

Hypertext Linking as a Model of Expert Indexing

Judith A. Tessier

Division of Library and Information Science
San Jose State University
San Jose, CA 95192-0029

The position of this paper is that hypertext environments provide a new domain in which to explore the cognitive processes of indexing and classification and that the study of hypertext links allows the classification researcher to test the adequacy of current indexing theory. The process of linking segments of text in a hypertext system is a process similar to development of classes and term relations in the indexing process. However, linking methods can allow for linking *without* articulating the nature of the relationship: two segments may be connected on an intuitive basis rather than on the basis of a stated relationship between segments. Exploration of these “intuitive” links may suggest new approaches to indexing, that is, alternative facets, or connections among text segments, that contrast to current indexing languages.

This paper begins to explore the question: how congruent are the relationships expressed through formal indexing systems with those relationships established by subject experts within the hypertext environment? Since hypertext is a relatively recent development, the user may well be less constrained to follow conventions in establishing relationships. Indeed, one of the challenges in hypertext is to obtain coherent views of the data and a navigational sense through the material. The argument then is that structures created by the expert in the hypertext environment might suggest approaches new to the indexing process and language design.

Traditional classification systems have been developed as a collaboration of the subject expert and the information scientist. The subject expert contributes a thorough understanding of the parameters, structure, dimensions, and semantics of a field of knowledge; the information scientist contributes the knowledge of organizational schema, index language design and retrieval system properties. The position of this paper is that connections among concepts in a body of literature are in some instances “intuitive” to the subject specialist. Traditional systems for indexing language and classification system design may fail to take these intuitive links into account.

In contrast, hypertext is designed to encourage movement from selected points in one document to other points in that document, or to selected points in other related documents. The hypertext data model is a network, a set of nodes and links between the nodes. The user can traverse to any one of a set of predetermined nodes. The links create a series of alternative pathways through the material. The hypertext author or “composer” takes a variety of textual or multimedia materials, and connects these “nodes” to each other by “links” or connections which allow the user to move through the material. The interesting cases are those in which the subject specialist sees a connection, yet may not be able to state a rule for that connection, or uses a rule that does not conform to current indexing language relations.

Two approaches suggest themselves in the investigation of “intuitive” links:

1. an analysis of hypertext files in which links are generated by subject experts to serve their own browsing and retrieval functions: what kinds of nodes are linked? what is the nature of the relationship? does it fall into one of the standard types of indexing relationships (such as action: actor; part: whole)? Or are there instances of linking which fall outside such classification? If links do fall outside such classification, are there patterns among these special cases?
2. direct observation and protocol analysis of the expert during the linking process: how does the expert describe the links in his own language? why are certain links established? what links are wanted but not generated? how does the system encourage or constrain the linking process? is the process iterative, i.e., does the expert return to material to refine or systematize the links?

This paper describes initial investigations using the first method of inquiry: a description of two hypertext files to identify the kinds of relations implemented; discussion of an interesting study that compares three hypertext implementations of the same textual files; and description of two hypertext systems with different capabilities for the generation, identification, and organization of links.

I. DESCRIPTION OF TWO HYPERTEXT FILES

In this section, two products are described in terms of their organizing principles. The first, *In the Holy Land*, is a video, audio, and textual database concerning the Middle East; it was developed from a series of ABC news presentations and stresses nonsequential access to information stored in video format. Its user interface was written in HyperCard¹. The second, *Imitatio Christi [The Imitation of Christ by Thomas a Kempis]* is a text only file, with emphasis on the text of a series of meditations by Thomas a Kempis. It is also implemented in HyperCard. These two are described in some detail because they exemplify the organizing principles commonly used in hypermedia files. Hypertext and hypermedia systems do accommodate several organizing principles within one package; each of these programs uses a variety of organizations to allow the user flexible, variable access. What organizing principles actually were used? and how systematic were the links?

For this investigation, the author identified the organizing principle from the material. Typically, a screen or series of screens would display user options. The "organizing principle" was the arrangement that was expressed either through the order of the list of choices or through the actual order of the screens when selected sequentially. In many instances, the principle was not made explicit to the user, i.e., labeled as "chronological" or "alphabetical", etc. Similarly, the author scanned through the files to identify what kinds of access keys were used: i.e., what kinds of choices or data elements were available to the user, e.g., events, or dates, or topics. Again, the type of access key was rarely stated within the application (except for format type choices, e.g., biography). In many instances, the organizing principle and the type of access keys were obvious enough that explicit labeling was unnecessary. In these descriptions, **bold represents the organizing principle**; underlining represents access keys.

1. HyperCard is a trademark of Apple Computer.

Example 1: Hierarchical organization; on historical, geographic, and topical dimensions: In the Holy Land. NY: ABC, 1988-89.

I. History and Timeline

A. Timeline with logical date spans and headings

1. finer timeline with date spans and events (text)

a. chronologically arranged date spans and events have topical links, e.g.:

.1938-1945 Holocaust (text)

.Perspectives (video interviews; viewpoints by various personalities)

.Holocaust images (photographs)

.Biography: Adolf Hitler (text and photo)

b. chronologically arranged specific events (text) have topical links, e.g.:

.1897 First Zionist Congress (text)

.Perspectives (video interviews; various personalities)

.Document: 1st Zionist Declaration (text)

.Document: 7th Zionist Declaration (text)

.1945 Holocaust and Immigration (text)

.Perspectives (video interviews; various personalities)

.Graphic: number of Jewish immigrants

B. Index by date (chronological list)

C. Index by named event (alphabetical list)

D. Glossary by term (linked to boldfaced text in above)

II. Geography

A. Map of world with Middle East highlighted

1. Map of Middle East with country names (text)

III. The three religions

A. List of 3 religions (Islam, Judaism, Christianity: text on each)

B. Timeline of Jerusalem

C. Map of Biblical locations

IV. Issues

A. Alphabetical list of issues, e.g.: Camp David, Destruction of Israel, Leaders & People, Mutual Destruction

V. Choice of Israeli, Palestinian children (video interviews)

VI. Text describing Intifadah

Example 2. Linear organization from original text: Brian Thomas and Philip A. Mohr, Jr., Imitatio Christi [The Imitation of Christ by Thomas a Kempis]. Rosemead, CA.: San Gabriel River Stacks, 1988 (available from BMUG).

I. Text of the work (a series of meditations, each 1-3 screens in length).

A. Sequential presentation from text

II. Sequential, hierarchical table of contents; listing of topical titles of meditations;

Introduction

Spiritual Life:

Vanities, Humility, Teaching Truth, etc.

The Inner Life: specific titles of meditations

Inward Consolation: specific titles of meditations

III. Topical browsing guide

A. **Alphabetical** listing of quotations: e.g., "A Better Way to God,"
"Anxiety about Salvation," "Being Born Again," etc.

B. **Sequential** listing of prayers: "For the Grace of Devotion," "For Love," "That
the Will of God be Fulfilled," etc.

IV. **Alphabetical** index to Bible references, e.g., Acts, 1:7.

V. **Random** access index to meditations (opens to a randomly selected screen

VI. Word search with HyperCard "Find."

VII. User defined links to sequential personal journal entries.

Discussion.

The two files described above each implemented a variety of organizing principles, including chronological, topical (arranged logically, alphabetically, and sequentially from the material), geographical, and division by format. They differed from standard printed material in the variety of alternative accesses to the material, i.e., in the browsability of the file. That is, more, alternative structures were made available to the user than might be expected in printed formats. Both files showed a high degree of organizational structure and made extensive use of standard indexing approaches.

In a way, this investigation was disappointing, because there were few access keys or arrangements that are not commonly implemented in book organization and indexing. Since the interest in the investigation is to find those elements which do not follow common indexing principles, it was the "surprises" that were sought. These examples, while they demonstrated variety, did not yield many surprises.

Two factors may have contributed to these findings. First, both systems were implemented in HyperCard, which may impose constraints on the designer that discourage the less systematic links. Second, the hypermedia design literature stresses the importance of simple structures that can be communicated easily to the user; this too may constrain the designer to simplify rather than to impose suggestive links. Each of these issues is discussed here briefly.

In HyperCard the designer establishes a link from one place on one screen to another screen by installing a "button": a hotspot on the screen that the user selects to move to the other node. The hotspot is associated with a physical location on the screen. The hotspot is labeled so that the user can recognize that there is a link and select it if he so chooses. These labels, in most instances, are verbal labels, such as a single keyword or a list of keywords, each representing a choice. Button labels can also be a graphic device, such as a small icon designating a common action, such as moving forward, or stopping, or a particular type of material, such as a photo or biographical entry. Alternatively, a word in the text, or a phrase, can be highlighted, often shown in bold, indicating a hotspot that will move the user to a definition, an elaboration, or material connected to that concept in some manner. The two examples above used keyword lists and standard icons primarily; in-text buttons were used quite sparingly. HyperCard may discourage textual links because the link is

associated to a physical location on the screen rather than to the textual matter itself; consequently, when material is moved around the screen, the hotspot, or button, must be moved separately. In short, HyperCard encourages links that look like lists or buttons, rather than links embedded in text. Yet the cognitive process of association of “intuitive” links may be encouraged by textual material, rather than lists.

Second, the hypermedia design literature for composers encourages simplicity and predictability. These qualities are seen to enhance user navigation and to protect the user from confusion concerning what material has been covered. Rubens states:

Since the developer, in all but a few hypermedia systems, creates the links, users still must trust the rational approach, as well as the mindset, of the developer....some intelligent structuring must occur early in the development cycle....Hypermedia also requires a governing rationale for links. Employing one linking rationale in one section of a product and shifting to another rationale later asks users to remember too much detail. (Rubens, 1989, p. 18)

Brockman (1989) suggests that hypertext authors make a trade-off between the risk of user confusion in navigation and the expressiveness of the hypertext linking structure. Sequential and grid structures essentially reduce confusion and expressiveness, while tree and web structures may increase both. In short, making systematic and explicit rules for links is one of the cardinal principles of hypermedia design. And linear structures are advocated for overall organization. Yet these may inhibit the designer from implementing those “intuitive” or difficult to define connections, and indeed, the expressiveness of the hypertext material.

II. ALSCHULER'S STUDY OF THREE IMPLEMENTATIONS OF THE SAME MATERIAL

Liora Alschuler (1989) reported a study of three hypertext implementations of the same textual material. She concluded that hypertext linking in this instance was highly “subjective,” and argued that simple keyword searching provided a better access tool than the links installed by subject experts. The study is interesting as an attempt at an experimental approach to the understanding of linking. Because both the systems and the composers on each system varied, her conclusions are not clearly attributable to system or composer, although her descriptions suggest that composer variability was at the heart of some of the difficulties.

Alschuler studied the ACM sponsored publication, *Hypertext on Hypertext*, three hypertext products in which 6 papers from the ACM *Hypertext '87* conference proceedings were published in three hypertext systems: HyperCard, Hyperties¹, and KMS². She defined the three systems as sharing “hand-crafted” hypertext linking: “the decision to link is in the hands of the composer.”

Each system allowed the composer to install “direct links,” jumps from one node to another. All three systems used direct links to establish a hierarchical view and access to the parts of the papers.

1. Hyperties is a trademark of Cognetics Corporation.

2. KMS is a trademark of Knowledge Systems Incorporated.

The segmenting of the papers and the resulting hierarchies differed radically among the three systems. Two of the systems segmented the papers into screen sized units and derived a hierarchical view complete enough to access the individual screens; the third system segmented along paper sections rather than screen size and developed a less detailed hierarchy. All three systems used paper titles and paper section headings for the hierarchical view; two systems (using screen size) subdivided more specifically according to subtopics within the sections.

One system, HyperCard, also used direct links to move the user directly from one node to another nonsequentially when two cards were linked; Hyperties and KMS moved the user to an index and through it to the related material. Alschuler found inconsistencies in the installation of the links in HyperCard. For example, a series of links to direct the user through various definitions of hypertext moved from general to specific and back to general definition when started at one place; but from specific to general when started at a different place; the user was also shunted back and forth between different articles rather than completing one before starting another. The other two systems, which essentially gave the user a listing of destinations before the jumps were executed, tended to be more comprehensive and more systematic (either logical or alphabetical) in the ordering of the entries.

Each system also provided the user with a alphabetically arranged topical index; again, there were marked differences in treatment. KMS and HyperCard each had 26 first level topics (analogous to headings in a book index) but only 3 of the topics matched. That is, indexer consistency for the two systems was low although they coincided on the number of entries. In contrast, the Hyperties system had only 5 first level topics--a rather sparse index by book indexing standards.

Alschuler identified a number of other differences in the three implementations. In most cases, the systems offered the capability of doing the same work; the differences came in how the systems were used to link the text, that is, in the human factor. She also compared the composer written indexes available in each system to the word search capabilities also available. In her test cases, such as a search on Engelbart, she found that the author indexing, and author installed links, were less comprehensive than the string search results.

III. HYPERTEXT AUTHORS AS INDEXERS

The position taken in this paper is that hypertext authors might provide a model for the expert as indexer. Do hypertext authors link materials in new or nontraditional ways that would be suggestive for indexing theory?

In each of the case studies reported above, the hypertext files did not show evidence of linking in ways uncommon in traditional indexing. The two files described in some detail above used traditional principles of organization. Alschuler's study approach was somewhat different, but her observations were that the linking was nonsystematic, or "subjective"; experts used common principles of organization but did not do so exhaustively. The "linker errors" were errors of omission of appropriate access keys; the hypertext authors "underindexed." At one level, the studies refute the original position of this paper, that hypertext files and authoring may provide new insights into the indexing process. The cases under study lacked the kind of "intuitive" or unusual linking that was sought.

This section suggests characteristics a hypertext application might have that would indicate unusual linking. That is, what kinds of case studies might yield insights into unconventional linking?

- a. **File characteristics:** Unusual linking may occur in files in which links are implemented in a network or weblike structure rather than in a sequential or grid-like structure (Brockman, 1989) that is, a highly “expressive” file in which regularity of the pathways is less critical. Each of the files in the cases described here placed value on regularity (the first two as published products with book analogies; the ACM files because they served as exemplars). Also, files in which links are plentiful and dense, i.e., a greater number of links to each node, may well indicate the situation in which links were assigned for expressiveness. The University of California’s Issue-Based Information Systems (Seyer, 1991) encourages density in working files by predefining useful link types such as “suggested by,” “questions,” “supports,” and “responds to.”
- b. **File purposes:** Files designed for work or collaborative purposes rather than for use by third parties may well contain more “ideosyncratic” links, since concern with regularity or rational design may be less important than the needs of the author-user. Irish (1989) described the use of NoteCards¹ by colleagues as an annotation system that encouraged such a development of a series of alternative pathways, each designed by different author-users for their own purposes. The process was supported by such techniques as different type-faces to denote different authors, date stamping, and notecards that documented and could replicate each of the alternative pathways through the data.
- c. **Linking facilities:** Linking takes additional author effort; it stands to reason that the author may be more inclined to “play” with the material through linking when the process requires little effort. Alschuler’s finding that even keyword searching was more exhaustive than author links suggests that automated methods for some linking and relatively effortless mechanisms are needed. Mechanisms for maintaining link records, displaying them to ensure completeness, and editing facility may also support more developed networks of links.

This paper has suggested that linking in hypertext is analogous to the indexing process, but that hypertext offers an opportunity to study how the expert indexes. By reviewing cases of linking in selected hypertext applications, some of the issues in the study of linking as an indexing process have been described. In the cases studies here, hypertext linking showed remarkable similarity to indexing. Factors that might characterize particularly useful cases were also discussed.

1. NoteCards is a trademark of Xerox PARC.

IV. REFERENCES

- Alschuler, L.** 1989. "Hand-crafted Hypertext-Lessons from the ACM Experiment," in E. Barrett, ed., *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*. Cambridge, MA.: The MIT Press, p. 343.
- Brockman, R. J., W. Horton, and K. Brock.** 1989 "From Database to Hypertext via Electronic Publishing, in E. Barrett, ed., *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*. Cambridge, MA.: The MIT Press, p. 162.
- Gaines, B. R. and Vickers, J. N.** 1988. "Design Considerations for Hypermedia Systems," *Microcomputers for Information Management*, 5, no. 1: 1-27 (Mar. 1988).
- Irish, P.M. and R. H. Trigg.** 1989. "Supporting Collaboration in Hypermedia: Issues and Experiences." In E. Barrett, ed., *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*. Cambridge, MA.: The MIT Press, p. 90.
- Rubens, P.** 1989. "Online Information, Hypermedia, and the Idea of Literacy." In E. Barrett, ed., *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*. Cambridge, MA.: The MIT Press, p. 18.
- Seyer, P.** 1991. *Understanding Hypertext: Concepts and Applications*. Blue Ridge Summit, PA.: Windcrest Books.