HYPERTEXT - AN AID TO INFORMATION SEEKING

D.S. Bedi

Defence Scientific Information & Documentation Centre, Metcalfe House, Delhi-110054.

ABSTRACT

Hypertext systems feature machine-supported links–both within and between documents–that open possibilities for using the computer as a communications and thinking tool. The article gives an introduction to the world of hypertext. Discusses the features of an idealized hypertext system, focussing on machine-supported links as its essential feature. Surveys some of the most important design issues that go into fashioning a hypertext environment. Outlines the major application areas of hypertext technology and the hyper-text systems developed for these application areas. Highlights the advantages and disadvantages of the hypertext systems.

INTRODUCTION

The exponential growth, increasing complexity and multidisciplinary nature of scientific and other forms of knowledge have exposed human limitations in handling such information. The hypertext and other electronic information systems overcome such human limitations by providing mechanisms for compact storage and rapid retrieval of enormous volumes of textual, numeric and visual data. The importance of these systems lies in their potential capacity to augment and amplify human intellect.

WHAT IS HYPERTEXT ?

The term hypertext, coined by Ted Nelson, describes a vast network of text fragments linked together, an electronic writing and reading system that uses the power of the computers for more than editing and display. A hypertext system has three prominent components :

- a database of text;
- a semantic net which connects the text components; and
- tools for creating and browsing this combination of text and semantic net.

In short, hypertext is a text database plus semantic net plus interface. Hypertext systems allow users to traverse complex networks of information quickly. Authors can easily link passages and references; readers can freely move among text fragments to find sources of quotations, journal article references, definitions, and related passages.

From the writer's point of view, hypertext systems are the next generation of word processing. In addition to word processing features like block moves, search and replace, and spell or style checking, the hypertext writing tools can extend the writing process with telescopic outlines, posted notes that do not affect the main text, electronic book marks, and browsing modes. From the reader's point of view, hypertext systems are a new generation of database management. Full text is accessible from multiple perspectives, for various purposes, and through different search strategies.

FEATURES OF A HYPERTEXT SYSTEM

The concept of hypertext is simple : windows on the screen are associated with objects in a database, and links are provided between these objects. The features of an idealized hypertext system can be outlined as :

- The database consisting of a network of textual nodes.
- The windows on the screen correspond to nodes in the database on a one-to-one basis, each having a name or title displayed in the window.
- The standard window system operations are supported : windows can be repositioned, resized, closed and put aside as small window icons. The position, size, column and shape of the window or icon give cues to the contents of the windows.
- The windows can contain any number of link icons representing pointers to other nodes in the database. Clicking a link icon with the mouse causes the system to find the referenced node and to immediately open a new window for it on the screen.
- The user can easily create new nodes and new links to new nodes for annotations, comments, etc. or to existing nodes for establishing new connections.
- The database can be browsed in three ways : (i) by following links and opening windows successively to examine their contents, (ii) by searching the network for some string or keyword, (iii) by navigating around the text using a 'browser' that provides a direct two-dimensional graphic view of the underlying database.

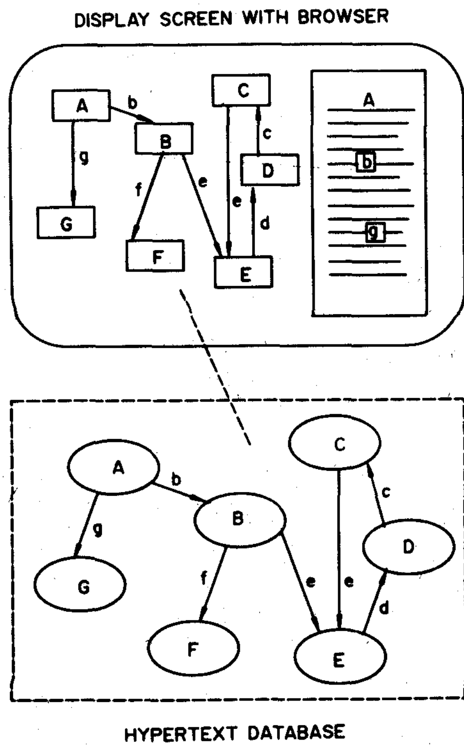

Figure 1.

Fig 1 illustrates how a hypertext browser provides a direct two-dimensional graphic view of the underlying database. Node 'A' has been selected for full display of its contents. The browser view in the Fig shows not only which nodes are linked to 'A' and by which links, but also how the subnetwork fits into the larger *hyperdocument*.

Delimiting hypertext from other electronic information systems

Some of the electronic information systems share features with hypertext, but do not qualify as hypertext systems for want of one or the other of the features of an idealized hypertext systems described above.

Some of these electronic information systems are :

- *Window systems.* Though such systems do have some of the interface functionality, they have no single underlying database–lack the database aspect of hypertext.
- *File systems.* File system has been claimed to be a database where one can move among 'nodes' (files) by simply invoking an editor by their names. But, a sophisticated notion of machine-supported links–a primary feature of a hypertext system–is missing.
- *Outline processors.* With their integrated hierarchical database and interface, the outline processors do approximate hypertext but they provide little or no support for reference between outline entries.
- *Database management systems* have links of various kinds, but lack the single interface to the database–the hallmark of hypertext.

LINKS AND NODES – BUILDING BLOCKS OF HYPERTEXT

Machine-Supported Links

The most distinguishing characteristics of hypertext are its machine support for tracing the references and the speed with which the system responds to referencing requests. The machine-supported link–the referencing device–forms an essential feature of hypertext. It is this linking capability which allows a non-linear organization of text. Links can be used for several functions :

- They can connect a document reference to the document itself;
- They can connect a comment or annotation to the text about which it is written;
- They can provide organizational information;
- They can connect two successive pieces of text, or a piece of text and all of its immediate successors;
- They can connect entries in a table or figure to longer descriptions, or to other tables or figures.

Links can have names and types.

Referential Links

Two points in a hypertext system can be linked by reference and by organization. The reference method is non-hierarchical. It uses referential links that connect points or regions in the text. Referential links are the kind of links that most clearly distinguish hypertext. They generally have two ends, and are usually directed. The origin of the link is called 'link source' which acts as the reference. The other end, 'link destination' functions as the referent. The two ends can be either a point or a region:

> *Link point* is referred to the icon indicating the presence of the link and shows the link's name and type or it may show the name and/or type of the destination node. The display of links can be suppressed to make the documents appear linear.
>
> *Link region* is a set of contiguous characters displayed as a single unit.

Organization Links

Organization links also establish links between points in hypertext, but differ from referential links in that they implement hierarchical information. They connect a parent node with its children and thus form a strict tree sub-graph within the hypertext network graph.

---

* An outline processor is a work processing program which is specialized for processing outlines, in that its main commands deal with movement among, creation of, and modification of outline entries.

Keywords Links

These links provide an implicit linking that occurs through the use of keywords. This ability is a valuable aspect of hypertext as one of the chief advantages of text storage on a computer is the ability to search large and complex documents for substrings and keywords.

Cluster Links

Cluster links are made of links connecting more than two nodes. These are useful for referring to several annotations with a single link, and for providing specialized organizational structures among nodes.

Hypertext Nodes

The hypertext nodes play a significant role in defining the operation performed by a given hypertext system. Most users favour using nodes which express a single concept or idea, and are thus much smaller than traditional files. In this way, hypertext introduces an intermediate level of machine support between characters and files.

Hypertext invites the writer to modularize the ideas into units in a way that allows : (i) an individual idea to be referenced elsewhere, and (ii) alternative successors of a unit to be offered to the reader (for instance, more details, an example, or bibliographic references). The expression of ideas into discrete objects to be linked, moved and changed as independent entities is another feature of the hypertext system that offers enhanced retrieval and recognition over computer processed flat documents. Nodes can be of various types :

Typed Nodes

Some hypertext systems sort nodes into different type. These typed nodes are useful as the types can be used to differentiate various structural forms. Hypertext systems that use typed nodes generally provide a specialized colour, size or iconic form for each node type. These distinguishing features help the user differentiate at a glance the broad classes of typed nodes that he is working with.

Semi-structured Nodes

These are typed nodes which contain labelled fields and spaces for field values.

Composite Nodes

The composite nodes provide another mechanism for aggregating related information in hypertext where several related hypertext nodes are glued together and the collection is treated as a single node with its own name, type, etc. This facility allows a group of nodes to be treated as a single node. The composite node can be moved and resized, and closes up to a suitable icon reflecting its contents. The sub-nodes are separable and can be rearranged through a sub-edit mode.

DESIGNING A HYPERTEXT SYSTEM – SOME ISSUES

Design decisions that affect information seeking in hypertext systems are related to defining access points, creating the user interface, and providing search strategy features.

Defining access points

A key decision for designers is what access points to define and how to link these points. They should be at least as rich as those available in books. Designers must decide how much more to offer and how much unrestrained jumping from point to point to allow. Designers should consider the targeted task domains and typical user population in deciding how fine the access points should be and what links among access points should be visible to users.

Creating the user interface

An interface enables users to perform their tasks by providing selection and feedback mechanisms and input/output devices. In a book, the default sequence is top-down due to the linear nature of the text. The interactive and flexible characteristics of hypertext require users to make more choices in searching for information. Further, each selection requires appropriate and understandable feedback to maintain a fruitful interaction. In view of non-linear organization of information in hypertext, mechanisms for selection and feedback are critical to good design.

Providing search strategy features

Search features like Boolean connectives, string search proximity limits and scope limits facilitate rapid access to information, but cause additional cognitive load on the part of the user and substantial preprocessing of the database itself. The systems that provide only browsing features allow casual, low cognitive load exploration, but are inefficient for directed search tasks. Development of a hybrid system that guides discovery should be considered in all design decisions, thereby striking a balanced analytical and browsing search strategies.

The flexibility/complexity trade-off

Systems transparent to one user may frustrate and impede others. Flexibility inevitably leads to complexity. Just as printed indexes and directories are organized to facilitate retrieval, so electronic information systems must be organized to suit the typical purpose of anticipated users. All designers must grapple with the issue of when to stop adding features. In general, each feature added to a system demands additional machine overhead. Many also add user cognitive load. These effects must be considered in all design decisions.

CLASSIFYING HYPERTEXT SYSTEMS

The hypertext systems can be classified on the basis of their application. There are four broad application areas for which hypertext systems have been developed :

Macro Literary Systems

These hypertext systems were developed for the study of technologies to support large on-line libraries in which interdocument links are machine-supported i.e. all publishing, reading, collaboration, and criticism takes place within the network. Bush's Memex, Engelbert's NLS/Augment, Nelson's Xandadu Project, and Trigg's Textnet are some of the hypertext systems implemented under this category.

Problem Exploration Systems

The hypertext systems falling under this category are highly interactive systems which provide rapid response to small collection of specialized commands for the manipulation of information. These systems are basically tools to support early unstructured thinking on a problem when many disconnected ideas come to mind. Examples are: Lowe's, UNC's WE, and outline processors.

Structured Browsing Systems

These systems are designed primarily for applications involving large amount of information or requiring easy access to information. These are similar to macro literary systems, but are smaller in scale and are applied for teaching, reference, and public information, where ease of use is crucial. CMU's ZOG and Knowledge Systems' KMS, Emacs INFO subsystem, Schneiderman's Hyperties, and Symbolics Document Examinee are notable examples.

General Hypertext Technology

This category includes the hypertext systems having one or more applications, but their primary purpose is experimentation with hypertext itself. These are general purpose systems designed to allow experimentation with a range of hypertext applications (for reading, writing, collaboration, et). Some examples: Brown University's, Intermedia, Textronix Nephena, Pitman's CREF and MCC's Plane Text, and Apple's Hypercard.

HYPERTEXT SYSTEMS - SOME EXAMPLES

Hyperties

Hyperties (hypertext based on the interactive encyclopedia system) enables users to easily traverse a database of articles and pictures by merely pointing at highlighted words or images. Highlighted or coloured words or phrases within the text become the menu items, selectable using a pointing device. Rather than isolated and explicit menu items, the content of embedded menu items provides information and cue for the selection of further information. This embedded-menus approach and simple user interface enable users to explore large databases easily.

Hyperties users merely touch or use a cursor to specify topics of interest. Users may continue reading or ask for details about the selected topic. An article about a topic may be one or more screen long and contain several pictures. As users traverse articles, Hyperties traces the path and allows them to return to previous articles, all the way back to the introductory article. Users can also select articles and pictures from an index.

Hyperties was designed to support easy browsing of text and graphic databases. Results of many evaluative studies demonstrated that even novices find it easy and effective to use. It is distributed by the Cognetics Corporations of Princetors Junction, NJ.

Electronic Encyclopaedia

Electronic Encyclopaedia is the full text of Groliers Electronic Encyclopedia on CD-ROM. The print version of the Encyclopedia occupies 20 volumes. The hypertext version consists of 60 megabytes of text and 50 megabytes of indexes that contain pointers to each occurrence of every word in the Encyclopedia, all occupying less than one-fifth of a single CD-ROM disc. The powerful search software for this system provides rapid access to all occurrences of any word or phrase entered by the user.

The experiments conducted with this system suggest that novices can successfully apply hypertext. The Electronic Encyclopedia is much more controlled than other hypertext systems in that jumps from article to article depend on a list of articles retrieved by a query. A user who wants to see an article not on the retrieved list must pose another query or enter a separate mode that allows look-up of single articles by title. On the other hand, the full text search feature is totally under the control of the user.

THE ADVANTAGES OF HYPERTEXT

Though the intertextual references have been a part of traditional literature, the importance of hypertext is that the references are machine-supported. Hypertext has offered new possibilities of authoring and design on one hand and for reading and retrieval on the other. Authoring deals with structuring of ideas, order of presentation, and conceptual exploration. The unit of this level of the authoring is the idea or a concept. Since the idea can be expressed in a node, this level of work can be effectively handled by hypertext. As the writer thinks of new ideas, he can develop them in their own nodes, and link them to existing ideas. The specialized refinements of hypertext environment assist the movement from an unstructured network to the final polished document. Hypertext may also offer new possibilities for accessing large information sources. The operational advantages of hyertext can be summarized as :

Ease of tracing references

Machine support for link tracing means that all references are equally easy to follow forward to their referent, or backward to their reference;

Ease of creating new reference

Users can grow their own networks, or simply annotate any document with a comment without changing the referenced document.

Information structuring

Both hierarchical and non-hierarchical organizations can be imposed on unstructured information; even multiple hierarchies can organize the same material;

Global views

Browsers provide table-of-contents style style views, supporting easier re-structuring of large or complex documents; global and local views can be mixed effectively;

Customized documents

Text segments can be threaded together in many ways, allowing the same document to serve multiple functions;

Modularity of information

Since the same text segments can be referenced from several places, ideas can be expressed with less overlap and duplication;

Consistency of information

References are embedded in their text, and if the text is moved, even to another document, the link information still provides direct access to the reference;

Task stacking

The user is supported in having several paths of inquiry active and displayed on the screen at the same time, such that any given path can be unwound to the original task; and

Collaborations

Several authors can collaborate, with the document and comments about the document being tightly interwoven.

PROBLEMS WITH HYPERTEXT

The problems with hypertext are :

Disorientation

The problem of having to know (i) where one is in the network and (ii) how to get to some other place that one knows exists in the network. In a traditional linear text documents, one can search for the desired text earlier or later in the text. Hypertext offers more degrees of freedom, more dimensions in which one can move, and hence a greater potential for the user to become disoriented;

Cognitive overhead

The additional effort and concentration necessary to maintain several tasks or trails at one time is the other fundamental problem with using hypertext.

Linking of data is a long process; and

Loss of familiarity with traditional book format.

HYPERTEXT – LINKING TO THE FUTURE

Though the concept and the advantages of hypertext were clear several decades ago, the widespread interest in hypertext has upsurged recently mainly because of cheap and readily available supporting technology. Today, due to the avilability of powerful, high-resolution personal workstations, there is growing interest in the extension of hypertext to the more general concept of hypermedia, in which the elements which are networked together can be text, graphics, digitized speech, audio recording, pictures, and animation. Towards this end, three of the largest ongoing research projects are: Notecards from Xerox PARC, Neptune from Tektronix, and Intermedia from Brown University. There are also commercial hypermedia products, such as Apple's Hypercard and Owl International's Guide.

FURTHER READING

1. Conklin, J. Hypertext - an introduction and survey. *Computer*, 206, 1987; 17-41.
2. Marchionini, G & Shneiderman, B. Finding facts vs browsing knowledge in hypertext systems. Computer, 21, 1988; 70- 80.
3. Rada, R. Writing and reading hypertext - an overview. *J Am Soc Inf Sci,* 40, 1989; 164-71.
4. Smith Karen E. Hypertext - linking to the future. Online, 12, 1988; 32-40.
5. Scacchi, W. On the power of domain-specific hypertext environments. J Am Soc Inf Sci, 40, 1989; 183-91.
6. Irish, PM & Trig, R H. Supporting collaboration in hypermedia: issues and experiences. J Am Soc Inf Sci, 40, 1989; 192-99.

We can be knowledgeable with other men's knowledge, but we cannot be wise with other men's wisdom.

Michel de Montaigue

Knowledge is a steep which few may climb while duty is a path which all may tread.

Lewis Morris