TEX & Hypertext — The Future of Electronic Publishing?

Christine Detig
Kranichweg 1, D-6074 Rödermark, FR Germany
Bitnet: xiticdet@ddathd21

Abstract

The usage of computers allows writing and publishing documents electronically. This article is oriented towards the support of authors.

Hypertext is a popular method of electronic document manipulation. It supports authoring by representing a document as a net of linked text fragments in which the structure can be visualized and manipulated. Publishing a hypertext means no longer printing. Instead, the document remains in the computer and is read electronically.

Can hypertext concepts be used for traditional electronic publishing, i.e., for the production of printed documents? And is it possible to integrate them in an environment that uses typesetting systems like TEX for the electronic authoring and for the printing? These questions are discussed.

Introduction

Talking about the rôle of TEX in the process of electronic publishing should reflect that TEX is only a tool for solving part of the problems that arise. We should not enlarge it to handle everything, but instead, integrate it in a set of adequate tools. This article will reflect on the usability of TEX in such a publication environment and under what circumstances it can be used. It shall be a contribution to the development of TEX towards its real power. Otherwise, TEX will be buried under all the other "easy-to-use" text processors.

Electronic publishing, the interwoven process of editing, viewing and revising that results in a printed version of the document, can be separated into electronic authoring and electronic printing. The preparation of a certain group of documents demands more from both parts than an editor and a formatting system. These documents can be classified as being large, long-living and determined to be under regular revision. They may exist in different versions and in several variations. It may also be that several authors work on them. Examples are manuals, lexica, tutorials, etc. They are best handled with a special authoring system that supports long-time management and flexible manipulation.

There exists a technique of document preparation, *hypertext*, that offers support for the interactive authoring of structured documents. In the following, a short overview about hypertext and some typical applications will be given. For more details, see Conklin [1987] or Meyrowitz et al. [1985]. Following the overview, the usability of the hypertext technique in an authoring environment for printed documents in connection with TEX shall be discussed.

What is Hypertext?

In addition to the "classic" electronic typesetting methods, there are other ways to use a computer for electronic document preparation. One of them, quite popular now, is hypertext. Hypertext is a technique that supports electronic authoring by offering an interactive environment that allows one to visualize and manipulate the structure of a document, which is built up from text *fragments* that are connected by *links*. Publishing a hypertext means no longer using printed paper, but only providing on-line access: The document remains in the computer and is read electronically, usually via a network. For visualization, the fragments are formatted, usually by formatters with graphics support.

Hypertext systems. The first concepts for hypertext were developed by V. Bush in 1945 who designed a machine that should manage documents like text, pictures or notes. It allowed the construction of paths between related information which could be used for browsing. This should allow working with stored information analogous to human thinking.

> "The human mind [...] operates by association. Man cannot hope fully to duplicate this mental process artificially, but he certainly ought to be able to learn from it. 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." (Bush [1945])

Hypertext systems may be classified by their intended usage. To give an overview about existing hypertext systems, I will give a short description of four categories of hypertext applications: bibliographic, idea, information and notecard systems.

Bibliographic systems. Bibliographic systems are used to maintain on-line libraries with various documents. They offer a simple and comfortable interface, not only for reading the documents, but also for working with them. This includes the creation and critical review of texts so that an original version and the reader's modifications may both exist in the system. Finally, this leads to the vanishing of the distinction between reader and writer.

Quite often, the propagators of these systems plead for the infinite computerization of the world and for the abolishing of printing. They do forget the problem of access, which may be easily monopolized and controlled. Financial resources, modern equipment and technical skills would be necessary requirements for taking part in academic (and social) life, resources that are available for only minority of mankind. I do not believe that the social circumstances of new technical systems automatically become democratic and fair (or, at least, free of charge), and any propagation of a single technique as a universal solution lacks realism.

Idea systems. A second group of hypertext systems is designed for the interactive work of one person or several people on a problem. They help collect ideas and organize them so that nothing gets lost and different views on a problem are possible. They may be compared with the notebook of an author or developer that offers possibilities for retrieving and grouping. (Of course, you need your computer next to you all the time in case you get ideas at unusual places.)

Information systems. Information or browsing systems are small bibliographic systems, used for special purposes such as demonstrations, providing for courseware or a collection of information about a special item (e.g., all information about a certain bomb or the time table of a railway station). Usually, they are made for reading only, so that after an initial phase of authoring the content is static. Quite often, they provide only a limited overview about the structure of the document, and you can only follow the keywords for more information. Disorientation is a frequent consequence of this system.

Notecard systems. These systems are not designed for special purposes, but can be used for various applications for which they provide reading, writing, maintaining, and browsing facilities. Systems that are currently popular are NoteCards by Xerox and HyperCard by Apple which are designed as electronic card-index boxes which allow several sorting criteria of the same data.

Concepts of hypertext systems. All hypertext systems have some basic concepts in common. Hypertext documents are non-linear. They offer information that can be dynamically arranged by a desired criteria. They allow the incorporation of ideas in a document and to connect them later in another step so that the author can add and rearrange components freely.

The fragments are the nodes of the graph structure that constitutes the document. They represent a single idea, a specific item. As they are meant to be arranged in various sequences, they should be independent from others, except by connections that can be represented by links. They may be named, can have attributes, or be of different types.

The system supports the management of the fragments and linking them together. Fragments can be viewed or manipulated (i.e., edited), and the structure is represented and can be browsed. When viewing a fragment, the reader can follow or ignore links to other fragments that are related, or entirely different paths may be selected in the document representation. Links may lead to texts that belong to a keyword, annotations, supplementary information, or successor fragments. They may be named and can have attributes, too.

One can use references that are non-hierarchical (e.g., cross references), but they may as well be hierarchical so that a substructure of related information is built. While the first ones are usually represented by (graphical) markup in the text, the latter ones exist on the management level and are represented in the document browser.

Using a hypertext system. Basic hypertext features are the editing abd viewing of the fragments. If they contain not only text, but also pictures, tables or video, more powerful processors are needed to represent the contents in a formatted form. Some

systems also offer multi media facilities like audio or video components.

Another aspect is the representation of the document structure. Maps of the graph are used in which one can navigate. They show the actual position and sometimes have different levels of detail to reduce the complexity.

Next to viewing and editing, the systems usually offer features for searching and selecting. Many hypertext systems offer query techniques for the retrieval of information. As well, views are possible that filter the whole document according to some criteria. Both are usually done by attributes of nodes and links, but there also exist content retrieval systems.

Another feature is the possibility of versioning. This allows one to maintain different versions of text or to keep modification logs. If the system supports access control and offers synchronizing mechanisms, co-authoring is possible as well.

For authors, hypertext offers support for design, structuring and presentation. Nevertheless, the necessary actions for adding new parts may also lead to diversion, and provisional structures may result that are not revised later. In general, human association is not explicit, but subconscious. It is not only determined by the problem's inherent logic, but also by aspects like social environment. Hypertext forces authors to build associations explicitly. This may be quite difficult and irritating (just like learning the first programming language) because a way of thinking must be practiced that is not "natural" and demands a higher degree of detail than the human is used to.

Readers (the author is also a reader) may browse through the text in any variation and with any level of detail they like, supported by tools for navigation. But that also means making a lot of decisions in a short period of time, not knowing whether they are sensible, because no "glimpse" is possible. Following a keyword or selecting a node only by its name may lead to many impasses. Following several paths at a time is quite stressful, too.

The visualization may be helpful for orientation, but in a heavily linked structure it will be difficult to find an adequate and clear representation. Clipping or reducing the complexity will be necessary which destroys information that might be needed.

Hyper documents. The above explanations show that hypertext is not equally useful for all kinds of documents. Hypertext is developed for non-linear documents, such as a notebook, a reference, or a lexicon. These hyper documents consist of information fragments that are more or less independent from each other, having a relation that is marked by a link.[2] The resulting document is dynamic. The reader builds his own text by finding his individual way through a database of interconnected information.

This means that hypertext systems are information systems with an interactive environment that supports structured authoring and dynamic reading of the available material. On the one hand, the reader no longer depends on structures designed by the author. He may freely find the structural representation and the parts of information he is looking for.

On the other hand, the author will group and structure the information in a representation he likes best, though he knows that the reader may look at the hyper document in a totally different way. He can try to avoid this by creating fixed structures that are absolutely necessary for comprehension.

The consequence is that hyper documents may offer a very different degree of freedom to the reader: The more the author fixes structures in it, the less variability is left for the reader, until we end up with totally fixed systems in which the paths cannot be modified, or the information is hidden from the reader (depending on his classification). These methods are often used in tutorial systems.

Hyper database systems. Next to applications that are based on documents, hypertext is being used more and more as the interactive interface for nonstandard database systems. It serves for retrieval, expert system requests, and the like. While there exist facilities for the transformation of normal database output into TEX to produce lists of the selected objects, the problem of publishing the contents of a non-standard database is yet unsolved because the mapping of objects to markup is not trivial. As more and more information of all kinds will be managed in databases because they support multi-access, versioning and consistency, this will be an important request of the future.

Hypertext and the Traditional Publishing

As we have seen, hypertext is typically used for a specific kind of document that is different from standard text. Nevertheless, the question remains: Can hypertext concepts be used for traditional electronic publishing, i.e., for the production of printed documents? And is it possible to integrate them in

[1] That this is not trivial is described by Raymond and Tompa [1988].

an environment that uses typesetting systems like TEX during the electronic authoring as well as for printing?

Hypertext and printing. We do not want to discuss here the usage of TEX as a formatter for a hyperbased on-line information system. Printing means that one plans to reach a widespread readership. Hypertext limits the potential reader to a small group with connections to the computer system that holds the text. Presently, interactive access over wide area networks which work with a lot of documents is not without problems. And reading from the screen is not always what we want, even if access and all technical problems are solved.

Hypertext documents are not made for printing. Of course, fragments may be extracted and printed on paper, but the non-linear structure of the whole document would force one to choose a mapping on a sequence that will be inadequate in many cases. The sequence of reading can no longer be chosen interactively, and a visualization of the links must be established so as not to destroy the structure. Contrary to this, a printed document usually has a linear structure and the author intends a sequence of reading. This sequence influences the contents of the text components because they may depend on information given before.

Database demands. New demands arise from the request of publishing the contents of non-standard databases. Hypertext features with their interactive concept are not sufficient. Paper versions are still needed, e.g., for the management. A markup is not enough either. It supports only the printing aspect and not the authoring. An integration of both markup and hypertext will lead towards maintainable systems for both the interactive maintenance of the contents and the publishing of it.

Authoring with hypertext. Now, let's take a closer look at the authoring aspect of hypertext. There are some good reasons to use hypertext for the authoring of printed documents. Hypertext allows the management of structured documents that are variable in both structure and content in a way that adds a new quality. It is this support of variability that makes it much more powerful than a structure editor.

The interactive environment is very helpful for the maintenance of a document, especially for those that are long-living and are intended to be modified and revised again. Annotations can be added easily, and versioning is usually supported or can be introduced by the author himself. Structural elements such as glossaries, footnotes or an index can be mapped to the hypertext concept. The problem of finding an adequate representation of the text structure in the hyper document will no longer be severe, because there are no longer readers of this document form. Problems are also reduced by the possibility of structure visalization. Navigation helps to locate points of interest that the author, who is more familiar with his text than a reader would be, may identify more easily. This is especially true because the named fragments offer better identification criteria than files with naming restrictions.

Hypertext is very useful for the authoring of a specific kind of document, that is, those that exist in different variations but are maintained and updated together. They are 'hyper-texts' with a non-linear structure and may contain free nodes, but the resulting printed document will be linear.

An example is standards in different versions with history and rationals. Hypertext allows one to realize such a document as a base of text fragments through which different paths exist with common parts. This considerably reduces updating problems, and the reorganization of the document structure can be done easily and coordinated by just modifying the paths. Another example is manuals of the same software for different computers, which leads to very similar documents, yet with different components (sometimes naming conventions only). For this application — maintaining a non-linear document that consists of several linear, printable documents — we developed the prototype of an authoring system with hypertext concepts at the Technische Hochschule Darmstadt.

An authoring system for printed documents that uses hypertext concepts needs some modifications and additions to standard hypertext features. The interactive graphical user interface will be used, but it must support the extensions and adaptions to the special requests of linear text. First, the whole document must be accessible, not only single fragments. Viewing the whole document is as important as printing it. But fragmentation offers more flexibility on this point. Features can be added for selecting parts of documents which can be viewed, selected, browsed, and printed. It must be assured that the selected fragments can be transformed to an acyclic, connected graph with one root.

Specific problems with linear texts in a hypertext based environment arise from the interdependence of the fragments in a traditional document. The possibility of rearrangements is restricted, but the restrictions can be specified in the hyper document by using special fragments for annotations or

by using attributes that lay them down. And the feature of (hierarchical) structuring and composition may be used to build components that belong together and should not be rearranged except with modifications in the contents.

TEX and Interactive Authoring Systems

Previously we talked about printing without thinking about the realization. Of course, what we want is to use TEX for formatting so that we can view and print documents (and parts of them) by using `dvi` drivers. To use TEX in an interactive environment, some reflections about the demands of TEX must be made.

Requirements. To use the TEX and driver facilities, the authoring system should contain features of a TEX menu so that the printing and viewing process can be parametrized, or search paths can be specified. This should be designed with the same "look-and-feel" philosophy as the rest of the interactive environment. This requires either a special TEX authoring system or one that can be enlarged by the formatting components.

TEX is a batch processor which requires complete documents for processing. Complete means that the grouping levels are balanced and that mode switches are reset again. All initializations like the selection of formats and style options or user defined macros must be available, and the termination sequences are needed, too.

Extensions. For the authoring system described before, that means that the viewing and printing mechanism must add special user defined pre- and post-information into a document for TEX to handle. This document must be extracted to a file as a whole so that TEX can process it.

To allow viewing and printing of any fragment (or document), the restriction of balanced TEX usage in fragments must be forced (probably by a structure editor), or composed structures must be built with components that belong together and cannot be handled separately. This is very difficult to control because of the various group symbols and the possibility of grouping via macros.

Conclusion

TEX was designed as a typesetting machine. The last section discussed the integration of TEX in an authoring system. It is this aspect of TEX which has the consequence that TEX and hypertext is *not* the future of electronic publishing.

TEX is both — optical and logical markup. The fragmentation and structuring of a document is incompatible with optical markup. Fine tuning of the output within the hypertext system is impossible. Problems will arise with error location and handling. Syntax error free TEX input from humans cannot be assured because a structure editor would be necessary that is a complete TEX interpreter to deal with the user changeable lexical analysis and the dynamic binding of token.

In a hypertext system, the user should not input original TEX. Instead, we should encourage the separation of optical and logical markup in TEX by providing markup systems like LATEX (or SGML with a TEX converter) and a variety of styles for the author. Mapping to the optical markup should be provided by the system, leaving the design to professionals.

The separation of author, typesetter and printer has proved to be useful over hundreds of years. Abolishing this splitting of responsibility for contents, form and realization has not led to satisfactory results in the desktop publishing area nor in hypertext systems. Modern technologies allow us to realize these demands as an interwoven electronical process with hypertext and logical markup on one side and TEX and drivers on the other. Let us fill this possibility with life.

Bibliography

The following articles give an introduction into the hypertext technology, present existing hypertext systems and discuss the advantages and problems of their usage.

Bush, V. "As we may think." *Atlantic Monthly* 176, pages 101 – 108, 1945. reprinted in: CDROM: The New Papyrus. Microsoft Press, 1986.

Conklin, E.J. "Hypertext: An Introduction and Survey." *IEEE Computer* 20(9), pages 17 – 41, 1987.

Meyrowitz, N., A. van Dam, N. Yankelovich. "Reading and Writing in the Electronic Book." *IEEE Computer* 18(10), pages 15 – 30, 1985.

Raymond, Darrel R., Frank Wm. Tompa. "Hypertext and the New Oxford Dictionary." *Communications of the ACM* 31(7) (Special Issue Hypertext), pages 871 – 879, 1988.