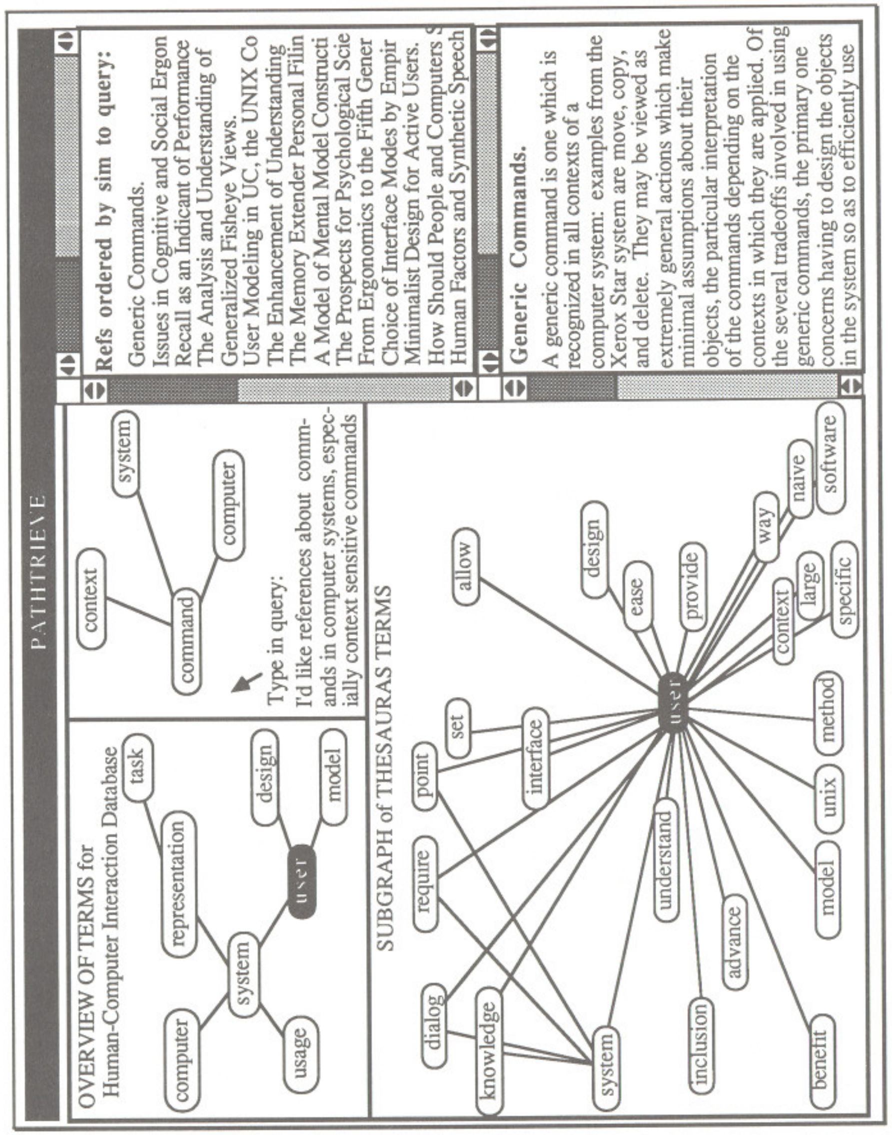

Figure 4. Pathtrieve display after selecting *user* node in Aerial View window. The thesaurus subgraph focused on *user* is displayed in the central Network window. The aerial view orients the subgraph to the complete thesaurus by tracking (inverting) aerial view nodes present in the subgraph.

infrequently occurring terms. Infrequently occurring terms co-occur less often with other low frequency terms than with the common terms. The resulting thesaurus PFNET has relatively few high-degree nodes (the frequently occurring terms) and many low-degree nodes (the infrequently occurring terms). These high-degree nodes are the nodes that are selected for the aerial view. In terms of path length the aerial view nodes are relatively close together. When traversing the complete thesaurus, an aerial view node is usually within the path length used for subgraph definition and so is tracked in the aerial view. This form of the thesaurus PFNET also allows traversal with relatively few changes of subgraph focus node.

Document Retrieval Techniques

Several different retrieval mechanisms might be usefully employed in an IR system to meet different types of information needs (Belkin & Croft, 1987; Jones & Furnas, 1987). Exact match Boolean searching, in which a query is formed by the combination of terms using the operators AND, OR, and NOT, is the retrieval technique most widely used in commercially available IR systems. Using this technique, query-document matching produces simply a set of retrieved documents. No use is made of relationships between query terms or relationships between individual documents, and the resulting unordered groups of documents are difficult for the user to evaluate. Because the match of query and document representation must be exact, the indexing vocabulary problem is particularly significant. Still, this approach is useful for information needs that can be articulated by the searcher in terms appropriate for the system. However, a precise specification of indexing terms is difficult or impossible when the subject area is not well-known to the user. Partial match techniques, such as term vector matching in which documents and queries are represented as vectors of terms, allow a range of match between query and document representations. The varying degree of match can be used to rank documents for presentation to the user with respect to how well each corresponds to the query.

IR systems use a wide range of techniques for representing the conceptual structures of queries and documents, from simple unweighted keyword lists to representations with unlimited expressive power, such as semantic networks. Our approach represents an intermediate position: PFNETs allow nodes to represent conceptual entities, but do not provide conceptually labeled relations as necessary for the expressive power of semantic networks. The choice of which retrieval techniques to include in a system depends on how all system components interact to meet those needs. PFNET-based components seem particularly useful when meeting information needs in subject areas not well-known to the user.

PFNET Similarity Measures. Providing similarity metrics for the PFNETs derived from documents and queries is an area rich with possibilities. We anticipate that Pathfinder network representations, which were designed with explicit consideration of psychological models of human memory, will lead to graph matching techniques useful in a wide range of document retrieval tasks. In comparing query and document PFNET representations, simply matching the node labels (stems) and making no use of the links is one form of query vector matching. A matching algorithm that takes into account both conceptual similarity, as might be measured by identifying common node labels, and structural similarity (Knoke & Kuklinski, 1982) should be more effective because link structure can be used to provide information about the relationship of terms.

The algorithm now employed derives metrics for two components of similarity. The first component is computed as the number of common term stems in the query and the document divided by the number of terms in the query. A second component provides a measure of structural similarity by considering the link structures of both query and document. The value of this component is increased when nodes closely connected in the query

PFNET are also closely connected in the document PFNET. For nodes common to both query and document, the measure of similarity of structure is increased by 2, if direct links exist in both query and document, between common node pairs; or by 1, if node pairs common to query and document are not directly linked in both, but have paths of 2 or less in both. This sum is divided by 2 times the number of links in the query so that the value is between 0 and 1, as is the measure of conceptual similarity. Conceptual and structural similarity measures are then weighted and summed to give the final similarity measure. The weighting of the different components provides one means for adjusting the algorithm's retrieval performance.

One major tractability concern in the system is the time required to apply a matching algorithm to document and query representations. Many forms of graph matching in the general case require a number of computations exponential in the number of nodes or links considered. Except for very small structures, time complexity of this order is unacceptable for the matching component of an IR system. Yet, in designing a matching algorithm for a specific application or problem domain, it is usually possible to exploit constraints so that a more tractable, but less general, algorithm can be found (Galil, 1986). Another consideration is that in IR the expected case complexity is often more important than worst case performance. Here, too, the nature of the domain plays a critical role in determining whether a particular algorithm is feasible for use.

Fortunately, well-known IR techniques can be applied to accomplish most aspects of the graph matching process efficiently. For example, in the matching scheme outlined above, the measure of conceptual similarity can be obtained using efficient vector matching techniques such as those provided by inverted file organizations (Salton & McGill, 1983). Determining similarity of link structures seems a more difficult task, but constraints of the problem common to many IR applications allow a tractable solution. In comparing query and document link structures in the above scheme, it is, of course, only necessary to consider documents with at least one query term. For the remaining set of documents only node pairs present in both query and document need be considered, thus typically eliminating most documents. In practice, PFNETs with $r = \infty$ and $q = n{-}1$, such as we have used tend to be quite sparse, usually the number of links is less than twice the number of nodes. Since the document PFNET link structure is static, it can be indexed to allow efficient access. The form of the indexing can be selected to be appropriate for the particular matching algorithm to be used.

For the user of an information system, the first step is often the difficult task of selecting a database likely to contain relevant information. The Pathtrieve system automates database selection using the natural language query. For each document collection database, a query PFNET is derived using the database's term set and the query network is compared to the thesaurus PFNET for the term set. The matching algorithm, slightly modified to provide a measure not normalized by the number of query concepts, provides a measure of similarity of the query network to each database term thesaurus PFNET, and the database with the highest similarity value is selected for use.

Document Collection PFNET: Browsing and Cluster Retrieval. Browsing among shelves in a library is exploration and search among documents guided by the classification system used in the library. Within an information system more search and organizational flexibility is possible. Access to documents can be based on a network of documents derived from interdocument relations. From some entry point the network can be traversed and documents selected. This technique has the advantage for some types of information needs of requiring little query formulation and knowledge of the subject area.

Finally, there is a PFNET reflecting relations among documents. In this PFNET, nodes are the documents in a database and the graphic labels are document titles. The proximities for deriving this network are obtained by applying the matching algorithm to all pairs of document abstract PFNETs in a database. The document collection PFNET is used for exploration in the network of documents similar to exploration in the term thesaurus. The PFG of documents can be displayed, as shown in Figure 5, at any time. As with the term thesaurus, an aerial view is provided as well as navigation tools for browsing and entry point selection. Selecting a title allows the user to view the document's abstract text and PFG. Functions using the PFG of documents provide an additional means of gaining domain knowledge and access to graphic network structures.

In addition to ordering documents by similarity to a query, a second form of retrieval is available in the system. Cluster-based retrieval uses the PFNET of documents and is based on traversing the network beginning at a particular document. The entry point can be directly provided by specifying a title, or by finding a document that best matches the query. Additional documents are then retrieved by following the links from the starting point. The sequence of retrieved documents is displayed to the user ordered by the number of links from the entry point document.

Summary

The Pathtrieve system has served as a testbed for the application of PFNETs in information retrieval and has allowed us to explore a number of uses of PFNETs. PFNETs have been used within well-understood IR techniques, such as retrieval based on document networks, as well as somewhat novel applications, such as direct manipulation of query network structures. The network representations are relatively shallow and, as such, contrast with current trends in IR for the use of deep representations more characteristic of AI techniques than traditional IR techniques. However, the associational representations may be appropriate to meet some of the more challenging user information needs by allowing multiple paths of access to information items. While our initial goal was to examine PFNET-based mechanisms solely as an adjunct to effective retrieval techniques in subject areas well-known to the user, our investigations suggest that graph matching retrieval techniques using PFNETs can provide a useful retrieval mechanism for a wider range of user information needs (Fowler & Dearholt, 1989).

The primary emphasis in development has been on mechanisms that allow the user to express an information need through the elicitation, elaboration, and revision of conceptual structures. In deriving the query PFNET from the user's natural language statement of the information need, an expression of the need using the system vocabulary terms and in the form used by the system for retrieval is automatically constructed. The interface provided for query revision and thesaurus use is designed to facilitate query elaboration and revision by providing mechanisms for graphic manipulation in order to reduce the perceived distance between user and conceptual structures of the system. The use of a uniform graphic PFNET representation is designed to reduce system complexity. These mechanisms are provided to exploit the potential of a visual, graphic interface for the display of knowledge during the highly interactive information retrieval process. The graphic PFNET structures provide representations that complement the user's conceptual structures. In many cases the Pathtrieve system guides the user in obtaining domain knowledge through exploration, rather than directly providing it. The graphic interface based on conceptual structure facilitates interaction for users at any level of expertise in a subject area and might extend information system use to a larger population.

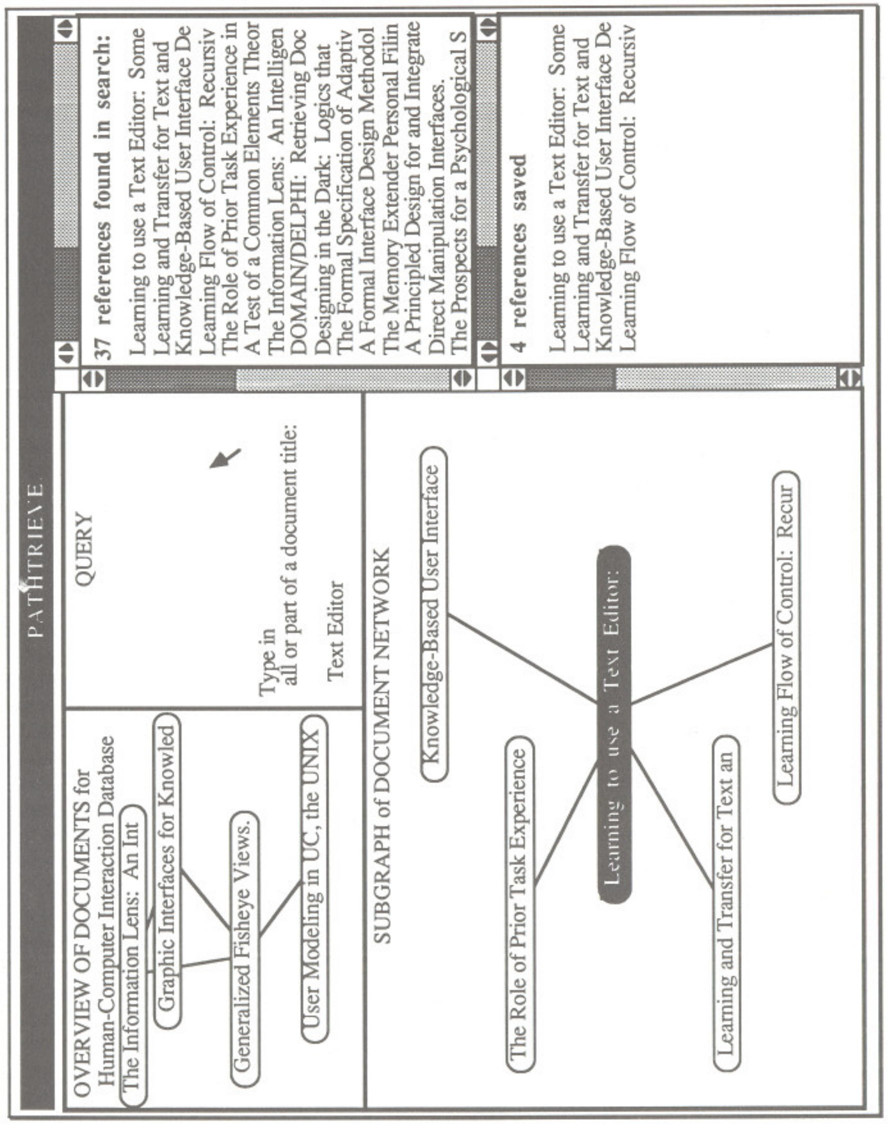

Figure 5. Pathtrieve display with PFG of documents subgraph in Network window and abstracted representation of the document PFNET in the Aerial View window. Node labels are document titles. The results of a cluster-based search are displayed in the upper-right window.