Report Number: WUCS-89-49

1989-11-01

Information Retrieval from Hypertext: Update on the Dynamic Medical Handbook Project

Mark E. Frisse and Steve B. Cousins Washington University in St. Louis

This paper attempts to provide a perspective from which to develop a more complete theory of information retrieval from hypertext documents. Viewing hypertexts as large information spaces, we compare two general classes of navigation methods, classes we call local and global. We argue that global methods necessitate some form of "index space" conceptually separate from the hypertext "document space". We note that the architectures of both spaces effect the ease with which one can apply various information retrieval algorithms. We identify a number of different index space and document space architectures and we discuss some of the associated trade-offs... **Read complete abstract on page 2.**

Information Retrieval from Hypertext: Update on the Dynamic Medical Handbook Project

Mark E. Frisse and Steve B. Cousins Washington University in St. Louis

Complete Abstract:

This paper attempts to provide a perspective from which to develop a more complete theory of information retrieval from hypertext documents. Viewing hypertexts as large information spaces, we compare two general classes of navigation methods, classes we call local and global. We argue that global methods necessitate some form of "index space" conceptually separate from the hypertext "document space". We note that the architectures of both spaces effect the ease with which one can apply various information retrieval algorithms. We identify a number of different index space and document space architectures and we discuss some of the associated trade-offs between hypertext functionality and computational complexity. We show how some index space architectures can be exploited for enhanced information retrieval, query refinement, and automated reasoning. Through analysis of a number of prototype systems, we discuss current limitations and future potentials for various hypertext information retrieval structures.

INFORMATION RETRIEVAL FROM HYPERTEXT: UPDATE ON THE DYNAMIC MEDICAL HANDBOOK PROJECT

Mark E. Frisse
Steve B. Cousins

WUCS-89-49

November 1989

Department of Computer Science
Washington University
Campus Box 1045
One Brookings Drive
Saint Louis, MO 63130-4899

Presented at Hypertext-89, November, 1989, Pittsburgh, Pennsylvania.

Information Retrieval From Hypertext: Update on the Dynamic Medical Handbook Project

Mark E. Frisse and Steve B. Cousins

Medical Informatics Group
Department of Internal Medicine,
Institute for Biomedical Computing
and Department of Computer Science
Washington University School of Medicine
660 S. Euclid Avenue (Box 8121)
St. Louis, Missouri 63130
frisse@wucs1.wustl.edu

ABSTRACT

This paper attempts to provide a perspective from which to develop a more complete theory of information retrieval from hypertext documents. Viewing hypertexts as large information spaces, we compare two general classes of navigation methods, classes we call local and global. We argue that global methods necessitate some form of "index space" conceptually separate from the hypertext "document space". We note that the architectures of both spaces effect the ease with which one can apply various information retrieval algorithms. We identify a number of different index space and document space architectures and we discuss some of the associated trade-offs between hypertext functionality and computational complexity. We show how some index space architectures can be exploited for enhanced information retrieval, query refinement, and automated reasoning. Through analysis of a number of prototype systems, we discuss current limitations and future potentials for various hypertext information retrieval structures.

"Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing."

—Vannevar Bush, As We May Think, 1945

LARGE INFORMATION SPACES

Hypermedia may be most useful as a means for exploring large collections of information recorded in electronic form. Vannevar Bush, for example, saw the Memex as a tool which could spawn a "new profession of trail blazers, ... who find delight in the task of establishing useful trails through the enormous mass of the common record." Nelson proposes a global "public access system to be franchised like hamburger stands" [Nels87]. Trigg's thesis concerned a "network based approach to text handling for the online scientific community" [Trigg83] and Schatz's Telesophy system was envisioned as a "system for the masses" united through a "transparent WorldNet" [Schat85].

We believe that these metaphors represent a particularly desirable vision for the future of information management in medicine. Workers in pursuit of recorded biomedical knowledge must confront a literature characterized by millions of scientific articles from over

3000 worldwide scientific publications [Warr81]. Adding to the confusion, many articles are never cited, some report flawed experiments, and some make exaggerated or deceitful claims. The rate of accumulation of scientific information is expected to increase dramatically as the biomedical community embraces large-scale research endeavors like the elucidation of the structure and function of the human genome [Bilof88, Came88, Sidm88]. Without advanced information-retrieval aids, we believe that the biomedical information-seeker will resemble an inhabitant of Borges' Library of Babylon - wandering through the madness of an infinite library whose books contain every conceivable combination of letters and words.

The Dynamic Medical Handbook Project has as one of its principal goals the discovery of more effective methods for information retrieval from large-scale biomedical hypertexts. Using text from a popular medical handbook, we wish to determine the degree to which these novel software systems can facilitate the effective use of recorded medical information [Frisse88a].

LOCAL AND GLOBAL ACCESS METHODS

Many people find hypertext most useful when they examine in a sequential manner a number of separate units of text. We call these units *information units* (IUs), and we use the term *links* to identify the structures connecting IUs. We use the term *document space* to refer to that portion of the information space consisting of the information units and links, and we use the term *navigation* to refer to the process of moving along links from one IU to another. Examining one IU, the user sees within the text "buttons" with labels which represent either the semantics of a connecting link or a summary description of the IU residing at the link's distal end. We use the term *link labels* to denote information about distal information units displayed on a proximal information unit. Links can connect information units which may be far apart when viewed from the perspective of some underlying "native" document structure and the process of traversing the link allows one to traverse great "distances" in a single step. We use the term *button-based browsing* to denote the entire process of examining an IU, activating a link label icon, and travelling to a new IU (Figure 1). Button-based browsing is a *local* method of hypertext navigation.

Indexes are data structures that facilitate *global* navigation. Because indexes are in one sense a set of "pre-compiled" links, they facilitate access to a needed information unit without the need for traversal through many intermediate information units, and the time necessary to find indexed information does not increase significantly as the size of the document space increases (Figure 2).

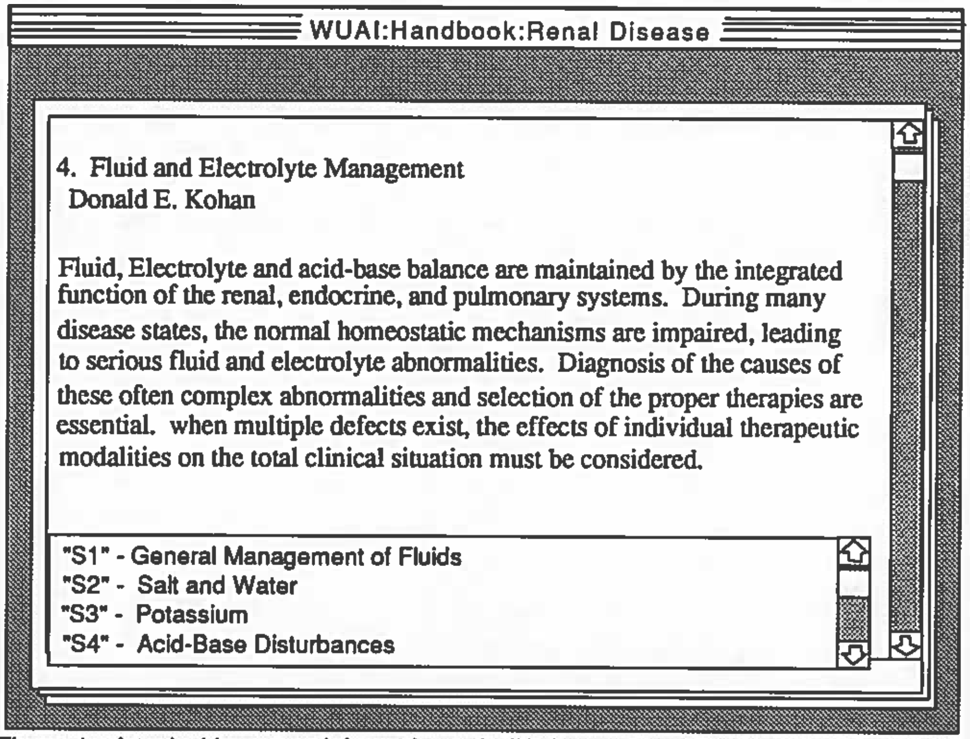


Figure 1: A typical hypertext information unit (IU) from a HyperCard medical handbook prototype. The four link labels at the bottom of the information unit represent IUs at the next immediate level of the hypertext's underlying hierarchical organization.

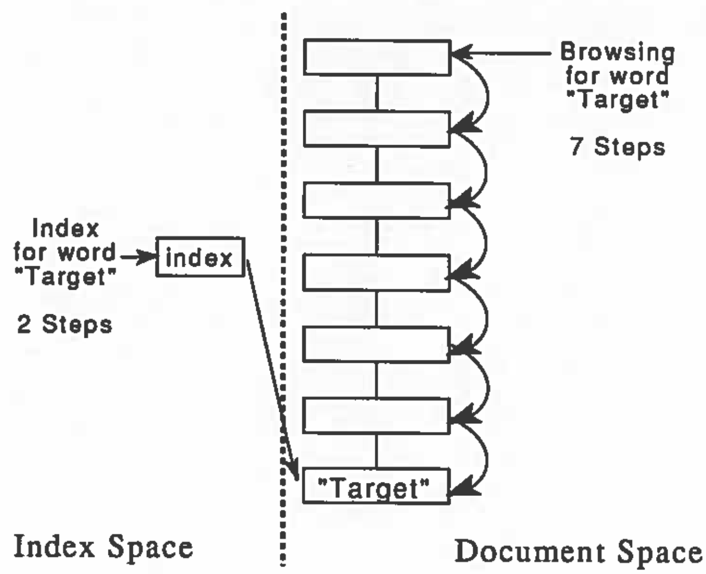


Figure 2: Local and Global methods. If button-browsing through a list, a reader at one IU can only test for the presence of a token on anotherIU by traversing the entire list because the token in question might reside only on the "last" IU examined. An index can generally answer the same question by means of a pointer which takes the user immediately to the desired IU.

All global methods for information retrieval appear to require the existence of a data structure "outside" the document space. We call this outside structure the *index space*, and we use the term *information space* to denote the union of the index space with the document space. In applications like HyperCard and the NeXT Digital Librarian, large and complex global index spaces are hidden from the user, giving the impression that global operations are being performed solely through manipulation of the document space. The presence of a data structure outside of the document space also allows one to retain some memory of user preferences and behavior. This memory can facilitate *learning* when the iterative models for information retrieval are adopted [Book83]. There remains much controversy over the efficacy of various methods for index-based information retrieval [Blair85a, Salt88].

Figure 3: A flat index space combined with a flat document space. This is the configuration commonly observed with an inverted index and a full-text document retrieval system. A strong sense of conditional independence is present both within the document space and within the index space. The probability that a specific document or term will appear at a specific location is independent of the location of other documents or terms.

INFORMATION SPACE CONFIGURATIONS

Configurations for both the index space and the document space can best be understood through an example. Consider first a traditional full-text document database system accessed by a conventional inverted index. The document sequence within the database generally is random and index space terms are usually arranged in lexicographic order. The lack of ordering within the document space gives testimony to the lack of explicitly stored relationships between documents. The arrangment of terms within the index space gives credit to their lexicographic ordering more than to their meaning. (Certainly the fact that the index term "ant" might immediately precede the index term "antelope" provides little insight into the order of nature!) The independence relationships observed both between documents and between index terms suggest that in traditional full-text document retrieval systems both the document space and the index space are "flat" (Figure 3). Location and meaning are not correlated. From the perspective of probability theory, knowledge about the utility of one document or index term does not provide one with information about the utility of any adjacent document or index term. This property, known as *conditional independence* explains in part the tractability of algorithms applied to full-text document retrieval systems. If two documents or index terms are independent, one can perform mathematical operations on either one without concern for the state of the other, and one generally can combine terms in a linear fashion.

The document space of the medical handbook used in our work is not flat. Our handbook, the Washington University *Manual of Medical Therapeutics*, is a hierarchically-organized document accessed in part by a traditional index [Frisse88a]. The information space of the *Manual* is characterized by a hierarchical document space and a flat index space (Figure 4).

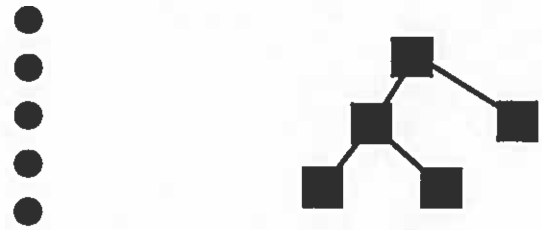


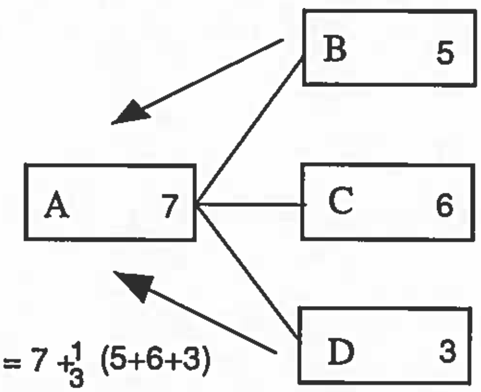
Figure 4: A flat index space combined with a hierarchical document space. This configuration exists when an inverted index is used to access a hierarchical document space

Imposing on the *Manual* the flexibility of a personalized hypertext creates a dramatic change in the structure of the document space. Incorporating cross-links into a directed hierarchy adds to the document space the potential for graph cycles, multiple parents, and multiple semantic relationships. Although some of these attributes expand capabilities for local browsing, they often do so at the expense of complexity necessary to support global access methods. By our terminology, document spaces that allow information units to have multiple parents are called " *networked spaces*".

Figure 5: A hierarchical index space combined with a networked document space. The terms in the index space are organized as a complete hierarchy. The information units in the document space are unconstrained and can be connected with multiple links.

DECOMPOSITION INTO SPACES

In our research, we have chosen to study how one can develop aids to hypertext navigation by exploiting the configurations of the index and document space of the hypertext. We believe that different conformations of index and document spaces allow for different reasoning mechanisms. To a large degree, the more complex the organization of a component of the information space, the greater the computational complexity of algorithms used to reason over the space. In this report, we summarize briefly our efforts to exploit the hierarchical document space of the *Manual*, and we discuss in detail our recent efforts to apply Bayesian reasoning techniques to hierarchical and networked index spaces in an effort to construct hypertext navigation aids that improve their performance with use (Figure 5).



$$Wt_A = 7 + \frac{1}{3}(5+6+3)$$

Figure 6: Propagating Information Unit Weights Upward. Propagating weights upward biases towards information units higher in the document space. This strategy is applied to the document space to direct attention toward parent nodes of cards with high weight. This approach may be useful to identify the "best" root for a graphical browser. (See [Frisse88a] or [Cons89] for additional details)

Operations on the Document Space

When complex documents are decomposed into smaller hypertext information units, valuable contextual information can be lost. If, in a hierarchical hypertext, an information unit about "complications" is a "child" of an information unit about a "disease", one should be able to conclude that the child information unit is really about a complication of a specific disease and not about complications in general. In our previously published experiments involving manipulations of hierarchical document spaces, we proposed a method to incorporate context at the time of query by passing a fraction of an information unit's value "upward" to its parent [Frisse88a]. Parent information units would therefore be credited for their possession of valuable descendant information units (Figure 7). Consens and Mendelzon summarize this approach from the perspective of the formal query language GraphLog [Cons89].

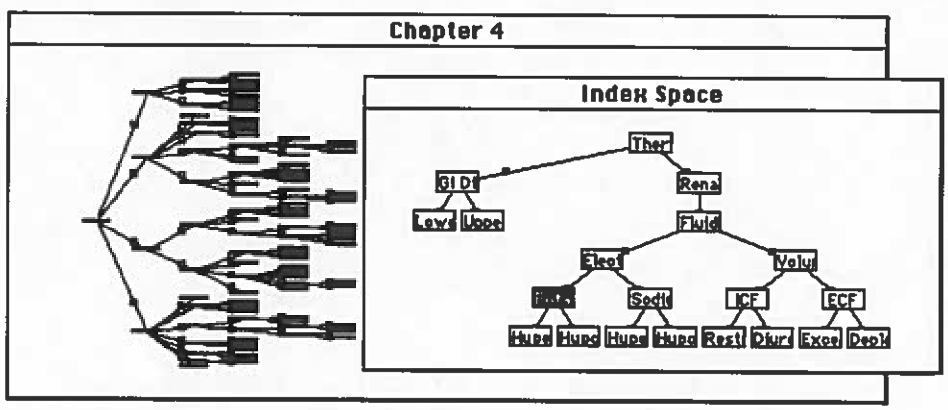


Figure 7: A portion of the index space and document space used in our experiments. The left hierarchy represents a handbook chapter. Information units (Figure 10) are accessed by activating individual nodes in the document space graph. The right hierarchy is a portion of an index space. Activating individual index space nodes displayes node belief values and probabilistic relationships between nodes.

OPERATIONS ON THE INDEX SPACES

Index spaces represent a rich source of organized information about a large information space. In our current research, we are investigating how one can represent index spaces as belief networks. If index spaces are represented this way, it is reasonable to determine whether or not reader feedback can be used to manipulate the networks in a manner which will lead to a distribution belief values for topics of interest to the hypertext user.

Inference Using Belief Networks

Using a hierarchical document space and a hierarchical index space, we are exploring methods to facilitate search through a structured medical handbook (see Figure 9). In the spirit of the RUBRIC system, we arrange index terms in a hierarchical manner [McCu85]. Unlike RUBRIC, our index terms are joined by probabilistic dependencies expressed and equilibrated by the method of Pearl [Pearl88]. This method assures that new information about a reader's likes and dislikes will be transmitted recursively from all information units representing concepts upon which the user has passed judgement to all related information units present in the index space (Figure 10). For example, if our graph has an index unit node "Volume", we assign a probability which expresses our degree of belief that articles classified under the term "Volume" will be of use to the hypertext user. If index unit terms "ICF" (intracellular fluid) and "ECF" (extracellular fluid) are stated to be aspects of volume, we define between both the "Volume" and "ICF" pair and between the "Volume" and "ECF" pair some conditional probability which expresses the likelihood that if a user likes information units classified under "Volume", she will also like information units classified under the index terms "ICF" or "ECF"(Figure 8).

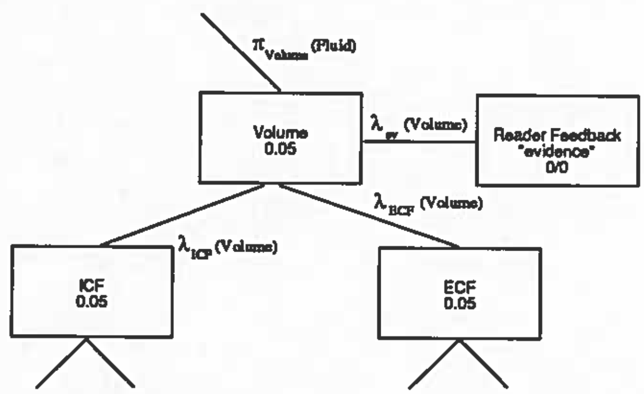
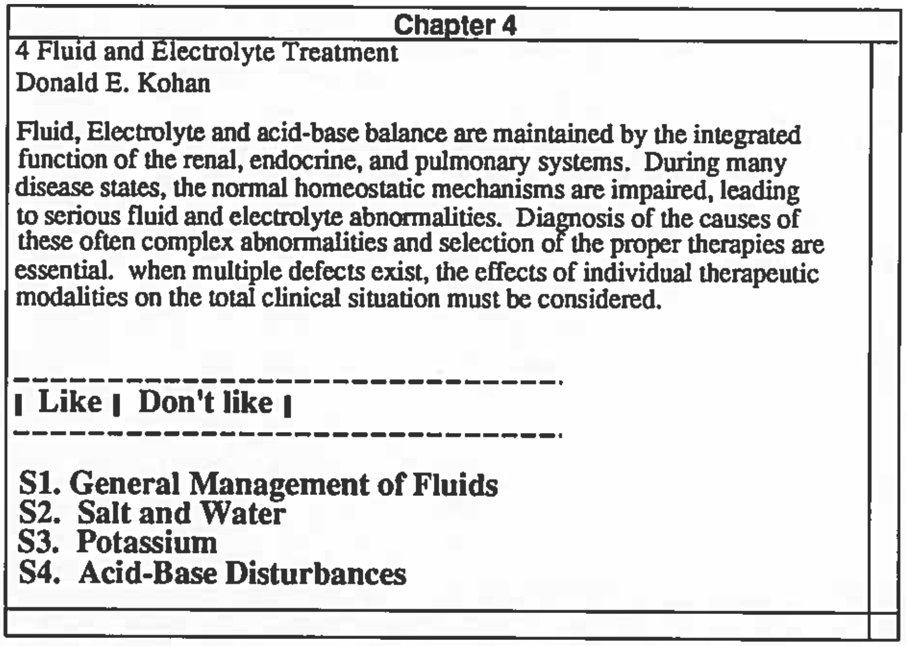


Figure 8: Probabilistic influences on node "Volume". $\pi_{Volume}(Fluid)$ represents the *predictive support* contributed by the single parent node "Fluid". $\lambda(Volume)$ represents the *diagnostic influence* contributed by the two child nodes ("ICF" and "ICF") and a single judgmental evidence node representing reader feedback. This structure is similar to the networked structures illustrated in the the work of Croft and Turtle [Crof89]. Pi and lambda messages influencing the value of other nodes are not represented in this figure. (This figure is adapted from Pearl, [Pearl88, p. 163]).



Chapter 4

4 Fluid and Electrolyte Treatment

Donald E. Kohan

Fluid, Electrolyte and acid-base balance are maintained by the integrated function of the renal, endocrine, and pulmonary systems. During many disease states, the normal homeostatic mechanisms are impaired, leading to serious fluid and electrolyte abnormalities. Diagnosis of the causes of these often complex abnormalities and selection of the proper therapies are essential. when multiple defects exist, the effects of individual therapeutic modalities on the total clinical situation must be considered.

| Like | Don't like |

S1. General Management of Fluids
S2. Salt and Water
S3. Potassium
S4. Acid-Base Disturbances

Figure 9: A typical document space information unit from a LISP handbook prototype. The lower portions of the IU represent "active text". The lower four buttons activate "child" IUs. The "Like" and "Don't Like" buttons provide feedback to the index space. This card is indexed under the terms "fluid","electrolyte", and "acid-base". Selecting the "Like" or "Don't like" button will increase by one the numerator or denominator respectively of the evidence node associated with each of the three index space nodes representing the index terms, changing the degree of belief value assigned to the corresponding three index space nodes. These changes are propagated throughout the index space network using standard Bayesian techniques. Compare this representation with that depicted in our earlier HyperCard prototype (Figure 1).

Calculating Degree of Belief in an Index Space Node

Belief networks present probabilistic relationships as a directed acyclic graph where nodes represent variables and edges represent probabilistic relationships. In many "sparse" networks, this approach significantly diminishes the worst case computational complexity because one states explicitly which variables have an effect on others. Using the methods and notation of Pearl [Pearl88], the belief distribution of the index space node "Volume" can be computed from the following three items of information:

1. The current strength of the support, $\pi_{Volume}(Fluid)$, contributed by the node "Fluid", the parent of "Volume", which in turn is dependent on all evidence (e^+) from ancestors of "Volume":

$$\pi_{Volume}(Fluid) = P(Fluid \mid e^{+}_{Volume})$$

2. The current strength of support, $\lambda_{Y_j}(Volume)$, contributed by each of the j children of "Volume", which in turn is dependent on all evidence (e^-) from descendants of "Volume":

$$\lambda_{Y_j}(Volume) = P(e^{-}_{Y_j} \mid Volume)$$

$$\lambda\,(Volume) = \prod_j \lambda_{Y_j}\,(Volume)$$

3. The fixed conditional probability matrix P(*Volume*|*Fluid*) that relates the node "Volume" to its immediate parent "Fluid". This relationship is needed to calculate an overall influence of "Fluid" on "Volume":

$$\pi(Volume) = \sum_{Fluid} P(Volume \mid Fluid)\, \pi_{Volume}(Fluid)$$

Because the node "Volume" "blocks" parents from children (Figure 10), these expressions can simply be combined to represent the total degree of belief in the value of an index space concept node:

$$\text{BEL(Volume)} = \alpha\lambda\text{(Volume)}\ \pi\text{(Volume)}$$

where α is a normalizing constant rendering Σ BEL(*Volume*) = 1. Similar relationships hold true for all nodes in a hierarchical network (Figure 11).

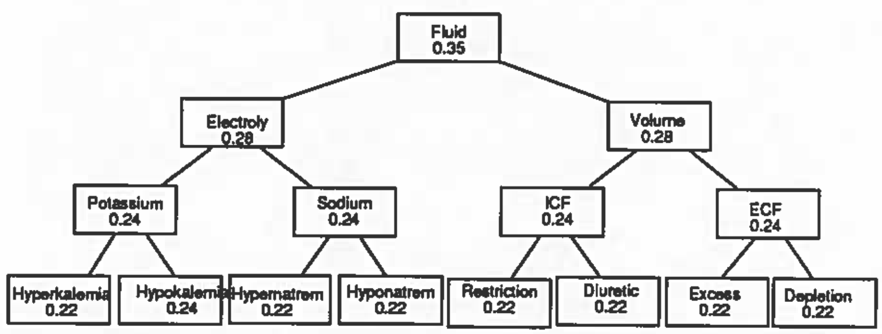


Figure 10: Reasoning over a hierarchical index space. The configuration of hierarchical index spaces allows Bayesian reasoning algorithms to compute in polynomial time. In this example, the node "volume" is the only parent of nodes "ICF" (intracellular fluids) and "ECF" (extracellular fluids). As the only parent, "Volume" *blocks* "ICF" and "ECF" from knowledge about the value of the parent node "Fluid" and all nodes in the left subtree. From a probabilistic viewpoint, complete knowledge about the value of the term "Volume" makes irrelevant knowledge about the value of all other non-descendant nodes like "Fluid" and those in the left subtree. The relationship between two nodes like "Volume" and "ICF" is represented by a conditional probability matrix. Values for the matrix are provided by observation of reader interests, empirical methods, or a heuristic function.

Updating Degree of Belief in an Index Space Node

If one is presented with many IUs indexed under "Volume", a favorable response to reading will increase a likelihood ratio consisting of the number of "Volume" cards read with a favorable response divided by the number of "Volume" cards read with an unfavorable response. The likelihood ratio is expressed through an evidence node associated with the

"Volume" term. This evidence will increase the degree of belief that one will favor other IUs classified under the term "Volume". The conditional probability relationships associated with the "Volume" node will also increase the degree of belief that the reader will like "descendant" terms like "ICF" (intracellular fluid) and "ECF" (extracellular fluid). By similar probabilistic relationships, one will also more likely favor "ancestor" terms like "Fluid" (Figure 11).

A change in the ratio of "favored" to "disfavored" cards indexed under "Volume" creates positive feedback by sending new lambda messages from the "evidence" node for "Volume" to the parent node "Volume". "Volume" in turn sends new lambda messages to its parent "Fluid" and new pi messages to each of its j child nodes Y_j. Pearl discusses this procedure in greater detail in Chapter four of Pearl's recent monograph [Pearl88].

$$\lambda_{evidence}(Volume) = \sum_{evidence} P(evidence|Volume)$$

$$\lambda_{Volume}(Fluid) = \sum_{Volume} P(Volume|Fluid)$$

$$\pi_{Y_j}(Volume) = \alpha\pi(Volume)\prod_{j \neq k} \lambda_{Y_k}(Volume)$$

In this case, the new evidence favoring "Volume" increases the likelihood ratio. This increases the BEL(*Volume*) to 0.79. Propagation increases BEL(*ICF*) and BEL(*ECF*) to 0.50, and increases BEL(*Fluid*) to 0.69 (Figure 12).

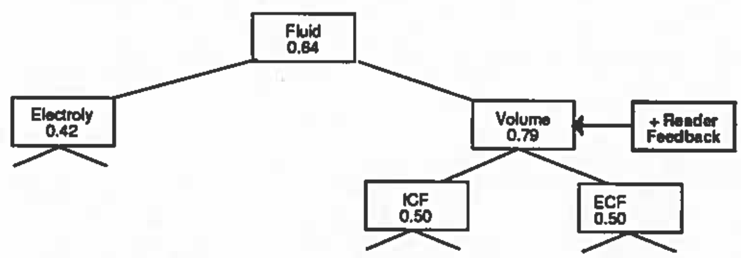


Figure 11: Positive feedback increases the likelihood ratio associated with the "dummy evidence" node associated with the term "Volume". This increase will lead to an increase in the degree of belief in the node "Volume". Increasing the belief that "Volume" is useful also increases the belief that terms near "Volume" in the index are useful. By the conditional probability relationships between "Volume" and both "ICF" and "ECF", the values of the latter two terms are increased. Because of the probabilistic relationship between "Fluid" and "Volume" the value of the term "Fluid" increases. By similar probabilistic calculations, the change in value of the term "Fluid" is propagated down through all values in the left subtree.(See Section 7.3)

Negative feedback will also affect the distribution of degree of belief values over the index space. If a reader favoring "Volume" IUs now encounters many unfavorable IUs indexed by the term "Electrolyte", the resulting negative feedback will decrease the degree of belief that the reader will like other "Electrolyte" terms. Immediate effects due to belief update propagation include a decrease in the value of descendant nodes like "Potassium" and "Sodium", and a tendency to decrease the value of the ancestor node "Fluid". Because the "Fluid" node now receives positive feedback from the "Volume" node and negative feedback from the "Electrolyte" node, its final value will depend on the relative degree of feedback from the two descendant nodes (Figure 12). In this case, a low likelihood ratio from the evidence node for "Electrolyte" decreases BEL(*Electrolyte*) to 0.13 and both BEL(*Sodium*) and BEL(*Potassium*) from 0.24 to 0.16. BEL(*Fluid*), facing opposing influences, is increased from its initial value of 0.35 to a new value of 0.51 (Figure 12).

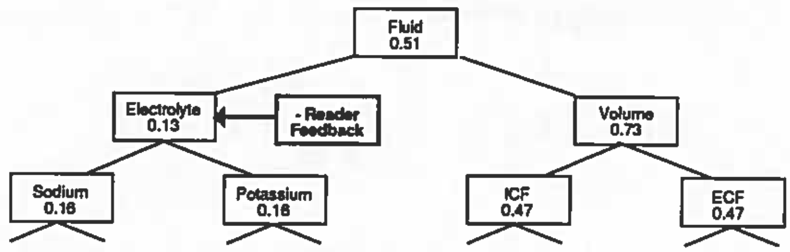


Figure 12: Negative feedback decreases the likelihood ratio associated with the "dummy evidence" child of the "Electrolyte" node. This decreases the value of "Electrolyte", which causes a corresponding decrease in the belief of neighboring terms.

Preliminary Evaluation

Our current investigations use only small information spaces but still suggest that probabilistic inference techniques can be applied to hierarchical index spaces in a manner that yields properties helpful to information retrieval from hypertext document spaces. Belief values of nodes of interest to a reader are increased in value, while nodes not of interest to the reader generally decrease in value. In some cases, the value of all nodes drifts upward, but the increase in value for nodes of interest predominates (Figure 13). In situations of repeated use, we find that personal preferences are better represented by the degree of *change* in belief values than by the *absolute* degree of belief values for index space nodes. In a sense, absolute values reflect the consensus of a group or of an individual over time. Relative changes reflect how one's current interests differ from the consensus.

Potential for Reasoning over the Index Space

Our approach offers many interesting possibilities. For example, the sharing by a collaborative working group of an index space "web" could allow a program to retrieve automatically those portions of a new hypertext that are likely to be perceived as valuable to the group. On the other hand, obtaining the difference between the index space values of an individual and the values of a group might allow for the identification of those portions of the hypertext which most likely characterize the difference in interests between the individual and the group.

Our use of belief networks differs from the method proposed by Croft and Turtle [Croft89]. These investigators describe a network containing nodes from both the index space and the document space. Their architecture is designed to support relationships between pairs consisting of any combination of index space term and hypertext information unit. One elicits probabilities of the description of a concept C given a document space node N. Probabilistic "link" representations relating pairs of document space nodes are also supported. All other supporting information (e.g., knowledge about the meaning of terms) is expressed as evidence nodes in a manner similar to that used in our program to represent reader feedback from document space to index space node likelihood ratios. The truth value of a query Q is calculated *sequentially* for each hypertext node N in the document space. As is the case in our program, Croft and Turtle's conceptually powerful proposal requires significant compromises to elicit probability matrix data and ensure computational tractability.

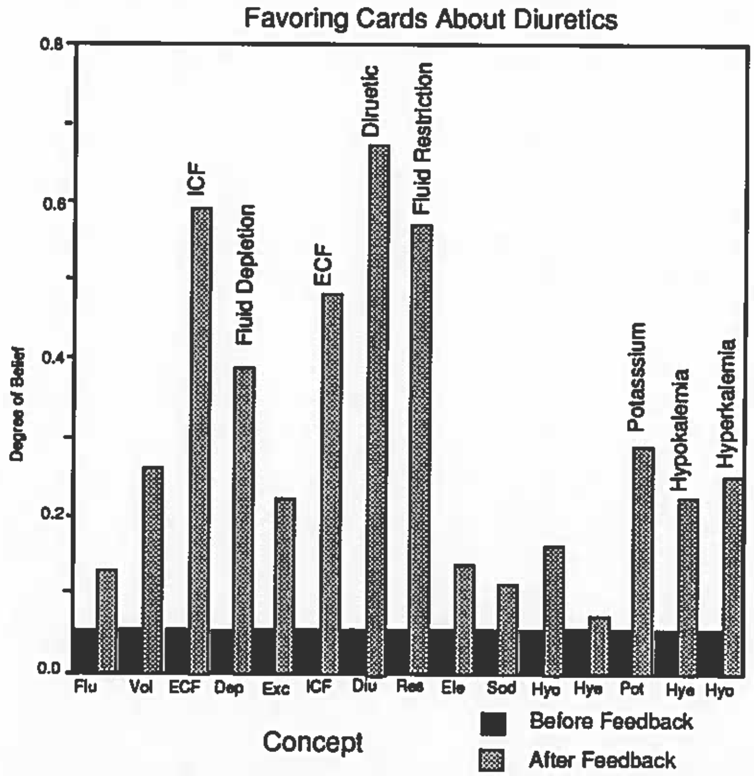


Figure 13: An Example using a Small Index Space. These data are the result of approving a number of information units concerned with diuretic therapy (See Figure 10). As expected, the preference for the "Diuretic" leaf node term is increased, as is the preference for the ancestor term "ICF" (intracellular fluid) and a sibling term "Res" (fluid restriction) . "Potassium", "ECF", and "Fluid Depletion" are increased in part because of co-occurrence in IUs and in part because of index space reasoning.

Reasoning Over Networked Information Spaces

Several characteristics of some hypertexts lead one to believe that it will be difficult to adapt a belief network approach to the networked information space architectures required by many applications. First, it can be shown that applying exact Bayesian computation methods to a complex graph is NP-complete [Coop87]. However, we believe that approximation methods like those published by Pearl and Chavez offer exciting potential for evaluation in an information retrieval setting [Pearl86,Chav88]. Second, the incorporation of multiple semantic link types makes probabilistic computation difficult. We believe that a transformation of an index space from a semantic net representation to a predicate calculus representation allows one to perform computation using the sound foundations of probabilistic logic, but the computational complexity of this technique is equally as intimidating [Nils86]. We believe that this problem will remain one of the challenges facing proponents of Bayesian approaches.

Of greater concern is the broader hypothesis underlying adaptive information retrieval: Can it be shown that a learning system will converge on the appropriate set of index space descriptors given imprecise information? Valiant argues that in some cases the computational complexity of this problem is insurmountable [Val84]. Our hypothesis is that even incomplete or mildly misleading results from the index space can provide an adjunct to other information retrieval and hypertext browsing techniques. We believe that most readers will benefit from imprecise results and that they will recognize circumstances where our algorithms are not performing effectively. The evaluation of this hypothesis awaits formal evaluation using larger hypertext information spaces.

Our approach also requires methods to assert "changing the topic". Medical use is characterized by episodic pursuit of a wide variety of topics, and it will be important to recognize when one has switched from one problem investigation to another. One will also have to accommodate *human* learning. As we learn, we often see an abrupt change in the desirability of information units. Information that was once invaluable can suddenly become obvious, and one must ensure that one's program recognizes when presentation of an information unit becomes a nuisance.

CURRENT RESEARCH STATUS

We currently are enhancing a number of prototype medical handbooks. Our current research centers on the index space, where we are attempting to provide automatic hypertext guidance by implementing algorithms of Pearl [Pearl86], Lauritzen [Spieg88], Chavez [Chav88] and Shachter [Shac86]. Our hypertext prototypes are constructed and tested using SUN-4, Macintosh, and NeXT hardware. In the near future, we will incorporate our medical hypertexts into the extensive radiology picture archiving and communications prototypes under development in the Electronic Radiology Laboratory of Washington University's Mallinckrodt Institute of Radiology [Jost89,Cox88]. Using these advanced workstations, we expect to evaluate the performance of our hypertext prototypes in "real-life" biomedical settings.

ACKNOWLEDGEMENTS

The authors wish to thank Michael Kahn, Tim Oren, Ingo A. Beinlich, H. J. Suermondt, and Steve Weyer for their helpful comments. This research is supported in part by the New Medical Foundation and through an equipment grant from Apple Computer, Inc. Dr. Frisse is a Teaching and Research Scholar of the American College of Physicians.

REFERENCES

[Belew86] RK Belew. Adaptive information retrieval: Machine learning in associative networks. PhD thesis, University of Michigan, Ann Arbor, Michigan, 1986.

[Bilof88] H Bilofsky. The genbank genetic sequence data bank. **Nucleic Acids Research**, 16:1861 - 1863, March 11, 1988.

[Blair85a] D Blair and M Maron. An evaluation of retrieval effectiveness for a full-text document retrieval system. **Communications of the ACM**, 28:289 - 299, March 1985.

[Book83] A Bookstein. Information retrieval: A sequential learning process. **Journal of the American Society for Information Science**, 34:331 - 342, September 1983.

[Came88] G Cameron. The EMBL data library. **Nucleic Acids Research**, 16:1865 - 1867, March 11 1988.

[Chav88] RM Chavez and GF Cooper. A fully polynomial randomized approximation scheme for the bayesian inference problem (working paper). Technical report, Stanford University, Stanford California, Fall, 1988.

[Cons89] MP Consens and AO Mendelzon. Expressing structural hypertext queries in GraphLog. In **Proceedings of Hypertext 89**, Pittsburgh, Pennsylvania, 1989.

[Coop87] GF Cooper. Probabilistic inference using belief networks is NP-hard. Technical Report KSL-87-27, Medical Computer Science Group, Stanford University, Stanford, California, 1987.

[Cox88] G Cox and SJ Dwyer. Picture archival and communication systems. In **RISC Ninth Conference on Computer Applications in Radiology**, June 1-4, 1988.

[Croft89] WB Croft and H Turtle. A retrieval model incorporating hypertext links. In **Proceedings of Hypertext 89**, Pittsburgh, Pennsylvania, 1989.

[Frisse88b] ME Frisse. From text to hypertext. **Byte**, 13:247 - 253, October, 1988.

[Frisse88a] ME Frisse. Searching for information in a hypertext medical handbook. **Communications of the ACM**, 31:880 - 886, 1988.

[Heck86] D Heckerman. Probabilistic interpretations for MYCIN's certainty factors. In L Kanal and J Lemmer, Editors, **Uncertainty in Artificial Intelligence Artificial Intelligence**, pages 11 - 22, Amsterdam, 1986. North Holland.

[Jost89] R Jost, W Wessel, G Blaine, J Cos, and R Hill. PACS - Is there a light at the end of the tunnel? In **Conference on Medical Imaging III**, SPIE, pages 1093 - 1125, Newport Beach, California, January 29 - February 3 1989.

[McCu85] BP McCune, RM Tong, JS Dean, and DG Shapiro. RUBRIC:a system for rule-based information retrieval. **IEEE Transactions on Software Engineering**, SE-11:939 - 945, September 1985.

[Nels87] T Nelson. **Literary Machines**, 87.1 edition. Project Xanadu, 8480 Fredricksburg #138, San Antonio, Tx. 78229, 1987.

[Nils86] N Nilsson. Probabilistic logic. **Artificial Intelligence**, 28:71 - 87, 1986.

[Oren87] T Oren. The architecture of static hypertexts. In **Proceedings of Hypertext87**, pages 291 - 306, University of North Carolina, Chapel Hill, November 13-15 1987.

[Oren88] T Oren. The CD-ROM connection. **Byte Magazine**, 3:315 - 320, December 1988.

[Pearl88] J Pearl. **Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.** Morgan Kaufmann, San Mateo, CA, 1988.

[Pearl86] J Pearl. Evidential reasoning using stochastic simulation of causal models. Technical report, University of California, Los Angeles, Los Angeles, California, September, 1986.

[Salt88] G Salton. Another look at automatic text-retrieval systems. **Communications of the ACM**, 29:648 - 656, July, 1988.

[Schat85] BR Schatz. Telesophy. Technical Report Bellcore TM-ARH-002487, Bell Communications Research, 445 South Street, Morristown, NJ. 07960-1910, 1985.

[Shac86] R Shachter. Evaluating influence diagrams. **Operations Research**, 34:871 - 872, 1986.

[Shas85] D Shasha. Netbook - a data model to support knowledge exploration. In **Proceedings of the Eleventh Conference on Very Large Databases**, Stockholm, August, 1985.

[Sidm88] K Sidman, D George, W Barker, and L Hunt. The Protein Identification Resource PIR. **Nucleic Acids Research**, 16:1869 - 1871, March 11, 1988.

[Spieg88] S.L.Lauritzen and D Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. **J. Royal Statistical Society**, 9:157 - 224, 1988.

[Trigg83] R Trigg. A network based approach to text handling for the online scientific community. PhD thesis, University of Maryland, Department of Computer Science, College Park, Maryland, November 1983. Technical Report 1346.

[Val84] L Valiant. A theory of the learnable. In **Proceedings of the 16th Annual ACM Symposium on the Theory of Computing**, Washington D.C., April, 1984.

[Warr81] KS Warren, Editor. **Coping with the Biomedical Literature.** Praeger Publishers, New York, 1981.