

# "Hypertext Engineering": Practical Methods for Creating A Compact Disc Encyclopedia

Robert J. Glushko  
Mark D. Weaver  
Thomas A. Coonan  
*Search Technology, Inc.*  
Norcross, GA

Janet E. Lincoln  
*Stuyvesant, NY*

## INTRODUCTION

The *Engineering Data Compendium* [1] is a multi-volume scientific and engineering encyclopedia that contains comprehensive reference data about human perception and performance. Its primary users are designers of systems with complex user interfaces, such as aircraft cockpits or power plant control rooms. We created a compact disc version of the *Compendium* in which hypertext features enhance and supplement the Table of Contents, Index, cross-references, and other parts of the printed *Compendium* as methods for finding information.

When we began this project in late 1987, we sought help by reviewing existing products and published literature. We hoped to find case studies and "lessons learned" reports from previous efforts to create hypertext encyclopedias. Instead, we found research on creating new documents with hypertext structures or so-called "proof of concept" efforts with small sets of notecard-sized documents. (We highly recommend [2] and [3]). Compact disc encyclopedias that contained realistic amounts of text to serve as examples for our efforts typically did not contain any figures or graphics. In sum, we could find little work that was directly relevant to the scale or scope of the problems we faced with a real encyclopedia.

Since no established methodology existed for creating an electronic encyclopedia from a printed one, we set out to develop our own approach to "hypertext engineering." Even though we have considerable experience in user interface design, information retrieval, on-line documentation, software engineering, and other relevant domains,

we have found significant and unexpected challenges in the application of the rapidly evolving hypertext concept to the new technology of compact discs.

In this paper we describe the design tradeoffs and pragmatic concerns involved in translating text into hypertext for practical applications with real-world constraints. One such constraint is the need for our compact disc encyclopedia to be used on industry-standard personal computers with small display screens.

We describe the relationship between the logical structure of the encyclopedia and its physical structure in printed form. Understanding this relationship is a prerequisite for defining the size and structure of the text and graphical units and deciding which hypertext links to create or not to create.

We begin with a detailed description of the the encyclopedia. The first principle of hypertext engineering is that without a comprehensive understanding of the logical and physical structure of the document, a project of this scale is impossible.

## THE ENGINEERING DATA COMPENDIUM

The *Engineering Data Compendium* is a highly specialized scientific and engineering encyclopedia containing nearly 3000 pages of technical information on human perception and performance. It provides comprehensive data on the capabilities and limitations of the human operator that can help design engineers achieve a better match between system specifications and operator characteristics. The *Compendium* contains basic and parametric data on important aspects of human sensation, perception, and performance, tabular summaries of data from related research, predictive models, and perceptual and performance principles. Also included are topic area introductions and other background information to aid users in understanding and evaluating the data.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

1.1 Measurement of Light

1.103 Range of Light Intensities Confronting the Eye

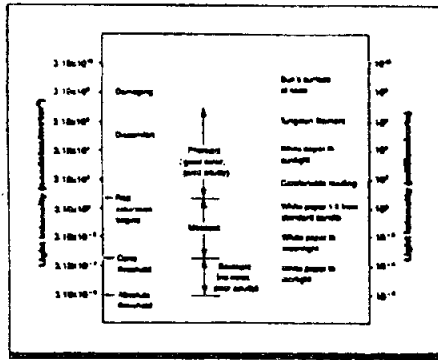


Figure 1. The range of light intensities that confront the human eye. (Adapted from C. R. Brubaker [196], Vision and Visual Perception. Copyright © 1966 by John Wiley & Sons, Inc. Reprinted with permission.)

Key Terms

Illuminance level; luminance;oscopic vision; photopic vision; scotopic vision

General Description

The human eye is sensitive to a wide range of light intensities, from a minimum visible level of  $\sim 0.00003 \text{ cd/m}^2$  to an upper tolerance limit of over  $300,000 \text{ cd/m}^2$ . Vision at very low levels of illumination (e.g., starlight) is termed scotopic vision and is mediated by the rods; visual acuity is poor with scotopic vision and an sensation of color (hue) occurs. Vision at high intensity levels (e.g., daylight) is known as photopic vision and is mediated by the cones; photopic vision is characterized by high visual acuity and

the perception of color. Mesopic (mixed) vision (mediated by both rods and cones) occurs with intermediate light intensities (e.g., moonlight). Figure 2 shows how outdoor brightness decreases during twilight. Dark adaptation of the eye with decreasing illuminance is at least as rapid as the normal decline in ambient illumination at evening. Figure 3 shows how the luminance of a test patch changes with the angular elevation of the sun above the horizon.

Constraints

- Sensitivity to light depends on the eye's state of adaptation. Maximum scotopic sensitivity requires  $\sim 1 \text{ hr}$  of dark adaptation even after as little as a few minutes' exposure to photopic light levels. The time course of light adaptation is similar for rods and cones and is much faster than dark adaptation, requiring only a few minutes' exposure at a high luminance level.

WILKINSON, J. L. Engineering Data Compendium, Volume 1, Photometry and Radiometry. McGraw-Hill, New York, 1980.

Visual Acquisition of Information 1.2

Key References

1. Barck-Gross, J. O., Conrad, L., & Maud, L. C. (1964). Daylight viewing of plates for night flying. *Transactions of Experimental Aviation*, 3, 9-29.
2. Green, C. H. (Ed.) (1963). *Visual Acuity*. New York: Wiley.

Cross References

1. 181 Range of visible energy in the electromagnetic radiation spectrum.

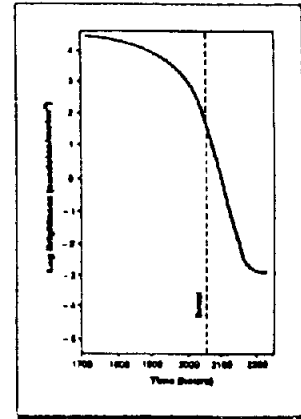


Figure 2. The decrease in brightness from daylight (cd/m²) to darkness (candlepower) on July 14, 1948 (adapted from Ref. 2). Sample of test patch to reference to test not shown.

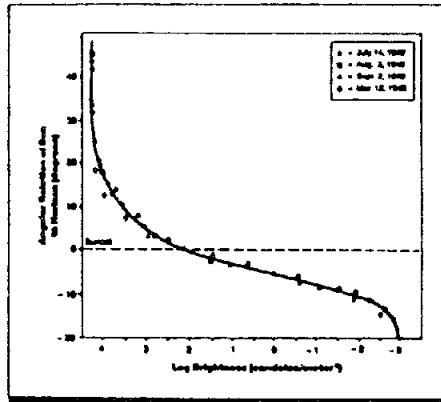


Figure 3. The relationship between the brightness of a calibration patch in a vertical position and the angular elevation of the sun relative to the horizon on four days during different seasons of the year. (From Ref. 1)

Figure 1. Typical entry from the *Engineering Data Compendium*, reduced from two facing pages in the printed version. Entries are highly structured and all text is contained under all or most of ten sub-headings (e.g., Key Terms, General Description, Constraints, etc. in this example). Up to six figures, tables, or illustrations are embedded in the text. Note that Figures 3 and 4 show this same entry as it is displayed from the compact disc.

The *Compendium* originated with the observation that designers of complex systems fail to exploit much potentially useful information about human perception and performance. A variety of "cross-disciplinary chokepoints" prevents designers from identifying, finding, and applying information. Among the most important barriers are the volume of information, its fragmentation across disciplines, and the conventions by which research results are packaged for scientific communication [4].

To overcome these problems, the *Compendium* introduces a novel format for presenting technical information that makes it usable by designers with little knowledge or experience in the subject area:

1) Information is segmented into small, self-contained entries that deal with narrow, well-defined topics. In the printed *Compendium*, entries typically appear as a single spread of two 8 1/2 x 11 inch facing pages. Figure 1 shows a typical *Compendium* entry (reduced). Note that this is the same entry used in Figures 3 and 4.

2) Each entry is centered around one or more figures, tables, or illustrations (about 2000 in all) that depict the key findings.

3) Entries are highly structured and typically contain all or most of the standard text parts listed below. This modular format allows users to locate the information they need rapidly and reliably, and supports a diversity of users requiring different types of information and levels of detail.

- Title
- Key Terms
- General Description
- Applications
- Methods
- Experimental Results
- Empirical Validation
- Constraints
- Key References
- Cross References
- Figure/Table Captions

In its printed form, the 1138 *Compendium* entries occupy three volumes. A fourth volume is the *User's Guide*, which contains the usual aids for helping users find and understand relevant information, as well as some novel ones. The usual aids are the Table of Contents, an Index with over 2000 top-level headings and 10,000 items in all, including thousands of cross-references, and an extensive Glossary.

The novel user aids are a composite index of Key Terms contained in the entries, a Design Checklist, and a set of Design Questions. The latter two, taken together, form a "application-oriented" hierarchical view of the *Compendium* that complements the "scientific" taxonomy of the Table of Contents. At the lowest level of this design-centered perspective on the *Compendium* are questions that a designer might ask about human perception and performance. These questions are indexed to the entry that answers them.

### THE COMPENDIUM ON COMPACT DISC

The compact disc contains all of the text and graphics from the printed *Compendium*. It can enhance the use of the printed version or can be employed by itself. The user interface for the disc was designed to exploit the careful organization and aids for finding information already designed into the printed version, augmented by hypertext and other features that became possible when the text and figures were converted from their printed form.

### Screen Design

Functions and displays are arranged in consistent and meaningful positions on the screen. The important areas of the screen are labeled in Figure 2 and described below.

- \* **MAIN MENU** -- Always on the screen to show functions that are always available.
- \* **CONTEXT LINE** -- Describes the user's current context; this context may be the name of the main menu function currently being invoked, the title of the *Compendium* entry being viewed, and so on.
- \* **HELP LINE** -- Contains a brief help message advising the user of expected actions or options.
- \* **VIEWER WINDOW** -- The workspace in which most program activity takes place, including browsing, searching, and display of entries. The scrollbar can be used to position the text within the window.
- \* **SIDEBAR** -- Presents various "function buttons" and information areas that are used by different functions on the disc. The sidebar area is context-sensitive; different sidebar displays accompany each function.

- \* **BOTTOM PANEL** -- Like the SIDEBAR, a context-sensitive area for "function buttons" and displays.

### Browsing for Entries

The user interface emphasizes browsing as a means of finding relevant entries. Browsing the Table of Contents, the Index, and other access aids is far more effective than Boolean search in teaching users about the rich structure of the *Compendium*. Browsing is akin to a floodlight that illuminates the entire document, in contrast to searching, which is like a flashlight that shines brightly but only on a small part.

Choosing BROWSE from the main menu lists four hierarchical databases that point to *Compendium* entries.

Table of Contents  
Index  
Design Checklist and Questions  
References

The first three access methods existed as part of the printed *Compendium*; the fourth was created by collecting and "inverting" the Key References contained in the entries after we determined that many references are cited by more than one entry.

We designed a hierarchical "outline browser". The initial display in the viewer window of each of the hierarchical databases contains only the highest level of the hierarchy. For example, the initial display of the Table of Contents lists only the 12 major topic areas into which

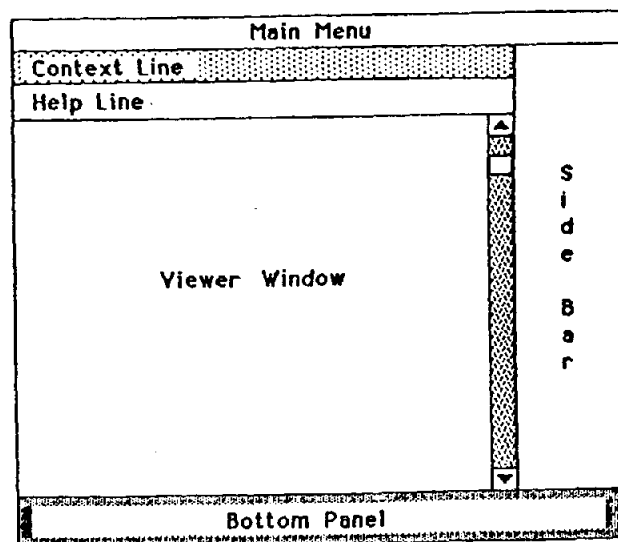


Figure 2. Schematic view of the user interface for the compact disc *Compendium*. Figures 3 and 4 depict the screen when text and figure portions of the entry shown in Figure 1 are displayed from the compact disc.

the 1100 entries are arranged, not the sections and sub-sections. Selecting an item expands it to show the levels below it in the hierarchy; at the same time, details in other parts of the hierarchy are hidden to focus the user's attention on the most recently selected item. If the item is an entry number, expanding it displays the entry in the viewer window.

To aid users when browsing, the bottom panel holds "function buttons" that enable them to find a word in the access database being browsed. The FIND function is especially useful when the information is in alphabetical order (the Index, Reference List, and Glossary) because it lets the user start browsing somewhere besides the beginning. If the FIND pattern is found, other functions in the bottom panel let the user locate the NEXT, PREVIOUS, or any specified occurrence.

Our browser design emerged from studies of several different browsers to identify the functions that they provide and to compare different methods for implementing them (these included *Thinktank* [5], *Document Examiner* [6, 7, 8], *NaviText SAM* [9, 10], *Guide* [11, 12], and *KnowledgeSet* [13]). We identified seven distinct functions in these hierarchical browsers:

1. Representing "where you are" in the hierarchical structure of the database being browsed,
2. Moving around in the hierarchical structure of the database being browsed,
  - A. Is there semantic support (e.g., are there distinctions between "move to the next item at the same level in the tree", "move to parent", "move to child", etc.)?
  - B. Is there purely syntactic support (e.g., "move cursor")?
  - C. Is there a way to specify a point in the database being browsed and search for it?
3. Expanding an item to reveal its next lower level of structure,
4. Closing an expanded item to hide lower levels of structure,
5. Representing whether an item in the database can be expanded or not,
6. Representing whether an item in the database is currently expanded or not, and
7. Designating the last item expanded.

We can not present our complete analysis here. However, it is worth noting that none of the browsers we studied supported all seven functions, and any two usually differed in their implementations of three or more of these functions.

### Searching for Entries

Most compact discs emphasize Boolean search of the full text as the primary method of finding relevant information [14]. However, Boolean searching does not easily discriminate among small and relatively homogeneous

documents like *Compendium* entries that contain many engineering and scientific terms with many distinct senses (e.g., accommodation, adaptation, threshold) [15]. For users with little knowledge of the information the *Compendium* contains, Boolean searches with simple queries yield too many entries to be useful including many that are irrelevant. Since the extensive (human-generated) Index distinguishes different senses of the same word, it can be more effective to browse the Index than to search the computer-generated inverted index from the full text.

Nevertheless, our compact disc supports full Boolean searching of the full text of the *Compendium*. Choosing the SEARCH menu item brings a set of "buttons" into the sidebar that enables the user to specify a query, select the parts of entries to be searched, display a list of entries that match the query, or select from a "history list" of queries to define a more complex query.

The regular structure of each *Compendium* entry allowed us to treat each of the ten text parts they typically contain as separately-indexed fields, which enables users to improve the precision of searches by limiting them to particular parts of entries. By default, users search only the Title, Key Terms, General Description, Applications, and Captions parts since these are most likely to contain information relevant to designers. Of course, users can specify any combination of fields as the scope of their search.

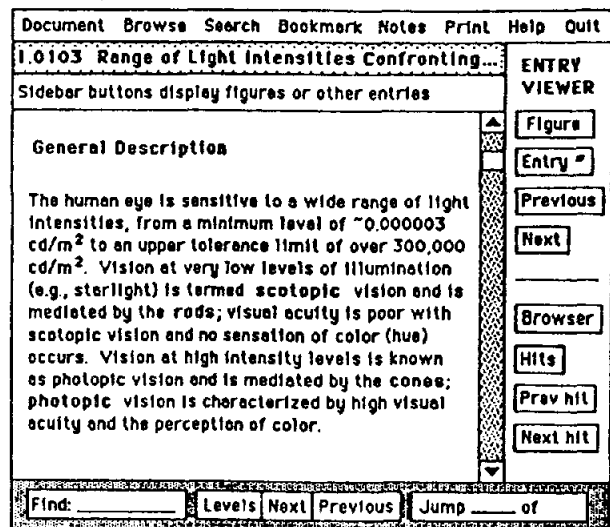


Figure 3. Appearance of the screen when the *Compendium* entry shown in Figure 1 is displayed from the compact disc. Sidebar and bottom panel contain context-sensitive buttons that implement hypertext functions. Boldface items in text are Glossary terms whose definitions can be "popped-up" in window.

## Viewing Entries

Selected entries (via browsing or searching) are loaded into the "viewer window." Entry text is displayed separately from an entry's figures or tables. Ideally, two side-by-side windows would be used, one for the text and one for the graphics. Instead, the size constraints imposed by a 13-inch display monitor make it necessary for the user to view one or the other and switch back and forth. The text portion of each entry is initially presented when an entry is loaded into the viewer. Figure 3 shows how the screen appears when the text portion of an entry is displayed in the viewer window.

Figure 3 also shows the set of navigation and information retrieval functions that are enabled by "buttons" in the sidebar when an entry first appears on the screen. The top button, labeled "Figure", displays a menu of the figures, tables, and illustrations. Selecting one of these figures displays it in place of the entry text, as shown in Figure 4. Captions are shown in a small scrollable window below the figure. Scrolling seems to strike the best balance between the two opposing design objectives of providing maximum display area for the figure and displaying both the figure and caption at the same time.

The other buttons that appear in the sidebar when the text of an entry is in the viewer window enable several other hypertext functions for displaying other entries. Three of these functions are always available:

ENTRY # -- Display any entry specified by number

NEXT -- Display next entry in Table of Contents order

PREVIOUS -- Display previous entry in Table of Contents order.

Four other sidebar functions are enabled if they are appropriate for the context in which the the displayed entry was found. These are:

BROWSER -- Return to the browser context in which the entry being displayed was selected

HITS -- Return to the "hit list" of candidate entries located by a search query

NEXT HIT -- Display the next entry on the search hit list

PREVIOUS HIT -- Display the previous entry on the search hit list.

In addition to these functions that are invoked by explicit "buttons" in the sidebar, several other navigation and information retrieval functions are available through actions on the entry text. Users may display other entries by:

KEY TERMS. Selecting a key term yields a list of titles and numbers of the other entries that share the key term. Choosing one of these candidate entries displays it in the viewer window.

CROSS REFERENCES. This part of each entry contains explicit links to related entries. Selecting one of these cross-referenced entries displays it in the viewer window.

In addition, the first occurrence of each Glossary term

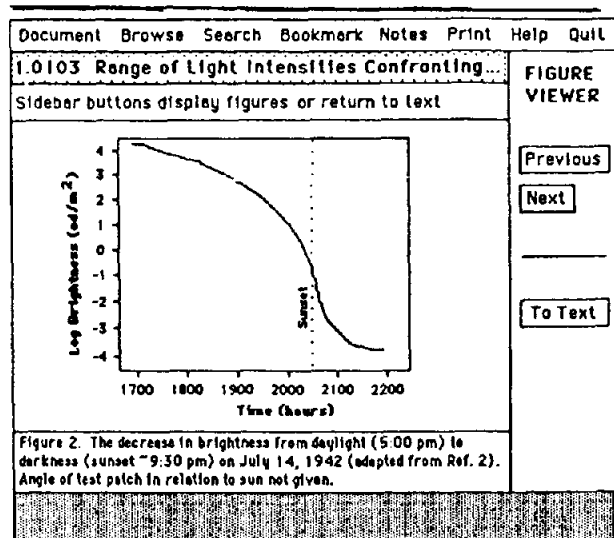


Figure 4. Appearance of the screen when one of the Figures from the the *Compendium* entry shown in Figure 1 is displayed from the compact disc.

within an entry part is highlighted in the text. Selecting one of these terms "pops up" a window that contains the definition.

## Other Functions

Other functions supported by the compact disc *Compendium* include bookmarks for returning to previously viewed entries, annotation, printing, and on-line help.

## IMPLEMENTATION ISSUES

The computing environment most available to our users was the industry-standard microcomputer (PC or AT-class) with a standard 13-inch display screen. These displays generally have graphics capability but not color. This target machine seriously constrains the user interface; experiments with several different hypertext representations of *Compendium* entries (using *Guide*) and various screen designs (using *In-a-Vision*[16] and *Document Examiner*) led us to abandon any desire to reproduce the appearance of the printed page layout.

We implemented the user interface and retrieval software in *Turbo Pascal* [17] using the *GEM* [18] operating environment. We chose *GEM* because it supports reasonable user interface standards for screen appearance, window management, and mouse or cursor control. Additionally, *GEM*'s performance is superior to many alternative DOS application environments, especially for low-end processors.

The complete text of the *Compendium* occupies about 15Mb, with the graphics requiring another 150Mb. Graphics are all bit-mapped, and extensive compression

was needed to enable timely retrieval and display. We cache the graphics in memory when the associated entry text is being displayed. We also cache the text portions of the next and previous entries in Table of Contents order.

The enormous storage capacity of compact discs has greatly expanded our view of computer-based books, even though we have years of experience with information on magnetic disks. Our CD contains 165Mb of text and graphics, we used another 150Mb for various indexes, and the disc is still half-empty. Imagine a printed book whose indexes were ten times larger than the text yet contained as many blank pages as printed ones!

### TRANSFORMING THE COMPENDIUM INTO A HYPERTEXT DOCUMENT

Several characteristics of the *Compendium* make it well-suited for hypertext representation as a compact disc encyclopedia:

- 1) Modular entries are uniquely named and numbered and possess a regular internal structure.
- 2) The entries contain thousands of explicit and implicit cross-references.
- 3) The entries can be arranged in several complementary hierarchies that can be browsed naturally to find relevant entries.

Nevertheless, creating an enhanced hypertext version of this document on a compact disc has presented some major challenges:

- 1) The *Compendium* contains a much larger volume of information than most previous hypertext encyclopedias (e.g., [19]).
- 2) The *Compendium* would not be useful without its figures and other graphics. Other compact disc encyclopedias that include realistic amounts of text to serve as examples do not contain any figures or graphics. For example, *Grolier's Electronic Encyclopedia* [20] seems to retain significant usability even though it includes only the text from the printed version.
- 3) The *Compendium* was developed as a printed document with a hypertext version on a compact disc arising as an afterthought. As a result, little emphasis was placed on version control or configuration management for the source text. Reliable versions of the text were exceedingly hard to come by because the publisher maintained the text in numerous dialects of a proprietary typesetting language.
- 4) The *Compendium* has only recently been printed so we have little experience with its actual use.
- 5) Even though the *Compendium* is well-structured logically, not all of this structure is clearly represented in the printed encyclopedia. Differences between the logical and "physical" structure of the *Compendium* preclude any simple translation from the printed to the compact disc form. In the following section we discuss this issue in more detail.

### Logical vs. Physical structure

The logical structure of a document defines the content and organization of its parts. For example, the *Compendium's* logical structure is as a hierarchical arrangement of entries containing text and graphical parts that are organized and accessed using several complementary schemes. These include the Table of Contents, Index, Design Checklist and Questions, Cross References, and so on. The physical structure reflects how the logical structure is represented within the constraints of the printed book; physical structure in the *Compendium* involves its page layout, typographic conventions such as fonts and line length, and other aspects of its appearance. In creating a compact disc *Compendium*, we identified some key issues concerning the relationship between its logical and physical structure that arise in the design of any computer-based document.

The first issue faced by designers is *How closely will the computer version preserve the physical structure of the document?* Screen appearance is part of this issue. Users might welcome a screen that looks like a book, but only expensive workstations or desktop publishing systems currently have enough screen "real estate" to approximate the printed page. Preserving exactly the appearance of the printed book may not even be desirable. Paper has many permanent advantages over screens, so the challenge is to use the screen space to add features to books that paper cannot provide. Another basic concern is whether page numbers appear in the computer version. For documents with little chapter or section structure, page numbers may be useful indicators of location.

The ease by which a document can be translated from paper to computer depends on a second analysis: *How closely does the physical structure of the document preserve the logical structure of the document?* Building hypertext links from the *Compendium's* cross-references was made easier because references are made using the logical structure of entry numbers, not the physical structure of page numbers.

Nevertheless, *before preserving all aspects of the logical structure of the document, consider the design and implementation tradeoffs involved.* For example, the Index in the printed *Compendium* contains five levels of headings, but only a handful of terms contain more than three. Design and implementation of the data structures used by the hierarchical Index browser would have been considerably simplified if we had used only the first three Index levels, at the cost of only a small loss of precision in retrieving entries.

It is always a challenge to *separate the aspects of physical structure that reflect the logical structure from those that are shaped by stylistic, aesthetic, or entirely arbitrary factors.* Different fonts, spacing, or arrangement on the page may be alternative means to signal the logical structure of the text. Having one of the editors of the *Compendium* as a member of our team saved us lots of time

testing hypotheses about the meaning of different aspects of the physical structure. Sometimes physical structure may obscure the logical structure or may suggest logical structure where none exists.

Understanding the relationship between the logical and physical structure of the document is the prerequisite for designing the basic elements of hypertext. These elements are the units of text and graphics and the links between them.

### Identifying "Natural" Text and Graphic Units

The approximately 1100 entries in the *Compendium* appear to be the "natural" units of the printed encyclopedia, and should therefore be preserved in a hypertext document. But what is it about the entries that makes them the appropriate level of structure to emphasize? What are the consequences of this choice? We identified three issues that underlie the choice of basic units in a hypertext representation of a document.

**Unique names.** A good design rule might be to choose the smallest logical structure with a unique name as the basic unit. Entries have unique numbers and titles that make it possible to refer to them and manipulate them easily in different contexts. For example, an entry number can be the "selection key" when browsing various tables of contents or indexes, a candidate in the results of a search, a "bookmark" of a document location that was previously viewed, or a cross-reference embedded in the text.

In contrast, physical structures of documents like pages or (unnumbered) paragraphs are less desirable units since they have less cognitive relevance for the user and are more difficult to manipulate in the user interface. For example, physical structures without names do not naturally accommodate a "bookmark" function because the user can not be reminded easily of their content. Instead, unnamed units are usually ordered in a stack of "places visited" that supports backtracking (by popping the stack) but not revisits to arbitrary items.

**More than one type of unit?** The *Compendium* entries contain both text and graphic parts (e.g., figures, tables, diagrams) but we do not distinguish them as different kinds of units. The graphic parts of entries are logically attached to the entry number and are retrieved when the entry is selected. This follows the structure of the printed encyclopedia in which the graphic parts are embedded in the text. If, on the other hand, the graphic parts were separately indexed and classified (e.g., as in an art encyclopedia or atlas), it might be sensible to treat them as distinct units.

**Lowest level in hierarchical classification.** If the basic units of the hypertext document are arranged only at the lowest level of the hierarchical classification (e.g., in the Table of Contents) the user interface can be considerably simplified. The *Compendium* user interface can be divided into functions that locate units by browsing or

searching and functions that display the units in a viewer window" or similar context. *Document Examiner* and *NaviText SAM* also use this scheme.

### Explicit and Implicit Links

We agree with Conklin [21] that the essence of hypertext is computer support for links within and between documents. The challenge for hypertext designers is to support the right links, those that follow from some model of the user's information needs. Arbitrary or undesigned links create a "spaghetti document" with unmanageable hypertext structure.

The number of potential links in a document as large as the *Compendium* is enormous. Each of the over 1100 entries has on average ten explicit cross-references and dozens more implicit ones based on shared glossary and key words. A pragmatic approach to linking requires the answers to these questions:

1) What is the existing link structure of the document? We were inspired by the studies by Raymond and Tompa [22] of the cross-reference structure of the *Oxford English Dictionary* to devise a new way to analyze the links between entries in the *Compendium*. We found it useful to view the links between entries as a square matrix whose rows and columns are the entry numbers. A reference from one entry to another increments the link count in their intersecting cell. This matrix representation makes it easy to determine if additional links might be useful. We found that the matrix for any type of link is sparse, but in a book as well-designed as the *Compendium* different types of links provide complementary structure. For example, the entries that are implicitly linked by membership in the same Table of Contents section (a "scientific" taxonomy) only partially overlap with those linked by the Design Checklist (a "design-centered" taxonomy).

In addition, the matrices are not always symmetric; i.e., links are not necessary bi-directional. We were puzzled at first to find numerous cross-references between entries in the "Display Interfaces" topic area and those in the "Vision" topic with few in the opposite direction. The simplest explanation turned out to be true: the "Vision" entries were written first.

2) Can additional links be easily identified? In the previous example it was straightforward to create the "cited by" links when an entry did not contain an explicit cross-reference to another entry that referenced it. Similarly, we created a browsable database of references to external documents so that a user could find all the entries that shared information from the same source. We have been exploring the automatic extraction of links by using natural language processing techniques to find related entries (e.g., [23]), but without careful analysis of how users follow the existing links in the *Compendium* we think it is premature to add new ones.

3) What link types should be allowed? Which ones

should not be allowed? Our emphasis on supporting the goal-directed, pragmatic nature of most information seeking behavior by designers discourages the creation of links merely because they are possible. The small display screen we were required to use already makes it a challenge to provide the user with contextual information, especially for occasional users [24]. We chose to limit the variety of links to limit the number of contexts that a user can enter. For example, a user can select Glossary terms when they appear in the text to open a window that displays the definition, but these are not links into the Glossary itself, which the user must explicitly choose to browse.

### TOWARD 'HYPERTEXT ENGINEERING'

Encyclopedias are inherently the easiest documents for representation in hypertext, as they contain entries that stand alone yet are designed to be enhanced by cross references to other entries. The *Engineering Data Compendium* was an exceptionally good candidate, but even it presented numerous problems of identifying and translating its logical structure into a different form.

Other documents with less modular structure and fewer explicit links may be inappropriate or impossible candidates for translation to hypertext. Our experiences in building a compact disc encyclopedia have suggested ways to determine the suitability of a document for hypertext before investing significant time and money, and we have tried to record these lessons in this paper. But these insights did not come cheaply. Including the start-up costs of learning about hypertext and compact disc technology, this project cost about \$250,000.

Given this expense, it is hardly surprising that few full-scale systems of this scale exist and that research is generally done with toy databases. We caution, however, that some of the lessons we learned were lessons of scale that did not emerge until we dealt with realistic amounts of text and graphics.

We also note that some of the hard problems we faced were difficult without being interesting as a compensation. These problems included obtaining reliable source text and writing numerous programs to translate obscure typesetting codes into usable forms.

In summary, creating a hypertext encyclopedia is an engineering problem as well as a research problem. A disciplined approach to studying the document, identifying the constraints in its structure and in the target environment, and making tradeoffs is essential. It is premature for there to be standard methods for "hypertext engineering," but we hope we have made a significant contribution.

### ACKNOWLEDGMENTS

This work was supported by the Air Force Designer's Associate contract, AF #F33615-86-C-0542. The

Designer's Associate will use artificial intelligence techniques to help designers formulate questions, find relevant information, and understand it in the context of their design of systems with complex human-machine interfaces. Someday the Designer's Associate will be a user of the compact disc *Compendium*.

Dr. Kenneth Boff of the Armstrong Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base, Ohio, is not only the sponsor of this contract but the principal designer and editor of the *Compendium*. We are grateful for his support and advice. Iopis Corporation of Columbus, Ohio (especially Mike Adams, Leonard Rivers, Jack Johnson, and Ken Constant) premastered the text and graphics and implemented the user interface. Discovery Systems of Columbus, Ohio pressed the disc. Jan Walker and Gary Perlman provided key conceptual advice. At Search Technology we received essential technical help from Brad Wiederholt and Dan Sewell and strong management support from Bill Cody.

### REFERENCES

- [1] Boff, K. R., and Lincoln, J. E. *Engineering Data Compendium: Human Perception and Performance*. AAMRL, Wright-Patterson AFB, OHIO, 1988.
- [2] Lambert, S., and Ropiequet, S. (Eds.) *CD ROM: The New Papyrus*. Redmond, WA: Microsoft Press, 1986.
- [3] *Hypertext '87 Papers*. Proceedings of a conference held at the University of North Carolina, Chapel Hill. November 13-15, 1987.
- [4] Boff, K. R. The tower of Babel revisited: On cross-disciplinary chokepoints in system design. In W. B. Rouse and K. R. Boff (Eds.), *System Design: Behavioral perspectives on designers, tools, and organizations*. New York: Elsevier, 1987.
- [5] *ThinkTank* is a trademark of Living Videotext.
- [6] *Document Examiner* is a trademark of Symbolics.
- [7] Walker, J. Document Examiner: Delivery interface for hypertext documents. *Hypertext '87 Papers*. Proceedings of a conference held at the University of North Carolina, Chapel Hill. November 13-15, 1987. (pages 307-323).
- [8] Walker, J. Supporting document development with Concordia. *Computer*, 21(1), January 1988, 48-59.
- [9] *NaviText* is a trademark of Northern Lights Software.
- [10] Perlman, G., and Moorhead, A. Applying hypertext methods for effective utilization of standards. *Proceedings of COMPSTAN'88 Conference on Computer Standards*, IEEE, 1988.
- [11] *Guide* is a trademark of OWL Software.
- [12] Brown, P. Turning ideas into products: The Guide system. In *Hypertext '87 Papers*. Proceedings of a conference held at the University of North Carolina, Chapel Hill. November 13-15, 1987. (pages 33-40).
- [13] *KnowledgeSet* is a trademark of KnowledgeSet Corporation.



- 
- [14] Carr, R. New user interfaces for CD ROM. In Lambert, S., and Ropiequet, S. (Eds.) *CD ROM: The New Papyrus*. Redmond, WA: Microsoft Press, 1986. (pages 185-196).
- [15] Frisse, M. Searching for information in a hypertext medical handbook. In *Hypertext '87 Papers*. Proceedings of a conference held at the University of North Carolina, Chapel Hill. November 13-15, 1987. (pages 57-66).
- [16] *In-a-Vision* is a trademark of Micrografx.
- [17] *Turbo Pascal* is a trademark of Borland.
- [18] *GEM* is a trademark of Digital Research.
- [19] Marchionini, G., and Shneiderman, B. Finding facts vs. browsing knowledge in hypertext systems. *Computer*, 21(1), January 1988, 70-80.
- [20] *Grolier's Electronic Encyclopedia* is a trademark of Grolier Electronic Publishing.
- [21] Conklin, J. Hypertext: An introduction and survey. *Computer*, 20(9), September 1987, 17-41.
- [22] Raymond, D., and Tompa, F. Hypertext and the new Oxford English Dictionary. In *Hypertext '87 Papers*. Proceedings of a conference held at the University of North Carolina, Chapel Hill. November 13-15, 1987. (pages 143-153).
- [23] Fox, E., Weaver, M., Chen, Q., and France, R. Implementing a distributed expert-based information retrieval system. In *Proceedings of the conference on user-oriented content-based text and image handling*. MIT. March 21-24, 1988. (pages 708-726).
- [24] Tombaugh, J., Lickorish, A., and Wright, P. Multi-window displays for readers of lengthy texts. *International Journal of Man-Machine Studies*, 26(5), May 1987, 597-615.