Addressing Hypertext Design & Conversion Issues

Robert J. Glusko

December 3, 1990
Houston, Texas
South Shore Harbour Resort & Conference Center

Co-Sponsored by
University of Houston-Clear Lake
NASA/Johnson Space Center
Hypermedia Working Group
HYPERTEXT DESIGN
AND CONVERSION
tutorial for Hypermedia '90

Robert J. Glushko
Search Technology, Inc.
Norcross, GA
(404) 441-1457

copyright 1989, 1990 Robert J. Glushko
COURSE OUTLINE

I. Introduction (:30)

II. Definitions and Basic Concepts (:20)

III. Design Issues (:60)

IV. Applications (:30)

V. "Off the Shelf" Software (:20)

VI. Converting Text into Hypertext (:30)
ABOUT THE INSTRUCTOR

Dr. Robert J. Glushko is a Principal Scientist with Search Technology, a consulting and contract research firm that specializes in user interfaces to complex systems. He received a BA (Psychology) from Stanford University, a Ph.D. in cognitive psychology from the University of California, San Diego, and a M.S. in Software Engineering from the Wang Institute. He previously worked at Bell Laboratories and the Software Engineering Institute, and has been involved with research and development in user interfaces, online documentation, and hypertext for over a decade.

Relevant publications:


WHAT IS HYPERTEXT?

A right-brain artistic panacea that is a revolutionary new concept

"Hype" + something else

An evolutionary concept for increasing the accessibility and usefulness of online text
MEMEX

Vannevar Bush (1945): "As We May Think"

"A device in which an individual stores his books, records, and communications"

* "scanner" with miniaturized microfilm as storage medium
* voice and written annotation
* most contents are purchased and inserted; "wholly new forms"
* view page by page or jump many pages at a time

Associative indexing; "the process of tying two items together is the important thing"

* a trail is the sequence of links between associated pages; each new trail creates a "virtual book"
* any item can be joined into numerous trails
* new profession of "'trail blazers'"
EXAMPLE: HYPERTEXT ENCYCLOPEDIA

ENGINEERING DATA COMPENDIUM

4 volumes / 1138 articles / 3000 pages / 2000 figures and tables

Complementary entry points

"Scientific" table of contents

Design checklist -- alternate table of contents

"Back of the book" index

Extensive cross references and external references

Regular structure for articles
HYPERTEXT FEATURES

Derived from existing structure
* Next, previous, ? entry functions
* Links to figures and tables
* Links to cross references
* Embedded glossary definitions

Derived from contextual structure
* Bookmarks
* Return to entry point
* Return to search candidate list

Derived from usability concerns
* Context-sensitive help
* "Sticky" notes

Photopic vision is characterized by high visual acuity and the perception of color. Scotopic vision and no sensation of color (hue) occurs. Vision at high intensity levels is known as photopic vision and is mediated by the rods; visual acuity is poor with mediate by the rods; visual acuity is poor with (e.g., starlight) is termed scotopic vision and is cd/m². Vision at very low levels or illumination cd/m² to an upper tolerance limit of over 300,000.

The human eye is sensitive to a wide range of light intensities, from a minimum level of ~0.000003.
Figure 2. The decrease in brightness from daylight (5:00 pm) to darkness (sunset ~9:30 pm) on July 14, 1942 (adapted from Ref. 2).

Log Brightness (cd/m²)

1700 1800 1900 2000 2100 2200

Time (hours)

Sunset

Original page is of poor quality
Abduction, 1.905. See also Eye movements
Aberration, optical
  chromatic, 1.203, 1.212, 5.934
  refractive errors, 1.204, 1.205
  spherical, 1.211
Abney effect, 1.707, 1.708
Abney’s Law, 1.109
Absolute threshold, 1.656
Absorption defect
  in color vision, 1.726
Absorption filter, 1.108
AC/A ratio, 1.231. See also Accommodation; convergence
Acceleration
  angular (rotary)
    acceleration detection threshold, 3.208
    effect on eye movements, 1.930, 1.958
    perrotary procedures, 3.205
    postrotary procedures, 3.205
    sensory magnitude, 3.208
    and visual acuity 1.603, 10.902
control of, 9.519
display of, 9.532
illusions due to, 3.210
  elevator illusion, 3.210, 5.504, 5.505
  oculogravic illusion, 3.210, 5.504, 5.505
  oculogyral illusion, 1.921, 3.205, 3.208, 3.209
See also Vestibular illusions
linear
  acceleration detection threshold, 3.207
II. DEFINITIONS AND BASIC CONCEPTS

Definitions of hypertext and hypermedia

Why is hypertext such a hot topic?

Hypertext history
DEFINITIONS

(NIST hypertext standards working group, 1990)

Hypertext

A network of information units connected by relational links

Hypertext System

A configuration of hardware and software that presents a hypertext to users and allows them to manage and access the information that it contains
WHAT IS HYPERTEXT?

Hypertext is a user interface concept that closely supports the ways that people use printed information

* more intuitive transformation of the conventions in print media that imply non-linear information presentation and usage within a document

  + entry points like tables of contents and indexes

  + relationships like footnotes and cross references

  + "usage history" like bookmarks and margin notes

* improves the integration of related information that is contained in multiple documents

Hypertext concepts encourage modularity and the elimination of redundancy in databases because information can be stored only once but viewed in any appropriate context
HYPERMEDIA AND MULTIMEDIA

Hypermedia

* Hypertext + audio, video, animation

Multimedia

* Personal computers + consumer electronics
* Synchronized presentation of full-bandwidth information in a preprogrammed way

Are there meaningful distinctions?

* "Hypermedia" emphasizes the application; "multimedia" emphasizes the technology
* Hypermedia usually implies more user control
* Multimedia usually implies more synchronization of events from different media
* Multimedia "standards" vs. hypermedia "vision"
CORE CONCEPTS FOR HYPERTEXT

Units (components, nodes, containers) -- the information content; may be text, graphics, or other media; may be "typed" by media or by content; ("card" is a user interface description)

Links -- connections between units; may also be "typed"

Anchors -- the locations (point, region, or span) in units to which links are attached

Link markers -- the manifestation of links that are presented to users

Navigation -- process of moving from one unit to another by following links

Trails/webs/guides/paths -- Subsets of units or links, created by user or as pre-defined route through the hypertext
WHY IS HYPERTEXT SUCH A HOT IDEA?

Enabling technology

* workstations and personal computers finally provide enough local processing power (for hypertext user interfaces)
* CD-ROM and other optical media for storage
* user interface software and concepts maturing

Information standards efforts with hypertext implications

* CALS for U.S. DoD
* ATA-100 for airline industry
* SGML for publishing industry

Market pressures

* incentives for digital information delivery
* Hyper-this, Hyper-that bandwagon

Academic interest

* ACM conferences
HYPERTEXT HISTORY

1945 Vannevar Bush describes the Memex
1964 Doug Englebart (SRI);
    Augmentation Research
1965 Ted Nelson; Xanadu concept;
    coins "hypertext"
1968 Englebart's NLS demo;
    1st hypertext system (outline viewer, mouse, bookmarks)
1968 van Dam & Nelson (Brown);
    Hypertext Editing System
1969 van Dam; FRESS (graphical views,
    history timelines, undo)
1972 Newell (CMU); ZOG
    (card metaphor)
1979 van Dam; Electronic Document
    System (graphical documents)
HYPERTEXT HISTORY (CONT.)

1980  Newell; ZOG on USS Carl Vinson
1981  ZOG commercialized as KMS
1983  Institute for Research on Scholarship (Brown)
1984  Halasz et al. (Xerox); NoteCards (programming environment)
1985  Walker (Symbolics); Document Examiner (on line manual)
1986  Guide (1st PC hypertext)
1987  HyperCard, HyperTIES
1987  ACM Hypertext conference
1987  Walker; Concordia (authoring tools)
1988  1st wave of hyperclones
1988  ACM "Hypertext on Hypertext"
1989  Hypermedia; 1st dedicated journal
1989  ACM Hypertext conference II
HYPERTEXT HISTORY:
EVOLUTIONARY VIEW

1960s -- Computer databases for limited storage and retrieval of abstracts, but no full text

1970s -- Text databases and information retrieval on mainframes and large minis emerge; Online documentation and online help emerge on micros and minis

1980s -- Workstations and high-end PCs have enough power and capacity to support text databases and better user interfaces --> hypertext functions

1990s -- Software and hardware support for hypermedia in "off the shelf" computing environments
III. DESIGN ISSUES

Hypertext functionality
Entry points
Units
Links
Navigation support
HYPERTEXT FUNCTIONALITY

Functions identified via task analysis

* information needed vs. information used
* need "friendly users" to make functions "user friendly"

Functions identified via document analysis

* existing and potential entry points
* existing and potential units
* existing and potential interconnections
* "front matter" and "gray matter"

Typical functions

* progressive display of structure
* create, edit, display, delete units
* create, follow, edit, delete links
* create notes or bookmarks
* search for units with specified attributes
ENTRY POINTS

(Places for users to enter the hypertext or to locate a starting unit)

Existing entry points

* Table of contents
* "Back of the book" index
* Glossary

Potential entry points

* Full-text inverted index
* Author list created from inverted references
* Timelines
Accounting

The purpose of accounting is to provide information about the economic affairs of an organization. This information may be used in a number of ways: by the organization's managers to help them plan and control the organization's operations; by owners and legislative or regulatory bodies to help them appraise the organization's performance and make decisions as to its future; by owners, lenders, suppliers, employees, and others to help them decide how much time or money to devote to the organization; by governmental bodies to determine how much tax the organization must pay.

Accounting provides information for all these purposes through the maintenance of files of data and the preparation of various kinds of reports. Most accounting information is historical—that is, the accountant observes the things that the organization does, records their effects, and prepares reports summarizing what has been recorded.

Accounting information can be developed for any kind of organization, not just for privately owned, profit-seeking businesses. One branch of accounting deals with the economic operations of entire nations. The remainder of this article, however, will be devoted primarily to business accounting.

The article is divided into the following sections:

<table>
<thead>
<tr>
<th>Financial statements</th>
<th>Management accounting</th>
<th>Problems of measurement</th>
<th>Net income</th>
<th>Owners' equity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Disclosure and auditing requirements</td>
<td></td>
<td>10. Asset cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Asset cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The liabilities are similarly divided into current liabilities and noncurrent liabilities. Most amounts payable to the company's suppliers (accounts payable), to employees (wages payable), or to governments (taxes payable) are included among the current liabilities. Noncurrent liabilities consist mainly of amounts payable to holders of the company's long-term bonds and such items as obligations to employees under company pension plans.

The difference between the total of the current assets and the total of the current liabilities is known as net current assets, or working capital.

The owners' equity of a U.S. company is divided between paid-in capital and retained earnings. Paid-in capital represents the amounts paid to the corporation in exchange for shares of the company's preferred and common stock.

The major part of this, the capital paid in by the common shareholders, is usually divided into two parts, one representing the par value, or stated value, of the shares, the other representing the excess over this amount. The amount of retained earnings is the difference between the amounts earned by the company in the past and the dividends that have been distributed to the owners.

A slightly different breakdown of the owners' equity is used in most of continental Europe and in other parts of the world. The classification distinguishes between those amounts that cannot be distributed except as part of a formal liquidation of all or part of the company (capital and legal reserves) and those amounts that are not restricted in this way (free reserves and undistributed profits).

A simple balance sheet is shown in Table 1. Because the two sides of this balance sheet represent two different aspects of the same entity—the corporation's capital—the totals must always be identical. Thus a change in the amount for one item must always be accompanied by an equal change in some other item. For example, if the company pays $40 to one of its trade creditors, the cash balance will go down by $40, and the balance in accounts payable will go down by the same amount.

The income statement. The company uses its assets to produce goods and services. Its success depends on whether it is wise or lucky in the assets it chooses to hold and in the ways it uses these assets to produce goods and services.

The company's success is measured by the amount of profit it earns—that is, the growth or decline in its stock of assets from all sources other than contributions or losses.
USER INTERFACES FOR ENTRY POINTS

Outline viewers

* Purpose is to provide progressive display of structure with detail where user requests

* Often called "browsers" but confusion with browsing as "wandering around"

* Essential component when units are hierarchically structured

* Can often be created semi-automatically from document markup, or by "extraction and inversion" such as for timelines or reference lists

When "browser" uses graphics, blurred distinction between entry points and navigation support after entering
From NavText SAM by Northern Lights Software, Westford MA. (508) 692-3600.

GUIDELINES FOR DESIGNING USER INTERFACE SOFTWARE

- User interaction with microcomputer terminals. Human hand copy and cathode-ray tube (CRT) computer terminals. Could not do computer work with CRT. Doing the same work with cathode-ray tube.

- Reading lengthy text on an electronic display may be 20-30 percent slower than reading it from a printed copy.

- Comment: When a user must read lengthy text, material, consider providing the text in printed form rather than requiring the user to read it on-line. Text in printed form rather than requiring the user to read it on-line.

- Consistent text format
- Preparing lengthy text displays
- Conventional text display
- Data displays
- General
- Text
- Data display
- I DATA ENTRY
- Line 1 of 15
IDENTIFYING UNITS

Units are both "containers" and "addresses" that have names to support functions like searching and giving directions, especially with large amounts of similar units.

Units may seem self-evident but usually aren't

* "Natural units" seem to be articles in an encyclopedia, items in a catalog, but...

* Is "page" an important unit? No, well maybe...; West Pub. vs. Mead Data

* Constraints on unit definition
  + delivery software
  + other (JASIS and ASIS Bulletin)

Definitions of "natural" unit

* "The smallest logical structure with a unique name"

* "A component that says something self-contained and comprehensible"

* "Whatever makes the links right"
5.4.2.2 Continuous adjustment rotary controls.

5.4.2.2.1 Knobs.

5.4.2.2.1.1 Use. Knobs should be used when low forces or precise adjustments of a continuous variable are required. A moving knob with fixed scale is preferred over a moving scale with fixed index for most tasks. If positions of single revolution controls must be distinguished, a pointer or marker should be available on the knob.

5.4.2.2.1.2 Dimensions, torque and separation. The dimensions of knobs shall be within the limits specified in Figure 7. Within these ranges, knob size is relatively unimportant, provided the resistance is low and the knob can be easily grasped and manipulated. When panel space is extremely limited, knobs should approximate the minimum values and should have resistance as low as possible without permitting the setting to be changed by vibration or merely touching the control. Resistance and separation between adjacent edges of knobs shall conform to Figure 7.

5.4.2.2.1.3 Knob style. Unless otherwise specified by the procuring activity, control knob style shall conform to MIL-STD-1348.

5.4.2.2 Ganged control knobs.

5.4.2.2.2.1 Application. Ganged knob assemblies may be used in limited applications when panel space is at a premium. Two-knob assemblies are preferred. Three-knob configurations should be avoided. Ganged knob configurations should not be used under the following conditions:

a. Extremely accurate or rapid operations are required.

b. Frequent changes are necessary.

c. Heavy gloves must be worn by the operator.

d. Equipment is exposed to the weather or used under field conditions.

5.4.2.2.2 Dimensions and separation. Dimensions and separation should conform to Figure 8.

5.4.2.2.3 Resistance. Resistance shall conform to requirements in Figure 8. Knobs should be serrated. Fine serrations should be used on precise adjustment knobs; coarse serrations should be used on gross adjustment knobs.

5.4.2.2.4 Marking. An indexing mark or pointer shall be provided on each knob. Marks or pointers should differ sufficiently to make it apparent which knob indexing mark is being observed.
User interface metaphors (Raskin)

* "Card sharks" -- creators
* "Holy scrollers" -- converters

Are units displayed as fixed size?

How many units can be displayed together?

Can text and graphics be displayed together?
FIGURE 1. Example Notebook with Embedded Link Icons


See Klimet's, "(P. 73) Accuracy after a journey of 1,600 kilometers. . ."

Tomahawk Characteristics

Guidance of Pershing

The new American Pershing II FABILITIES OF NEW MISSILES (IN QUESTION)

Even though the weapons in question

not have sufficient range to make

replace older weapons (in Pershing

out reach the U.S.S.R. in the case of

vulnerable installations and cities in

also have sufficient range to make

Guidance of Pershing II

P 2. Without a matter of minutes.

The Western USSR in the case of

their predecessors. This is the

capable of more destruction faster

and the Vietnamese bomber, both are

MISILE FANFARE (MICHAEL FANFARE)
LINKS

Existing interconnections -- usually easy to identify because of conventions in printed information

* Index terms
* Cross references
* External references
* Footnotes

Potential interconnections

* Inverses of existing ones
* Lexical relationships; indexing and clustering
* Conceptual relationships (the hypertext vision!); AI and natural language processing
* Emphasis or complement in alternate medium; e.g., connect a picture to its description
readable product of these efforts. A more precise use of the term "program" would limit its meaning to source code and machine code (often referred to as "object code"). Source and machine code are similar in that both are sets of detailed instructions setting forth the order in which the hardware of a computer is to execute its primitive functions in order to carry out a particular task. Source code, however, is a written text in a human-readable computer programming language. Machine code is the set of electrical pulses that, more or less, correspond to the source code and make the program instructions "readable" by the computer. Machine code is not readable by human beings. In general, only machine-readable forms of programs are

70. A bill, H.R. 6983, 97th Cong., 2d Sess. (1982), was introduced in the House of Representatives by Congressman Kastenmeier on August 12, 1982, that would have amended § 101 of 17 U.S.C. to redefine "computer program" in a more precise way and to define separately several program-related terms, often loosely referred to as manifestations of programs. The bill did not become law, however, and was not reintroduced in the next session of Congress.

71. The computer program cases tend to refer to "object code" when referring to machine-readable forms of computer programs and often imply—when they do not say so outright—that source code and object code are the only forms to be considered. See, e.g., Apple Computer, Inc. v. Franklin Computer Corp., 714 F.2d 1240, 1243 (3d Cir. 1983). There are, in fact, several intermediate stages possible between source code and the machine-readable code that can be executed. See, e.g., I. D. Kuck, The Structure of Computer and Computations 10 (1978). Depending on which hardware and which operating system one uses, object code may be one of those intermediate forms, not the executable form of the program. See, e.g., R. Hunter, The Design and Construction of Compilers 11 (1981). Because the cases have involved appropriations of machine-readable versions of programs, which may not be the same as their object codes, this article will focus its analysis on what it will call "machine-readable programs" or "machine code." It will refer to machine-readable programs as "object code" only when the terminology of another source under discussion requires use of that term for consistency.

72. The CONTU Final Report defines source and object code as follows: "A source code is a computer program written in any of several programming languages employed by computer programmers. An object code is the version of a program in which the source code language is converted or translated into the machine language of the computer with which it is to be used." CONTU FINAL REPORT, supra note 1, at 21 n.109.

73. It is possible to write a program directly in machine-readable form, but this is rarely done because of the difficulty of writing in machine language. See infra note 74.

74. In source code form, the ideas of the program, as well as the particularities of the expression of the ideas in the program, will be apparent, that is, capable of being read by someone who understands the language or symbols the program author has used to describe the program. Source code, like poetry, may contain some abstruse words whose precise meaning might be open to interpretation by readers of the source code, but whatever content there is in the program is there to be discerned. With machine-readable code, neither the ideas nor the expression of the ideas can be "read" in any meaningful sense by one who has no access to the earlier written form of the program.

Examining an encoded ROM chip with an oscilloscope, see supra note 32 for an explanation of ROM chips, one can detect the presence or absence of the electrical pulses which constitute the machine code. But machine code is so unreadable that the Copyright Office cannot even identify whether a particular encoded program is an original work of authorship. See infra note 218 and accompanying text.
LINK ISSUES (AND TYPES)

Link anchors -- what is being linked (granularity)?

* Links from unit to unit
* Links from points, spans or regions to units; vice versa

Is the anchor a place (bit position) or an object?

Links as structure connections

* from table of contents or index to unit

Links as relationships

* cross references or semantic (typed) relations between text units
* present related graphics, sounds, music, videodisc, animation

Links as functions or computations, often not resolvable except at runtime

* compute where to go based on history or current context
LINK ISSUES (CONT.)

How do you indicate links to the user?

* Source and destination markers
* Link type markers

Marker options

* Link symbols or icons
* Reverse video
* Surround boxes
* Bold, italics, underlining, color
* Cursor shapes
* Flashing or other motion ("moving ants")

One marker attribute vs. two delimiters?

Overloading of print conventions?

How many marker types can users distinguish?

Markers for non-text media?
From a HyperCard program written by Jakob Nielsen of the Technical University of Denmark. (The content is a trip report from the 1987 ACM Hypertext Conference.)
IDENTIFYING NAVIGATION AIDS

Purpose

* finding the right place
* finding the next "right place"
* returning to some previous place

Unit names, especially numerical ones

"Roadmaps" (usually in preface)

Structure landmarks

* Volumes, tabbed dividers, color cues

Sequence and cohesion signals in text

* "First," "next," "another," etc.

Running heads

Navigation aids defined by user

* Bookmarks, turned pages, incidental marks
Part 2. Word Processing and File Access

| Chapter 4: An introduction to text editing and formatting and to automatic typesetting | Trained computer-science students and computer professionals | may be skipped by persons experienced in text editing |
| Chapter 5: Statistical language analysis and basic text-compression methods | Computational linguists and language-processing researchers | Students in library and information science or science and technology programs |
| Chapter 6: Introduction to text encryption and a review of basic text-encryption methods | | |
| Chapter 7: File-access methods for single-key and multiple-key search statements | those familiar with data structures may start with Section 7.8 | |

USER INTERFACES FOR
NAVIGATION SUPPORT

Extremely important aspect of hypertext design because without it applications are unusable

Maps

* global vs. local
* active vs. passive
* handmade vs. system-generated

Places visited

* "you've been here" markers
* Backtracking for in-order return to "places visited"
* Bookmarks for out-of-order return to any previous place
* History lists

Places to visit

* Paths, scripts, paths, guides
IV. APPLICATIONS

Survey of hypertext applications

Creation examples
Conversion examples

Common characteristics

Relationship to similar non-hypertext applications
HYPERCARD HELP

Graphical hierarchical entry point

* enter hierarchy at top (Introduction) or second level
* hides significant detail (414 cards)

"Filebox" metaphor

* but why is it upside-down?

Active navigation map (hand-drawn)

* select a location to move to it
Click the Browse tab below after you've seen the introduction.
From HyperCard by Apple Computer.
"Document Examiner" contains complete reference manual for Symbolics workstation software (Jan Walker)

* uses book metaphor, not network metaphor
* consists of 10,000 text modules, corresponding to 8000 printed pages
* readers find relevant information using outline viewer or by query

System-generated active local maps for navigation

Bookmarks for returning to previously viewed places
Conversion of existing notebook into NoteCard chunks

* made easier by strong hierarchy of existing material

* but not always good names for sections, so use first few words as unit name; works in recognition situation for expert users but probably not in casual use application

Problem of "hierarchically distributed keywords" typical of conversion applications

* what a paragraph is about not always clear when viewed out of context

* use search algorithm that combines intrinsic weight of unit with weight of its descendants in hypertext hierarchy
Endotracheal intubation and tracheostomy.

IV. Endotracheal tubes are classified by size according to their internal diameter. Since the resistance to airflow is proportional to the fourth power of the tube radius, a large tube (e.g., > no. 8) is preferable to minimize airflow resistance and the work of breathing. A large tube also allows easier suctioning when indicated.
IMIS

Integrated Maintenance Information System being developed by Air Force (Dave Gunning <513> 255-2606)

* designed to merge diagnostics from airplane with maintenance documentation to reduce extraneous information and tailor instructions

* goal is to put it all in a portable computer that can be carried out to the plane

Lesson: hard to convert existing documents; should require contractors to deliver information in "neutral" form that is free of paper and formatting conventions

* IMIS is being watched carefully by CALS program!!
(2) Generator Low 20dB 50-2000M

(3) Signal From Signal Generator IN/OUT

Supplied with AN/UPM-137A

(d) Connect 18-inch RG-62A/U cable

(4) Aux. Mod IN

Amplifier (I) to Signal Generator

SIS Generator CHAL/TAG YRF

(c) Connect RG-62A/U cable from

Preliminary Control Settings and Connections

2.5.6.3c
Research project at University of Waterloo to convert OED to digital form and assess potential for hypertext version

* Big document => big database! 20 volumes, 320,000 entries, 2.4 million quotes, 56 million total words

* 21,000 printed pages => machine-readable form by manual keyboarding. Twenty "tags" in entries later expanded to sixty.

Seems like a perfect candidate for hypertext, but...

* What are the units? Dictionary entries vary in size from a few words to thousands and have complex internal structure. Better to dynamically construct units as views from entry "database"

* Where are the links? 600,000 explicit cross references, many imprecisely specified. Fast full-text search can simulate linking with much less implementation effort
SIMPLIFIED STRUCTURE OF OED ENTRY WITH COMMONLY OCCURRING TAGS

ENTRY

HEADWORD GROUP
- Headword Lemma
- Murray Pronunciation
- IPA Pronunciation
- Part of Speech
- Homonym Number

END OF HEADWORD GROUP

VARIANT FORM LIST
- Variant Date
- Variant Form

END OF VARIANT FORM LIST

ETYMOLOGY

SENSE(S)
- Sense Number
- Definition
- Quotation Paragraph
  - Earliest Quote
    - Date
    - Author
    - Work
    - Text
    - End of Earliest Quote
- Quote(s)
- Latest Quote (Obsolete Entries Only)

End of Quotation Paragraph
- Sub-Entry (Preceded by "Hence")
  - Bold Lemma (+ similar tags to those following Headword Lemma)

End of Sub-Entry

END OF SENSE(S)

END OF ENTRY
CHARACTERISTICS OF GOOD HYPERTEXT APPLICATIONS

Large amounts of multi-media units related more by content than by structure

* hierarchical or procedural structures especially useful

Need for non-sequential partial access to units

* not cover-to-cover
* intermittent or unpredictable use
* "right answer" easier to recognize than to specify

Pragmatic considerations

* document being written or rewritten
* source text already available, usable markup (SGML as an ideal)
* if existing text, you own the copyright or not copyrighted
RELATIONSHIP TO SIMILAR NON-HYPERTEXT APPLICATIONS

Databases -- more appropriate when relationships among units are highly regular or text only; "link types" inflexibly fixed in db schema; but distinction blurring rapidly as dbs adopt hypertext front ends

Electronic mail or conferencing systems -- units related through fixed header fields or by simple question-answer relationships

Multimedia (e.g., video, film) -- synchronized media (like soundtrack) not selected and invoked by user

Scanned image storage and delivery with attached keywords -- can be useful intermediate stage in transition from pure paper environment (with lots of drawings and pictures) to hypermedia system
V. "OFF THE SHELF"
HYPERTEXT SOFTWARE

What's available

How to evaluate it

Issues and features

Alternatives
WHAT'S AVAILABLE

PCs

Black Magic
Guide
HyperDoc
HyperPad
HyperTIES
KnowledgePro
LinkWay
NaviText
SmarText
ToolBook
Window Book
WHAT'S AVAILABLE (CONT.)

Mac

ArchiTect
Document Examiner
Guide
HyperCard
HyperGate
Intermedia (Mac Unix)
Plus
SuperCard

Workstations

Document Examiner
KMS
Knowledge Broker
NoteCards
EVALUATION CRITERIA

Environment(s) it runs in
Support for text structures
Support for non-textual components
Link types and granularity
Access methods/entry points
Navigation and session support
Extensibility

(Manual, support, marketing)

"Purpose"
ENVIRONMENTS

Programs run on PCs, Macs, Workstations

* but few run on more than one platform
  + Guide on PC and Mac
  + Document Examiner on Symbolics, MacIvory

* some vaporware rumored to "compile" HyperCard stacks to run on PC

Typical software architectures make interchange formats a long way off

* most programs do not separate back and front ends

* most programs use proprietary back ends and special formats anyway
TEXT STRUCTURES

Typical program limitations

* total number or size of units
* no scrolling text for "card" or "page" metaphor programs

* number of units that can be displayed simultaneously
  
  + just one
  
  + one, with temporary pop-up overlay window

* single font or type size
  
  + makes it hard to reproduce look of printed document
  
  + some programs claim character-only display is asset!
Some programs work only in character mode (and claim that lack of graphics is an advantage!)

Other programs "launch" graphics programs to display graphics that take over the screen, so graphics can't be mixed with text or made the source of a link.

Most programs can display graphics in standard formats

* many can cut-and-paste graphics to merge with text
* fewer have capability to edit or resize graphics once imported

Installed base of small display screens is a constraint that will eventually go away

Other problems with graphics displays await cognitive rather than technological solutions
LINK TYPES AND GRANULARITY

Many programs have built in link types with conventions for representing source, destination, and link semantics

* Guide uses cursor changes for different link types
* HyperCard conventions and button examples
* HyperTIES built-in link preview

Most important practical distinction is link "granularity"

* unit to unit links only (often closely tied to card or page metaphor) impose pressure to keep units short which is a significant constraint if no text scrolling

Can graphics be made the source of a link?

* essential for active navigation support and "exploding" diagrams
ACCESS METHODS AND ENTRY POINTS

Most programs lack good support for existing entry points like Table of Contents or Indexes

* Some programs automatically build outline viewers from article or unit titles

* For other programs outline viewers can be hand-crafted from link repertoire, but this can be tedious

Most programs have some simple text search capability, but this mechanism is often inadequate for large hypertexts
NAVIGATION AND SESSION SUPPORT

Many programs have built-in navigation functions

* Commands for next, previous, first, last units in current "stack," "folder," or whatever name given to current unit context

* Backtracking functions are more common than bookmarking functions because the former don't require that the visited units or screens have names

* HyperCard has "graphical bookmark" facility with active pictures of recently displayed cards

* Useful but seldom-seen feature is creating bookmark reminder without actually having to go there
EXTENSIBILITY

How can programs provide functions that aren't contained in the off-the-shelf version?

* programming language (NoteCards, Document Examiner)

* scripting language (Guide, HyperCard, HyperDoc, HyperPad, LinkWay, SuperCard)

* limited ability to invoke external programs or control serial port (Hyperties)

Sometimes the vendor or a 3rd party can write custom software

Some programs are simply closed environments and not able or willing to interact with other programs
PURPOSE

Basic distinction between authoring and conversion usually has scale implications

* which is the harder problem?
* few programs are designed explicitly to support conversion

+ Concordia (authoring tools for Document Examiner) sets the standard
+ other announced and unannounced programs look promising

Some "hypertext" programs are much better used as user interface prototypers than as hypertext delivery vehicles

In contrast, many database programs or expert system shells have successfully added some hypertext features

Does it matter if it is "really" hypertext if it helps the user?
VI. CONVERTING TEXT INTO HYPERTEXT

Why conversion is an important problem

Why converting existing documents can be harder than creating new documents

Conversion questions

Conversion philosophies

Hypertext engineering
WHY CONVERSION IS AN IMPORTANT PROBLEM

An enormous installed base of paper makes some amount of conversion inevitable!

Sometimes you can't create

* Qualified writers or subject matter experts not available
* Not enough time or money

Sometimes you must convert

* Distribution and maintenance costs
* The "productivity paradox" -- desktop publishing vs. CASE -- that results in lagging documentation
* Regulatory or market pressures (e.g. EC 92)
* Paper form of information paper won't fit
  + submarine, space station
WHY CONVERTING EXISTING DOCUMENTS CAN BE HARDER THAN CREATING NEW DOCUMENTS

The structure of existing documents must be identified from partial or ambiguous information.

The author's explicit or implicit design rationale may not be available to aid in the conversion (while the author is always available to help in creation projects!)

In conversion, issues of scale are critical from the outset; in creation, complexity is incremental.

Software tools for creation are numerous, but few are available that support conversion; using creation tools for conversion projects is difficult.
CONVERSION QUESTIONS

What kinds of documents make the best candidates?

What aspects of documents pose problems for conversion?

How much conversion can be done automatically or semi-automatically?

Into what format should the document be converted?

What tools and methods are needed to support conversion?

Can you live with the constraints imposed by "off the shelf" hypertext software? Should you?

What are the alternatives to using "off the shelf" hypertext software?
CONVERTING GRAPHICS

Screen size and resolution is often significantly inferior to that of the printed source materials

Except for simple line drawings or pictures, display limits make scanned bit maps illegible

* replacing scanned text "extends the viable range"

Vectorizing reduces storage space and display time, supports zooming and panning

* tradeoff between cleanup and redrawing

Sometimes it is necessary to redesign the graphics

* multiple panels => "animation"
* data plots => "data viewer"
Figure 1. Perceived roll of single produced by infranfield rolling visual stimuli as a function of head position. Symbols indicate individual subject data and solid lines connect median values. (From Ref. 1.)
ALTERNATIVES

Hypertext conversion services
  * HyperTRANS (Texas Instruments)
  * Window Book

Text Managers
  * Views (Folio)
  * Topic (Verity)
  * Knowledge Broker (Apollo)
  * HyperSift + AskSAM (AskSAM Systems)

Desktop Publishers
  * Interleaf
  * Arbor Text

Visual Programming
  * Layout

Other
  * Wingz
CONVERSION PHILOSOPHIES

"Hand-crafted" hypertext

* "Hypertext requires creativity; only if you build a hypertext by hand can you add any real value."

"Computer-generated" hypertext

* "If you don't convert automatically, it will never be cost-effective."
* "I don't care if it isn't really hypertext; it's more usable than it was before."

"Engineered" hypertext

* "I'll begin with automatic conversion and custom design the parts that don't convert well by computer."
* "I'll try to influence the 'upstream' processes so that the documents are easier to convert next time."
* "I'll figure out how much hypertext I really want and am willing to pay for."
"HYPERTEXT ENGINEERING"
(Glushko, et al 1988)

Hypertext is an attractive vision, but practical hypertext applications are hard to build.

Hypertext is not a revolutionary new idea; it is the natural extension of decades of work in computer storage and retrieval of text now that enabling technology and user interface concepts have arrived.

Successful hypertext projects are those that take a cautious approach to problems of scale and that make the right tradeoffs along the way.

A disciplined approach to analyzing information, identifying constraints in its structure and in the task environment, and using the appropriate implementation technology are required.