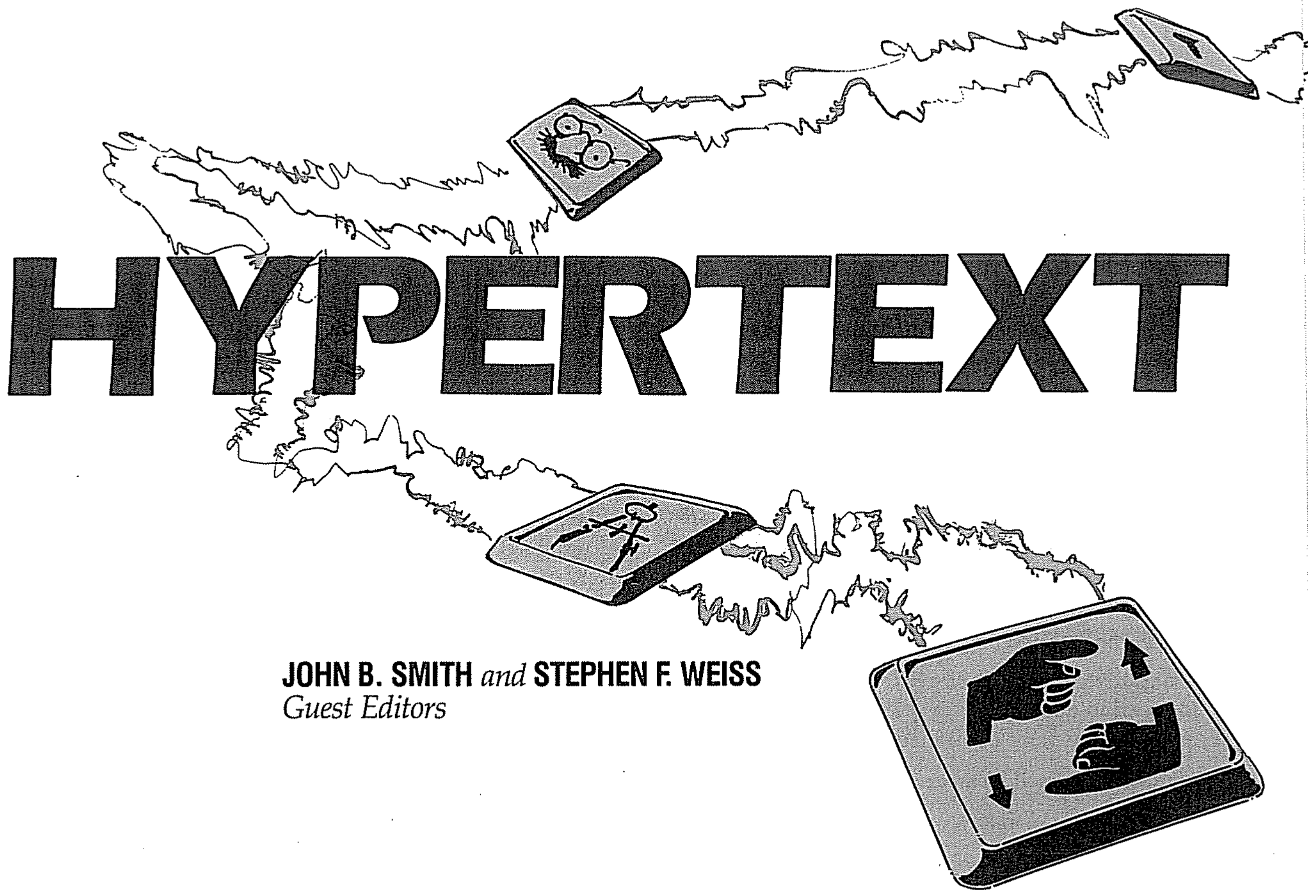

HYPERTEXT

JOHN B. SMITH *and* **STEPHEN F. WEISS**
Guest Editors

This special issue is devoted to a form of electronic document called *hypertext*. More precisely, hypertext is an approach to information management in which data is stored in a network of nodes connected by links. Nodes can contain text, graphics, audio, video, as well as source code or other forms of data. The nodes, and in some systems the network itself, are meant to be viewed through an interactive browser and manipulated through a structure editor.

While the term, hypertext, was coined by Ted Nelson during the 1960s [5], the concept can be traced to Vannevar Bush's 1945 description of "the memex":

> A device in which an individual stores his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory [1].

What distinguished Bush's concept from other forms of data storage was its associative structure that closely modeled the structure of human memory:

> The human mind . . . operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain.
>
> Selection by association, rather than indexing, may yet be mechanized. One cannot hope . . . to equal the speed and flexibility with which the mind follows an associative trail, but it should be possible to beat the mind decisively in regard to the permanence and clarity of the items resurrected from storage.

The first serious attempt to build a memex did not take place until 20 years after Bush's description. In 1968, Doug Engelbart, then at Stanford Research Institute, conducted a dramatic live demonstration of his Augment system at the Fall Joint Computer Conference in which he worked collaboratively on a hypertext document with a colleague 500 miles away [4]. During that session, Engelbart also demonstrated two of his other inventions—the mouse and the chord key set.

In the 20 years since Engelbart's demonstration, both interest and activity in hypertext have grown steadily. Jeff Conklin's excellent survey traces this history in detail [2]. Some of the more important milestones include the following:

- **ZOG**, a high-performance system developed at Carnegie-Mellon University and used aboard the USS *Carl Vinson*. ZOG is the predecessor of KMS, a commercial system described in the article by Robert Akscyn, Donald McCracken and Elise Yoder that begins on page 820.

- **Intermedia**, and several earlier systems, developed by a research group at Brown University that traces its ancestry to Nelson [8].
- **NoteCards**, the most ambitious system of the past decade, developed at Xerox PARC. NoteCards is described in detail in the article by Frank Halasz beginning on page 836.
- **Document Examiner**, a beautifully engineered high-performance system by Symbolics that provides on-line access to their user documentation [7].
- **Neptune**, a hypertext system for computer assisted software engineering, developed at Tektronix [3].
- **WE**, a hypertext authoring system developed at the University of North Carolina that produces conventional paper as well as electronic documents and closely models human cognitive processes [6].

During the past year interest in hypertext has accelerated sharply. No one factor explains it. More powerful workstations, high-resolution graphic displays, increased network communications, and decreased costs for large on-line storage all contributed.

Two specific events, however, seem to have played a particularly strong role. The first event was Apple's introduction of HyperCard. While this system is relatively primitive compared with some earlier ones, Apple's aggressive promotion of it is changing hypertext from an esoteric concept known to a few hundred people to a household staple of computing being used by millions.

The second event was Hypertext '87, the first major conference devoted entirely to hypertext. This workshop brought together an extremely broad spectrum of people. It drew participants from five continents. The meeting included computer scientists from a number of different areas, among them information retrieval, text and image handling, software engineering, VLSI design, graphics, and human-computer interaction. Academics came from a number of additional disciplines, including classics, philosophy, psychology, English, foreign languages, religious studies, and medicine. A number of commercial interests were also represented, including system developers, on-line information vendors, and people planning to develop hypertext databases in specific subject areas. And there were the users, those wanting to use hypertext to deliver medical information, train writers, and manage software development. This meeting of individuals from such a broad spectrum of backgrounds and interests has generated considerable momentum that is appearing in many different forms: hypertext meetings in specialized areas, new commercial ventures such as hypertext publishing companies, and a number of publications on the topic, including this special issue.

To understand why hypertext is attracting such attention, one must understand how a hypertext "document" differs from a conventional paper document.

In most conventional paper documents—such as journal articles, specifications, or novels—physical structure and logical structure are closely related. Physically, the document is a long linear sequence of words that has been divided into lines and pages for convenience. Logically, the document is also linear: words are combined to form sentences, sentences to form paragraphs, paragraphs to form sections, etc. If the document has a hierarchical logical structure, as do many expository documents such as journal articles, that hierarchy is presented linearly: the abstract or overview of the whole comes first, followed by the introduction, the first section, the second section, etc., until the conclusion. This linearity is easy to see if one imagines the hierarchical structure represented as an outline, with the sections of the document appearing in the same order as they normally do in the outline. Such documents strongly encourage readers to read them linearly, from beginning to end following the same sequence.

A few conventional paper documents—such as encyclopedias, dictionaries, and other reference works—separate logical structure from physical structure. Physically, these documents are linear sequences of independent units, such as articles on specific topics or entries for individual words. Logically, they are more complex. The reader seldom reads such documents from beginning to end, but rather searches them to locate the article or entry of interest (a form of random access), and then reads that portion sequentially. However, the reader is likely to encounter various cross references to other entries while reading as well as a list of "see also's" at the end of an article. To follow those pointers, the reader must locate the appropriate volume, find the appropriate entry, and then the relevant portion. The logical structure of reference and other similar documents is, thus, more complex. They have a sequential structure that aids search, but the logical path of the reader is a network that can crisscross the entire document or set of documents, from one item to another, to another, etc. Such documents are more flexible but they are also cumbersome, particularly when they appear in large, multivolume formats.

Hypertext electronic documents provide most of the flexibility of reference works as well as add a number of new features.

Hypertext electronic documents provide most of the flexibility of reference works as well as add a number of new features. Earlier, we described a hypertext as a document in which information is stored in nodes connected by links. Each node can be thought of as analogous to a short section of an encyclopedia article or perhaps a graphic image with a brief explanation. The links join these sections to one another to form the article as a whole and the articles to form the encyclopedia. These links are usually shown for each node as a "from" link pointing to the node just read and a set of "to" links that indicate the (usual) multiple nodes which one may select to read next.

SPECIAL ISSUE

Many systems also include pointers embedded in the text itself that link a specific portion to some other node or portion of text. Thus, one moves from node to node by selecting the desired "to" link, an embedded cross-reference link, or the "from" link to return to the previous node. For many documents, the "to" links can be thought of as organizational. Collectively, they frequently form a hierarchical structure analogous to the hierarchical logical structure of many conventional documents. From this perspective, the embedded, cross-reference links cross the main organizational structure.

While we can establish a rough analogy between the two, hypertext documents are much more flexible than conventional documents. For example, one can read the hypertext article just as one reads the conventional paper article by first reading the overview node, then the first section node(s), the second section, etc. However, one can also read the sections in different orders. For example, if the hierarchical structure of the article is viewed as a two-dimensional tree or organization chart instead of as a linear outline, one can easily imagine that if the "to" links were shown as the children of the current node, selecting the second section before the first or perhaps skipping the first entirely. Hypertext documents are also much more convenient. To follow the cross-references in a modern encyclopedia often means moving among thirty of more (heavy) volumes. Readers do it, but it is a slow, frequently laborious, task.

While conventional publications are limited to text and graphics, hypertext nodes offer sound, video sequences, animation, even computer programs that begin running when the nodes in which they are stored are selected.

While hypertext provides greater flexibility and convenience than conventional documents, its power and appeal increase dramatically when it is implemented in computing environments that include networked microcomputers and workstations, high-resolution displays, and large on-line storage. While following a cross-reference in a 30-volume encyclopedia can take several minutes, many hypertext systems can deliver the next node in less than a second and from a much larger body of information that might take thousands of volumes in print. While conventional publications are limited to text and graphics, hypertext nodes offer sound, video sequences, animation, even computer programs that begin running when the nodes in which they are stored are selected. While the organizational and cross-reference structures of conventional documents are fixed at the time of printing, hypertext links and nodes can be changed dynamically. Information in individual nodes can be updated, new nodes can be linked into the overall hypertext structure, and new links added to show new relationships. In some systems, users can add their own links to form new organizational structures, creating new documents from old.

We hope the reader, beginning with a narrower but deeper understanding of hypertext, will be well prepared to read further in the literature, consider using hypertext for a specific application, or think about the many issues raised by this technology.

Each of these changes represents an incremental difference between hypertext and conventional documents, but when considered together, they are producing a qualitative change in the way some people are conceptualizing information resources. It is this shift in perspective that is creating such excitement and such a wealth of new possibilities in the minds of some. This special issue was assembled with the goal of providing readers unfamiliar with hypertext with sufficient information so that they can see the potential for hypertext in their own fields and, perhaps, share this intellectual ferment.

The six articles included in this special issue were all presented at Hypertext '87 held at the University of North Carolina at Chapel Hill last November. In selecting them, we have not attempted to give an overview of work in the field—again, we refer the reader to [2] for this. Instead, we have chosen articles that go deeply into the characteristics of a particular hypertext system, attempt to generalize major concepts that define hypertext, or describe in detail a particular application. We hope the reader, beginning with a narrower but deeper understanding of hypertext, will be well prepared to read further in the literature, consider using hypertext for a specific application, or think about the many issues raised by this technology.

The first two articles describe specific hypertext systems. However, they extend their discussions to a number of design issues that apply to all current and future systems. Akscyn, McCracken, and Yoder describe KMS, a large system that is commercially available. In doing so, they examine the 26 key design decisions they made. This discussion not only provides a detailed view of KMS and its rationale but a very general view of the design space in which all hypertext systems fit. In the second article Halasz describes NoteCards, the system he helped to develop at Xerox PARC. In considering possible extensions to NoteCards, Halasz identifies seven key problems that currently limit hypertext systems but which also provide goals for future research.

The articles by Brad Campbell and Joseph Goodman and by Pankaj Garg attempt to generalize key concepts inherent in hypertext. Campbell and Goodman describe the Hypertext Abstract Machine (HAM) developed at Tektronix. This system is a back-end hypertext server that supports operations on directed graphs with typed nodes and links. Consequently, it can be used to support a number of different interfaces and applications. Garg offers a formal mathematical description of hypertext. This work could link hypertext systems to other areas of research, such as relational databases and formal search languages, as well as lead to even more general tools for building hypertext systems.

The last pair of articles consider two applications. Mark Frisse describes work underway at Washington University in St. Louis to develop a hypertext system for delivering medical information. Darrell Raymond and Frank Tompa describe a project at the University of Waterloo to build a hypertext version of the Oxford English Dictionary.

During the conference, these six articles were among eight that provoked considerable discussion in terms of issues raised by hypertext. We overheard much discussion of legal issues. How can copyright issues be resolved to permit development of large hypertext databases that can be searched, relevant material found, and reused in new documents? How can intellectual ownership be traced in such an environment? Others were concerned with social issues. As networks link more and more individuals and hypertext databases, will we create a further division between the included and the excluded? Most of this discussion, however, occurred outside the formal sessions. However, we hope this special issue will stimulate similar discussion among readers and their colleagues.

John B. Smith
Stephen F. Weiss
Guest Editors
The University of North Carolina
Department of Computer Science
Chapel Hill, N.C.

REFERENCES

1. Bush, V. As we may think. *Atlantic Monthly 176*, 1 (July 1945), pp. 101–108.
2. Conklin, E.J. Hypertext: An introduction and survey. *IEEE Computer* 2, 9 (Sept. 1987), 17–41.
3. Delisle, N. and Schwartz, M. Neptune: A hypertext system for CAD applications. In *Proceedings of ACM SIGMOD International Conference on Management of Data* (Washington, D.C., May 1986), ACM, New York, 132–143.
4. Engelbart, D.C., and English, W.K. A research center for augmenting human intellect. In *Proceedings of the 1968 Fall Joint Computer Conference* (Montvale, N.J., Fall 1968), AFIPS Press, 395–410.
5. Nelson, T.H. Getting it out of our system. In *Information Retrieval: A Critical Review*. G. Schechter, ed. Thompson Books, Washington, D.C., 1967, 191–210.
6. Smith, J.B. WE: A writing environment for professionals. Tech. Rep. 86-025, Department of Computer Science, University of North Carolina at Chapel Hill, August 1986.
7. Walker J.H. The Document Examiner. *SIGGRAPH Video Review*, Edited Compilation from CHI'85: Human Factors in Computing Systems 1985.
8. Yankelovich, N., Meyrowitz N., and van Dam A. Reading and writing the electronic book. *IEEE Computer 18*, 10 (Oct. 1985), 15–30.