Enabling Extremely Rapid Navigation in Your Web or Document

Michael Hoffman
3/21/96

This article presents information design techniques that apply to web sites, help systems, hardcopy, and online documentation. When the standard document navigation structures are provided, readers can rapidly survey the scope of a web or document and jump to the pages of greatest interest. This article explains the nature and benefits of detailed outlines and recommends that web authors provide a reasonably detailed and structured outline of their web site. Surfing the web can be speeded up greatly by loading fewer irrelevant pages and by giving users an (additional) alternative to page-by-page exploration, thus avoiding the lost-in-hyperspace syndrome. The distinctions between overviews, tables of contents, full-text searches, database keyword searches, and topical indexes are explained, to justify providing multiple approaches for the reader.

Contents

The Absurd Inadequacy of the Current Conventions for Web Navigation

It has often been said that when desktop publishing came out, poor page layout was rampant. Now with the Web we're seeing the results of poor online document design, resulting in difficult and restricted means of navigation. The extreme hypertext approach has actually *restricted* our ability to use documents, rather than *surpassing* the features of hardcopy.

Once you get the terminology right, you find that in fact, typical web sites today have:

- An Overview (called "Home Page" or "Contents")
- No Table of Contents
- No full-text Search
- A poor Index (called "Search") built by automated guesswork

This lack of lookup and surveying methods is not a difficult design problem. Familiar, flexible devices for information access already exist, such as a detailed outline, which for a book is called a "Table of Contents". Most readers would greatly appreciate having a fast and organized way of surveying all of a site's pages to decide which pages to jump

to. A related problem is poor results from search tools. This is easy to fix by creating appropriate index keywords at the top of each page. This article evaluates the navigation power of various devices for accessing information in documents. When these basic structures and principles are understood, the full potential will be realized for navigating the Web and for accessing information in printed documents.

Goal: Minimize Lookup Time

Lookup time is the crucial factor to optimize when designing the navigation features of a web or document. There is a fairly recent book by a disenchanted technical writer who feels helpless against the onslaught of unhelpful info-access devices in documentation. The book is critical of the usability of online information, because the online access mechanisms fall so ridiculously far short of what they could and should be. In several places, he says: "Perform an experiment of how long it takes to accomplish [some trivial computer task in an unfamiliar program] by looking it up in the documentation. Hint: you won't need a watch with a seconds hand."

The Web needs to wake up to the power of complete, hierarchical ToC's and Indexes. Currently, few information-access mechanisms are provided. We need good examples to copy. Study Windows Help/95 very closely. Study well-written traditional nonfiction books: *grok* the power of the hierarchical Table of Contents (ToC) and hierarchical, topical Index. I emphasize "hierarchical", "well-written", "complete", "well-worded", and "well-conceived", because almost every time I click a button labelled "index" or "contents", it takes me to yet another completely inadequate page that lists only the level 1 or maybe the level 2 pages in the site, in alphabetical or random order.

The 5 Navigation Devices for Documents: Overview, Detailed ToC, Full-Text Search, Database Keyword Search, and Topical Index

There are 5 standard ways which can be used to access information in books or webs:

1. **Overview**, or **Brief Table of Contents** -- with links. Includes Navigation Bar. These approaches show heading level 1, or 1 and 2.
2. **Detailed Table of Contents** -- Hierarchical tree of topics showing heading levels 1 through 4 or 5.
3. **Full-Text Search** -- Finds a string within the body text of the main documents.
4. **Index** -- Topical, indented, alphabetical. Handcrafted entries for each subsection.
5. **Database Keyword Search** -- Searches a special "index" built by automatic

guessing.

With the current web design conventions, the user is very limited in the choice of lookup mechanisms. They have methods 1 and 5, plus full text search within the current document (by using the browser's Find).

Which of these devices would people use if they could, if all these alternatives were available? The more access mechanisms, the better. I think the detailed Table of Contents would be very popular.

Windows Help in Windows 95 is the best example of the difference between true 'Contents', true 'Index', and true 'Search'. The Web search engines are a little different than any of these 3 devices. Part of the confusion among all these devices is that there are two types of Web searches: full text search engines, and keyword search engines. The full-text search engines are the same as full-text search in WinHelp. The keyword search engines are harder to compare to traditional book structures or full-text online Find's. functions. A keyword search engine is more of a classic database lookup... a fourth category.

The **HTML Reference Library** is a **WinHelp** file, so it enables you to look up HTML coding information using a genuine Contents, genuine Index, and if you have Win95, full-text Search. It's a 5-cow program available from tucows.

Another great document for studying info-access mechanisms is Sun's Java API Documentation in WinHelp/95 format.

If you are writing WinHelp3x files, Blue Sky Software has great extensions to give your WinHelp file all the WinHelp95 features including full-text search. The Contents and Index features are greatly improved too. Now, WinHelp is obviously equivalent to traditional books -- WinHelp continues to lead the way in hypertext design, by embracing and extending book devices. WinHelp95 also supports ultra-granular, 1-item popup help for dialogs.

Offer a Variety of Views and Access Methods

To make your web site *usable*, so that visitors can find information as quickly as possible, offer the user a variety of views and access methods. The best place to start is to add a Table of Contents (ToC). Tables of Contents seem familiar, but are poorly understood by Web authors.

Of course it takes a lot of time to add headings and diagrams, and summarize ideas into bulletted lists or "executive summaries". The time that the author spends is time saved for the reader. It takes time to update a ToC page to reflect and organize the many new pages added to a site. But the more organized the ToC is, the more powerful it is -- the faster the visitor can scope out the site and find a desired page.

A detailed, organized ToC translates to *speed* for visitors -- that's the bottom line. It's not

free; it takes effort to make and maintain such a useful, powerful navigation accelerator. It also takes time to add good index keywords. But given that programmers and HTML hackers go to the index before trying the Contents, it is time well-invested if you really do care about helping the user. If you deliberately add keywords, you will help users find the right type of pages when using a keyword-oriented search engine.

Confusion over the 2 types of 'Search'

A button labelled "Search" is ambiguous. Does it take you to a page that enables you to do a *full text* search... or just a "search" within an index document? You cannot know. A button labelled search is ambiguous: it might provide full-text search, or actually provide *index* functionality.

You can do a "full-text search" that is not a full-text search of end-documents, but rather, a search within an index document. A database keyword search effectively does a full-text search on an index that is built from document titles, hypertext hotspot text, headings, and deliberately defined keywords. But the index document it searches is not presented to the user as a visually scannable document. Instead, the list of actual pages is returned. When you look something up in a book's index, an item might have several pages listed:

Cat Scratch Fever - see *Amboy Dukes*
cats - iii, 15, 184, 203, 482-484
    feral - 15, 483-484
crashing -- see *cars*

Most of the web search engines skip this view and list the linked titles of the pages, so you cannot see neighboring entries.

Confusion over 2 types of 'Contents'

The term "Contents" should be *reserved* for a full, hierarchical ToC listing every section at least down to the page level. Otherwise, call it an "overview" or a "map".

Confusing 'Contents', 'Search', and 'Index'

The terms 'index' and 'contents' have become vague and generic, often resulting in something less useful than traditional Indexes and Tables of Contents.

Windows Help was confused, at first, about the meaning of "Contents", "Search", and "Index". But finally WinHelp has evolved fully. And you know what it looks like? A book. This suggests that a fully mature hypertext design comes back around to the book paradigm for information structuring and access. The traditional book contains these 2 info-access devices: Contents, and Index. Online documents are well-suited for full-text searching, but this feature is not generally available for the Web.

WinHelp/95 finally has the full inventory of info-access mechanisms:

1. Hierarchical, complete Table of Contents. Might not list small headings within a "help topic", but lists at least every page. As a user, I want *every* page listed, for almost all types of documents.
2. Search: finally, a full-text search. (Before, "search" actually showed the *index*. And the "Index" button brought up the *Contents*!)
3. Index -- a *real* index! It's *indented* and visually scannable. It shows distinct *levels* of entries. Every page in the document has several well-conceived entries by which the user is most likely to look up the page.

The Web information design standards currently have the same problems as Microsoft Windows Help used to have -- confusion over the meaning of 'outline', 'contents', 'index', and 'search'. Most "search" boxes are not full-text search -- they actually only search the "keyword index" page, which is not itself shown. The Keyword list is built by automated guesswork and looking for defined keywords, which no one uses. No wonder the Search engines so often return no results. The web standards for retrieval paths are very poor, almost nonexistent, compared to the traditional ToC and Index.

Every HTML heading should enable you to define true 'Index' terms, for multi-level indexes -- and not just for database keyword lookup "indexes". So far the only place you can really put index entries deliberately is in the poorly documented tag <meta type=keywords contents=" "> which can only go in a single location, at the top of an entire page. This is not granular enough.

Doing a 'Search' -- within a Table of Contents or Index Document

When you do a search in Yahoo, you are doing a search on both a Table of Contents (ToC) structure and a database keyword index structure. The search returns a mix of types: topic entries from the topic tree, and specific site-description entries. This is where a lot of confusion can enter about "doing a search". This is a full-text search not of the end-documents, but of the Contents and Index documents. If Yahoo had whimsical topic titles, this search would return bad results. It's important to use clear, unambiguous terms in your headings, because they show up in the ToC, which is searched visually or automatically. People would use ToC's more if they were clearer.

ToC's reach their full power when put online, because you can do a 'Find' -- a full-text search limited to the Contents page. You can then scan the structure -- the "chapter" -- to scope out the adjacent, related topics, to get the powerful benefit of serendipity.

Clarifying the Meaning of "Contents": A Hierarchical Tree of Headings

A screen of 10 or 20 links is far less helpful than a traditional, systematic, hierarchical ToC and index. The words "Contents" and "Index" must not be used lightly -- they are heavy, serious words that imply specific promises of completeness, order, and usability.

A genuine Contents list is not presented alphabetically, like the many web pages mis-labelled "Contents". A ToC or Outline is organized by topical order, using a hierarchical tree structure. Every medium-length page should appear in the site's Contents page. If a page is long and contains many headings, then you can place a partial table of contents (PToC) at the top of the page, and add '#' links to automatically scroll down. I manually generated the PToC for this article. A tool should have generated it for me automatically!

The Organizing Power of the Outline

A Table of Contents is a form of outline. For the theory of outlining, check out the ideas behind MaxThink, an idea-processor or outliner. It's one of the few programs that resulted from the excitement about outliners that occurred around 1986. I think it's still a DOS tool. If enough people encourage Neil Larson, the MaxThink developer, maybe he will port it to Java. Putting aside the marketing solicitations, Larson's goals for MaxThink sound very interesting.

- Outlines for Success: Computer Outlining and Teaching Philosophy
- The Philosophy of MaxThink

Outlining is powerful; outlining is organizing information into a structure. When non-bookish hypertext enthusiasts criticize "linear writing" it shows that they don't really understand that books and online documents both contain a defined structure, and a defined, comprehensible, tangible structure is the essence of information. If people like jumps and fragments of information so much, they should converse with a lunatic.

These guidelines are not meant to apply to every single case, but they are valid for most systems I've dealt with. There are a lot of so-called "Contents" and "Index" pages out there (on the Web and in online document systems) that fall far short of usefulness because inadequate detail. For example, a page labelled "Contents" or "Index" that only contains 10 items, in alphabetical order. (Thanks for trying...) As a user, I have wanted to see a hierarchical view of websites and online help systems in just about every case.

The hierarchical, indented, detailed, text-based outline is so powerful for finding information fast and for identifying the scope of a document or web site, every author should use it. It's well worth the effort to maintain a complete Contents page, because the page is the both the author's and the visitor's handle on what the heck the site contains, and what the total layout is. It's also an extremely convenient single page with direct jumps to each page in the site. Whether for books or online documents, the detailed ToC is a dependable way to organize a complete information structure for both the author and the reader. Hopefully, detailed tables of contents are the future!

Definitions of the Standard Information Access Mechanisms

These poorly understood terms should be totally obvious, but even Microsoft completely botched their terminology for WinHelp 3.x. Hypertext is not so absolutely unprecedented that people can abandon the conventions of terminology. Hypertext can not progress beyond book structures by abandoning book structures. Hypertext must embrace book structures and enhance them beyond what they have been. On the other hand, we have much to learn about taking full advantage of the potential of hardcopy information devices. Here is a fresh look at the essential structures or technologies that constitute both books and online documents.

These are structure and navigation components of all documents, whether hardcopy or hypertext. Understanding them may help you design effective documents that can be used quickly. Every component can be implemented in either medium -- hardcopy or hypertext -- and you should know how to translate between the two.

Information Retrieval Devices

Table of Contents (ToC, Outline)

This appears at the front or top of a "document". Should include H1 *and* H2, H3, possibly H4 and H5, for *every* page or section in the document, book, or web site. It is *hierarchical* and it is *long* and *complete*. Not just 5 or 10 items! For that, see "Overview".

High-Level Table of Contents (Contents at a Glance, Contents Overview, High-Level Outline)

Indented outline of a large document, placed near the front of the document. Shows heading levels 1 and 2. Each item is a phrase or section heading. A terse equivalent of the Overview. Appears before the Detailed Table of Contents, if both are included.

Overview

Page in the front matter that provides one prose paragraph per chapter, or per level 1 heading. Lists only about 10 entries, typically the level 1 headings; that is, the chapter titles. Example: home page, or survey of a book placed in the front matter before chapter 1. A home page is a cross between an Overview and a Short Table of Contents. A sentence form of the High-Level Table of Contents.

Detailed Table of Contents

Indented outline of a large document, placed near the front of the document. Shows heading levels 1 through 4 or 5. Appears after the Detailed Table of Contents, if both are included.

Partial Table of Contents (PToC)

Occurs at the start of a chapter. Online equivalent: a page full of links, a Gopher menu page, a high-level WinHelp page that serves to direct you down to the destination pages.

Heading

A phrase or concept that is crafted to be a meaningful pointer to a section *when viewed in isolation* or when viewed in the context of a table of contents or a help lookup screen.

Full-Text Search

A string-search within the main body of the document. Can find a specific multi-word string anywhere in the book or web site. There are other several types of search, though -- search within a Table of Contents, within a traditional topical Index built from handcrafted entries for each section heading, or within a database keyword index built by automatic guessing.

Database Keyword Search

A string-search within a keyword file that is generated based on automatic guessing of keywords in the body of the document. In the first stage, the titles, headings, and links of all the body pages of the web site are searched for important-looking terms, and an "index" file is generated. This is different than a traditional topical, indented index. In the next stage, a string-search is done within the "index" file. Thus, this is neither a full-text search, nor a traditional index.

Index

Occurs at the end or bottom of a document. To build an index, go to each topic, page, or section of a document, and think of 3 terms by which a user would try to look up that topic. An "Index" is a long, visually scannable list of keywords by which a user attempts to guess which keywords the writer has assigned to each page, section, or topic. The first-level items are left-justified. The second-level items are shown interspersed with the level-1 entries, but indented. Sometimes there are level-3 entries, but they should be rare, because that becomes overly complex and gets in the way.

Master Index

Helps you find what document (book or web site) to search in. Broader and shallower than any one document's index.

Pieces of Information at Various Scales

Library (Collection, Bookshelf)

This comprises multiple documents. For example, a large web site.

Document

This comprises multiple topics, sections, or pages. For example, a medium-small web site, a short book, or a long article.

Book, Help System, Web Site

A large, highly organized document. Includes many pages and the requisite access and organization features.

Major Section

A book with many chapters can be divided into 4 or 5 major sections. The Navigation Bar of a web site jumps to the major sections or areas of the site.

Chapter

A set of pages forming a fraction of a book or other long document. Can begin with a partial table of contents.

Page

An easily scannable rectangle of information, viewable all at once. May contain section headings and dividing lines but presents a solid, consistent visual arrangement of data.

Other Universal Document Devices

Front Matter

Cover, copyright page, table of contents, overview, preface, list of figures, list of illustrations.

Footnote, External Link

A very fast link from the middle of a paragraph to the bottom of a page or to another site. The primary continuity of the page stands without the footnote text. The footnote is secondary to the page and hangs off of the page independently of the organizing structure of the document. It's an independent bit of supporting information or a pointer that is a diversion from the primary organizing theme of the page. Footnotes appear on the same page as the reference to them, so are instantly scannable. They are placed "near" the reference, in terms of lookup time and lookup effort. Footnotes have a low cognitive overhead compared to endnotes.

Endnotes

Footnotes that have been removed from the page on which they are referenced and sent so far away, to the back of the book, that they are prohibitively difficult to use. Endnotes work well enough in online systems, where you can quickly jump there and back.

Back Matter

Appendixes, Endnotes, Bibliography, Glossary, List of Acronyms, Index.

Every Web Site Needs a Detailed Table of Contents

Every web site, no matter how disorderly, contains a hidden implicit *structure*. A simple, efficient tool for representing information structure is the *Outline*. Any set of pages and links can be logically and comprehensibly represented by an outline, as surely as it can be represented by box-and-arrow diagrams. Any document structure (collection of pages) that can be represented by box-and-arrow diagrams can be represented helpfully and reasonably well by an outline. The most common form or application of the outline is the *Table of Contents*, which is practically synonymous with "outline". The Table of Contents (ToC) is a structure-encapsulation or structure-extraction device so powerful that it can present an organized view that reveals the implicit structure of any web site or document.

Practically every web adheres to the common structural design standard of a home page that serves as a Cover and Overview. And there they stop, without providing a ToC. This standard design is all around us, and it needs improvement to bring it into line with traditional powerful lookup devices.

There is *no reason* why hypertext must be associated with "lost in hyperspace"! All it takes it proper understanding and respect for traditional hardcopy devices for information lookup and organization. With navigation maps such as a traditional hierarchical outline -- a good ToC with clear, dry headings -- you *can* always know where you are. No more lost in hyperspace, wading through page after slow page, wondering what's contained in a site.

In *The Web Design Cookbook*, Bill Horton recommends that every web site with more than a handful of pages have a Contents page. He shows an example from his own site, and it lists every page in the site.

It's not that difficult to provide users with a reasonably detailed ToC. Even if arbitrary, imperfect, and incomplete, a detailed ToC is an ideal that we can easily take a major step toward. This suggestion in no way "destroys" the hypertext structure or limits or forces the user. This amounts to adding an additional choice of lookup mechanism for the user, if they want it. I am sure that other users besides me want more detailed ToC's for web sites.

Structured outlines are extremely powerful. Give me a rat's nest of topics, and I can produce an orderly outline. Tables of Contents are the main example of structured outlines. They provide hierarchy and order. People try to use hypertext to liberate themselves from the oppression of order and structure. They say that hypertext "is" an exploration space, like Myst. But what about usability! Don't the readers *want* an obvious, firm order and structure? If you care about usability, then provide a good, detailed, well-worded ToC. You can have a pretty home page, with its overview of main areas of your web. But the first link on the home page should be Site Map or ToC. In the

navigation bar, the first item should be Home. The very next item should be Map or "Contents".

Structured ToCs have a strong serendipity factor. You can see adjacent, related topics, while an online keyword search often hides that valuable structural information from you.

This is "applied theory" of information structuring. Even books have far more potential than people have yet discovered. Authors need an audience-oriented *sensibility* for an audience who wants to rapidly scan through massive amounts of information. To maximize comprehensibility and rapidity of information transfer, you must take the viewpoint of a user who is *in a great hurry to assess, comprehend, and navigate the site*. Ultimately, documents should be like a chip you plug into your brain for instant expertise.

This applied theory of rapid navigation could be addressed in the newsgroups in which I began this topic:

- alt.hypertext
- comp.infosystems.www.authoring.cgi
- comp.infosystems.www.authoring.html
- comp.infosystems.www.authoring.misc

Web authors have almost completely ignored the indented, hierarchical outline device. Not through a careful design decision, but through thoughtlessness, a lack of insight driven by the excitement of novelty. The power of the detailed outline is unsurpassed. Hypertext in no way obviates or supersedes the traditional index. In almost all sites, this type of device would greatly enable me to find my way around. I am slowed down, hindered by the lack of robust use of this type of outline/contents device. The site designers provide only a high level outline.

Nearly all web sites force visitors to stumble from page to page, never sure what-all is really included in the site at any decent level of detail. All these sites would greatly benefit from a detailed, structured list of what pages are contained in the site. Exactly what level of detail they put in the ToC or PToC is up to them, just like the decision of whether to include H4's and H5's in a book's ToC. The addition of a true ToC is not an entire design methodology for the total layout of all sites. This is only a recommentation of adding one particular device to whatever structure web sites have, to speed up surveying the site and accessing the particular pages the visitor is interested in.

Larger sites or collections of sites will need a complete ToC for each area of site. For example, the IBM web is a whole collection of areas. This would be comparable to a library of books, rather than a single book. Any 100-1000 page book needs a single ToC in the front, as well as possibly partial ToC's at the start of each chapter.

Information Access in the 3 Types of Documents: Fiction, Nonfiction, and Reference

Some book users find that they use Indexes rather than Tables of Contents (ToC's) in books, so they assume that ToC's in webs would also go unused and that the current navigation pages and search engines are sufficient. But a more detailed comparison of the info-access mechanisms in books and webs reveals that a search engine is not a replacement for an Index, and does not fulfill the need for a detailed ToC for most web sites.

Book conventions could use a lot of improvement, and mediocre book Tables of Contents should not be treated as an ideal model for hypertext maps. Many books have insufficiently detailed Tables of Contents, listing only heading levels 1 or 2 instead of going all the way to levels 4 or 5. It's no wonder people have given up on Tables of Contents, since so many only show the chapter titles, or the level 2 headings with cleverly ambiguous wording designed to be "mysterious and intriguing".

Part of the wrong assumptions about "how books are used" is based on assuming that all books are accessed like computer reference books. What "books" are all about, and how the information in them is accessed and mentally organized, depends on what type of books you have in mind.

The usefulness of a ToC depends on the type of document.

The three primary types of documents:

1. **Fiction**
2. **Nonfiction**
3. **Reference**

Of these categories, general nonfiction most needs a detailed ToC or outline. Most webs are more like general nonfiction than fiction or reference. So, most webs would benefit from a detailed outline as much as general nonfiction books do. I don't go to the ToC when I am looking for a specific piece of information. But I read the ToC in the store and I study the ToC of most nonfiction books. Without a ToC, it's hard to know what topics a book covers. I read and buy books on many nonfiction topics, such as cognitive science, philosophy, psychology, religion, computers, cybernetics, philosophy of science, history, postmodernism, and business. The ToC is a main way of determining the scope of such a book. If you read mostly fiction and reference books, you might think of navigating webs the same way you think of reading a fiction or reference book, which would be largely a mistake. That is probably where the half-baked ideas of the pop hypertext theorists come from. If you primarily think in terms o f fiction and reference books, you might emphasize the wonderful "exploration" metaphor and the "powerful database keyword search" metaphor. But for general nonfiction scholarship, neither the hyper-fragmentation & exploration, nor the keyword search metaphors are so relevant. Organizing and explicitly representing structures of information becomes more relevant, and a detailed

ToC then seems essential.

Hypertext design conventions have been established by two types of people: fiction-oriented literary types, who like exploring and wandering, and technical types, who need to look up terms. But for general purposes of scoping out a site, visitors do not want to be lost in hyperspace, nor do they want to go straight to a keyword entry field. They first want a complete and reasonably detailed synoptic overview of the whole site in see what's there, before they set out to specific destinations. The home page or navigation bar does not provide this detailed overview of the site -- it only shows the names of the 4 major regions of the site.

Exploring Is Too Slow, Searching Is Too Unstructured

A more detailed ToC would enable everyone to cut down on the number of irrelevant pages that a visitor is forced to explore, at 15-30 seconds per link. There is surely a universal demand for this. It's not just an aid for looking up information you are seeking out, but rather, an aid for assessing the overall scope of a site, and then picking out and jumping directly to the specific pages that sound most interesting.

The current search engines are helpful for some purposes, but they can't provide the clear view offered by a structured, detailed map that shows levels 1 through 4 or 5 of a web site.

The Need for Better Site Navigation Tools - Such as a Genuine Table of Contents and Index

Even the "well-designed" sites are very badly designed as far as retrieving information and finding your way around. These sites are elegantly mute about what they contain. 5 pretty pictures on the home page does not constitute a navigation tool. This mysterious challenge is really easy to solve, as follows.

Every site should offer a reasonably detailed ToC and Index. No one would be lost in hyperspace. The ToC can also be called a detailed Site Map or detailed Outline. Providing a reasonably useful ToC and Index would be easy with tools to automatically generate the ToC and Index. The "keyword search" tools can be used but people should actually *show* the Index document, because visually browsing the index is much better than having to blindly enter various keyword search terms. The tool should generate well-formatted Contents and Index html pages. People like first seeing a simple home page with a few main links. The next page offered should be the detailed ToC. At the bottom of each page should be a 'search' and/or 'index' link, taking you to a true, browsable Index page.

Some sites take a feeble step in this direction with an "outline" page... but instead of the 5 icons on the home page, they list about every fourth document -- this is equivalent to a ToC that shows levels 1 and 2 but not 3, 4, or 5.

The ToC or Outline is a well-established, powerful, necessary type of device that has been abandoned by the more flakey hypertext enthusiasts, without good reason or fair consideration. They have not risen to the critical standards they should uphold as innovators in information technology. "Contents" means a topically structured outline, an organized topic tree with sufficient detail to get you very close to the page you want, by rapidly skimming this tree with your eyes, as quickly as your eyes can move. You would destroy the lookup speed of this tree if you broke it apart and put each branch on a separate page, which is exactly what the web convention has done.

If web authors are looking for a way to help the visitors find their way about as quickly as possible, then they must provide some sort of sufficiently detailed and structured map of the site. Call this device what you will, it is functionally a ToC or Outline. The author is free to experiment with some other approach. We've tried these other experimental approaches, which amount to providing a very shallow and unstructured free-form collection of 10 or 15 links serving as an "Overview". Then they label this overview "Table of Contents", but the label promises far more than it delivers.

This experimental convention is a failure. You can decide "I'm a creative designer, I'll eliminate the obsolete Table of Contents device and cleverly invent a new, advanced approach!" but it won't work. Authors are free to experiment with novel layouts of hypertext navigation pages, but they are deceiving themselves about the novelty of their device and about the accessibility of it. The hypertext homepage is not a new device -- it is equivalent to what has been properly called an "overview" for centuries, and suffers from all the limitations of an overview, compared to a detailed ToC or Outline. In their misguided effort to make their site "accessible" by eliminating the ToC and providing only an overview, web authors have obscured what's in their web, they've hidden it from their visitors, thinking that they are so clever, forcing the visitor to click around, groping in the dark at 30 seconds per page.

Detailed tables of contents are just as universally appropriate for all web sites as they are for all books. We do not call book designers unimaginative because they all include a ToC. Web sites are largely equivalent to books, and all books, at least nonfiction books, need a ToC. If you take the ToC out of a book, the usefulness of the book would drop drastically. The same is true of webs. The ToC is a universally important info-access device. The ToC is not an odd device only appropriate in a handful of cases. Abandon this ideal device, and you abandon global rapid, organized access to your pages. You're welcome to experiment with leaving out the ToC, but we've seen the results of this. We've given the exploration design-metaphor a chance, the experiment has been run on us guinea pigs. The experiment is a failure. Now there is no doubt that we could find the pages we're looking for in any site much faster if there were a detailed, structured, and detailed ToC that goes down to the page level o r at least the H4 level relative to the home page.

Adding a ToC or Outline page is not a total site layout design to make every site structurally identical, but merely a navigation tool that should be added to whatever structure you have chosen already, for those users who want to see that view. This page is

a universal type of organized view which should be available for any body of information.

Use Unambiguous Headings, Meaningful Out of Context

This is important in Tables of Contents, Windows Help systems, and Web Links. The reader often sees your heading out of context, away from the content of the page. In the Index window of Windows Help (a true Index window), when you click on an entry, if there are several topics covering the entry, a list of the topic headings appears. Those headings need to be clear, so you know which one to pick right away, without having to take the time and mouseclicks to explore both.

Even computer books sometimes use "cute" and "clever" headings, which are irritatingly ambiguous when you are trying to look up a piece of information that is stopping your whole project. One of the *main* principles of technical communication is *eliminate ambiguity*. For example, the book *Foundations of WWW Programming with HTML and CGI* contains good information, but it's obscured by the use of clever double-entendres in the section titles. Headings in a reference book should not be treated like newspaper headings; the headings should have exactly one, literal meaning. Otherwise, it's just annoying, additional cognitive overhead. In the effort to make the book chatty and pleasant, they make it more difficult to identify the relevant topics.

The chapter "Testing and Installing CGI Applications" contains the following headings:

- It's a Dirty Job, but...
- The Root of All Testing
- An Alias by Any Other Name
- The Act of Creation is Best Done in Private

Thinking once is bad enough. But ambiguous headings make me think twice, or thrice. The competitor book, *HTML and CGI Unleashed*, uses refreshingly clear headings.

The slightest change of phrasing can resolve or obscure which meaning you have in mind. Another principle is "be engaging", but often people try to be engaging by being cute and chatty. Apparently, some readers like this. But for reference works, it gets in the way. The O'Reilly computer books and the magazine *Scientific American* are good models of fulfilling both principles: Be Unambiguous, and Be Engaging.

Include Both a Short and Long Table of Contents

Technical books often start with a short ToC, called Contents at a Glance, listing levels 1 and 2, and then show a Detailed Table of Contents listing levels 1 through 4 or 5. The Short ToC only shows level 1 and 2 items, as a kind of overview. This is distinct from the Overview section, which effectively presents a level 1 ToC using prose paragraphs.

For web sites, you can effectively provide this by using the home page and a few other high-level pages as an overview. The navigation bar also acts similar to a very high-level

ToC or Outline.

Include Level 3 and 4 Headings in the Contents, Not Just Level 1 and 2

Most important is the detailed ToC or Outline. If you don't want two tables of contents, just include the detailed version. Do not include only a short ToC -- this would impede the ability to scope out the book. The ToC should be informative on its own. I was dismayed to find that Kevin Kelly's book about systems theory, *Out of Control*, lacks headings other than H1 chapter titles. Where there should be headings, there are only decorative dividers. The structure of his book is therefore impossible to comprehend without careful study and analysis. The book probably has a detailed structure but he has not revealed what it is. Thus the book reads like a lecture which is not accompanied by an outline -- paragraph after paragraph of sentences, with almost all the structure implicit, not explicit. There are no cognitive handles at the middle levels, only the very high level chapter titles and the detailed sentences. You could skim the index to see a jumbled list of key words that the book contain s, but this cannot substitute for an organized ToC that visually represents the topical structure. It is difficult to respect the caliber of ideas in a systems theory book when the author doesn't think it worthwhile to take advantage of the power of headings and a detailed contents. I would sooner buy a book with a detailed ToC than with only a list of chapter headings. A ToC is so valuable, I have often written out the full ToC to comprehend the structure of a book.

Making Your Visitors Lost, to Force Them to View Your Ads

Some people are afraid that if you show visitors how to jump through your site without delay, you will lose them. You should withhold navigation tools and hide external links from your visitors. This way, you trap your users and force them to view more of your pages. Your visitor is caught, and forced to click through many of your pages. Perhaps we should design our webs so that the user is forced to purchase a product before they are permitted to leave. Myself, I will just press Alt, g, h -- Go Home -- to go to Yahoo and then out wherever I want. So much for the strategy of deliberately hiding your site structure from your user to "force" them to look at pages they do not want to see.

But Providing a Map Would Ruin the Thrill of Unstructured Navigation

So this is what it comes down to. The thrill of exploring a space and building your own mental model of the site layout, hacking through the jungle. Helping people find their way around would ruin the "fun" of click-and-wait, click-and-wait...

Would you drive your car without maps in the glovebox? Even the shoot-em-up maze game Doom has maps available -- in fact, it was designed so that you need a map to play

it. Exploration is fine for video games, but we should not make routine document navigation more work than it should be. For a few webs, making the visitor lost and making them grope about, struggling to comprehend the site layout, is the main purpose. But for most webs, people are not there to enjoy the challenge of building their own mental model of the site. They would rather than the author do that work for them.

For the majority of web sites, comprehending the layout should be effortless and transparent, not obscure and challenging. The exploratory-hypertext-novel approach is inappropriate for most sites and merely annoying, when you are just trying to figure out what the site contains and how to jump straight to a particular page.

If you are in doubt, survey your users: "Would you like us to offer some sort of detailed outline view of the site, or would that ruin the fun of exploring the site?"

Tools for Automatic Generation of Detailed Tables of Contents

I'm looking for CGI tools to automatically generate well-formatted Table of Contents (ToC) and Index pages. I might even write or modify some code myself. I'm really into organizing information. There are probably some CGIs out there that I could modify. For all those sites out there that should have these navigation tools, I would like to be able to point a spider at the site and generate a ToC and Index page to find my way around and get a good handle on what's really there. The spider could also transmit these two pages to the site's webmaster and encourage them to include the pages so people can find their way around the site. People will be almost forced to do this as web sites become *much* larger than they currently are. Everyone's site is going to become like a substantial book. The current navigation tools are not nearly organized enough.

HTML tools so far have been *page* tagging tools. But the next generation of tools, such as FrontPage, is comparable to FrameMaker: tools for handling an entire *site* or book as an integrated entity. These tools will automatically generate a detailed ToC or Outline, and authors should include that genuine contents page to speed up navigation within their site.

- **FrontPage** generates comprehensive lists of pages to help you manage the site.
- **CyberPilot** Pro includes **NetCarta WebMap**, which produces a WebMap(tm) listing all the elements constituting a site, in a logical order. All graphics, applets, pages, and so on are listed.
- **CLEARweb** (Windows) organizes URLs into outlines and flowcharts. This is one mechanism that people can use to publish outlines of their web site or other document collections. A beta test is available at http://www.clearsoft.com/clearweb.
- **htmltoc** from http://www.oac.uci.edu/indiv/ehood/htmltoc.doc.html is a perl program written by Earl Hood to generate linked ToC's based on heading tags.
- Also look into **Vizion** and **WebAnalyzer** from The Ultimate Collection of Winsock Software. I installed them but haven't had the chance to see if they really

map out a site.

- For a Java implementation of a detailed site outline, check out the "Navigator" at the Gamelan Java applets site.
- **Pandect.pl** is an indexer tool similar to htmltoc, written for a class as part of a paper on multiple indexing. It allows for /multiple/ indices as well as glossaries and related features. It's still in its very primitive stages (more features are planned), but is quite operational. It's a small PERL script (4K).

Hypertext Should Embrace and Extend Book Devices, Not Reject Them as Superseded

The topic of information design, information structuring, and information access mechanisms is extremely interesting. Few people seem to perceive the untapped potential of common and ordinary information-access mechanisms. Instead, they approach hypertext as a silver bullet that will completely replace all previous structures associated with evil, supposedly linear books. It is particularly disturbing that the pop hypertext theorists do not understand how books work, so they prohibit by definition any information access structures that have been used in books. The results fall short of what online documentation should be. "Hypertext" has been both oversold and *online document access* has been underutilized -- because it has been treated as a "total alternative" to "linear" books and all the structures associated with them, instead of a powerful extension of book devices. Hypertext has been dissociated from the power of book structures. Hypertext should not compete against the old book devices, but embrace and extend them.

In the effort to promote hypertext, all traditional info-lookup structures have been uncritically tossed aside, rejected in a fervor of experimentation with the radically new. But this radical experimentation has left the users lost groping about through documents, wandering lost in hyperspace with no effective map. Spaghetti links and hyperfragmented pages act like a book with no glue -- open it up, and all the pages fly out into a chaotic pile on the floor. A rat's nest of gimmicks, with no firm and definite structure, is fine for the hip literary criticism professors and the Myst explorers. But most readers want to quickly comprehend the layout of a site or hypertext document, to make a coherent assessment of which pages match their interests.

When something is better than something else, that usually means that the new thing does everything the old thing does, but with additional capabilities. The new does not supersede the old by entirely discarding the old in all aspects, but rather, the new picks up all the capabilities of the old, and then adds much more capabilities. The new should be

backwards-compatible with the old. The hypertext systems we've had so far, which have only been experimental, make the mistake of abandoning not only the limitations of the old, but the strengths, too. Hypertext systems can and should implement hypertext-based contents and indexes that in fact work at least as well as book contents and indexes. The pages that have been called "contents" and "indexes" have not measured up. The fanatical contingent of hypertext mavens have claimed that hypertext is inherently an 'advance" over books. But hypertext is *not* an advance over books if the user cannot find information as *fast* as they could with a well-designed book.

The Myth of the Linearity of Information in Books

It is not true that books are basically linear and online documents are basically nonlinear. The information structure in both cases is nonlinear. It exists in information-structure space. An outline or ToC is a universal way of organizing this information. There are more than one different possible "linearizations" of a basically nonlinear or multi linear document whether that document delivered via hardcopy or softcopy. A linearization by no means forces the reader to follow a sequence -- readers in fact do not read hardcopy linearly. This linear sequence is not necessarily even a suggestion for reading order. The purpose of a ToC is not to show a reading-sequence. It is to reveal the logical structure of the body of information by providing a partly arbitrary but at least definite and comprehensible portrayal of the sections.

There are some popular, bunk ideas going around about the "linearity" of books. A hardcopy book is not linear. The sheets of paper are numbered linearly, but the information structure is distinct from the pages that the information happens to fall on. An information structure *whether delivered via hardcopy or online* is nonlinear. An info-structure exists in abstract space. You can and should provide a "linear" tree-structure, or even multiple alternative structures. Just because the outline/ToC is linear, and the sheets of paper are linear, does not imply that the information structure in a book is linear. People who claim that hypertext is nonlinear "unlike books" do not understand books.

An Outline Is a Tree, Revealing a Logical Structure -- It's Not Linear

A ToC doesn't necessarily have anything to do with a "good logical reading order". A ToC is a structure of topics, not a sequence of navigation. A good ToC doesn't imply some reading order, so much as show a general logical relationship among all the topics. Sequence of reading is only one item to consider when arranging the topics in a logical structure. The important issue for automatic ToC is, how good of a job has the author done at arranging a logical tree of topics. The automatic ToC then portrays this tree structure, for better or worse.

There are multiple possible outlines. It is futile to try to provide the absolutely best outline. Every outline of any book or online document is somewhat arbitrary. You can

even provide multiple outlines, arranged by different heuristics. This is similar to the idea of printing database reports sorted various ways. Just because there are multiple presentation schemes does not mean that any one scheme is invalid or detracts from the value of the database system. Putting headings into a definite order in an outline does not ruin the hypertextual nature of the document. It has been claimed that by definition, a legitimate hypertext document does not, and may not, reveal any particular ordering or structure. This nebulous, open form of document is supposed to be the most empowering form of hypertext, but in practice it can end up as an impractical, unusable, postmodern chaos. As long as you provide a reasonably good outline, this will help you and the reader handle the document. The only hypertext q uality you will lose by providing a detailed outline is the adventure of being lost.

Readers never read nonfiction or reference books from cover to cover; the ToC *never has been* the "reading order". This is the Great Hypertext Lie: that hardcopy is rigidly, linearly read from front cover to back cover. This never was true -- hypertext prides itself on liberating us from a jail of linearity that never existed.

Emphasize the Logical Structure of Information, Not the Mere Medium

Web authors are not doing anyone a favor by withholding the details of the logical structure of the site. Most hypertext enthusiasts have taken a mistaken and unfounded attitude. You should not think of providing a specific map as "forcing a reading order". It is enabling the user to scope out the document, to then jump straight to the most desired page or region. Providing a road map does not force people to drive anywhere in particular. It enables them to pick where they most want to go, and get there. Talk about force; it's when you do not provide a map, that you force something upon the user: you force them to flounder around and "explore" -- even when they want to have the *option* of a map.

Adding a detailed ToC or Outline is not a complete site design structure, nor does it amount to forcing readers to take a certain view of the site. If a visitor should happen to want to see a logical, comprehensive outline of the site, they should be able to access such a view. Many visitors would gladly view such a detailed, structured outline page, because it's a fast and effective way of comprehending the scope of a site, comparing the headings, and jumping straight to the heading you want.

Revealing a logical structure of a site does not "force" something upon the readers. In fact my proposition gives the reader more freedom -- they can choose to view or ignore the detailed ToC or Outline. It's when you leave out a detailed ToC or Outline that you "force" the reader, by restricting their navigation choices and making them have to explore and view pages they might not want to view, at 15-30 seconds per page. Providing a ToC is about enabling the reader and giving them more choices, not about forcing them to view something they don't want to.

A Table of Contents *Helps* Users Jump All Throughout a Document

Providing a "linear" outline permits more hyperlink jumps, not fewer. The link page contains more links than most other pages. Adding linear handles does not eliminate hypertextuality. The Table of Contents cannot constrain readers to follow a sequence. Rather, it inherently helps readers jump all throughout the document.

Whether for books or online documents, the ToC is somewhat arbitrary, but at least it is comprehensive and enables immediate comparisons of all available headings or topics, and direct jumps to those topics without being forced to navigate through less interesting pages while the clock ticks. If readers don't want to look at the ToC, fine, they don't have to. You should permit them access to some kind of orderly, comprehensive view of the site or document, if they want it.

A "Linear" Detailed Outline Provides More Information and More Choices for the User than No Detailed Outline at All

Despite the common implications to the contrary, adding a linear organizing view (an outline) cannot detract from a hypertext site or "force" the reader to view a particular sequence. Authors think they should not "force a reading order on the user since the document is intended to be read as a hypertext document rather than as a linear document". This popular conception of document structuring is deeply mistaken about the nature of information structuring, writing, and documents. Providing a ToC does not mean that the reader is going to read the document from cover to cover. Hypertext theorists imply that nonfiction books are actually intended to be read from cover to cover, which has never been true.

A very high, top-level ToC or Outline contains far less structural information than a ToC that also shows more levels. It's fine to highlight your level-1 topics on the home page or contents-at-a-glance, but you should also add a more detailed ToC for those who want it. If you give people the choice of whether to view the ToC, and then monitor their path, you will find that they all will freely choose to view the ToC, because this is easier than the tedium of having to explore. People don't want to explore. They want to quickly, effortlessly scope out and go directly to the pages which have headings that most interest them.

How to Deliberately Define Index Keywords in a Web Page

Use the META Keywords Tag

Use the META tag. The search tools build indexes by looking at the title tag, heading tags, links, and meta/keywords tag. This element is poorly documented and is unsupported on many servers.

<HEAD>
<TITLE>Observations about Felines and Autos>/TITLE>
<META HTTP-EQUIV="Keywords" CONTENT="cats, cars">
</HEAD>

Pack Your Titles and Headings with Unambiguous Keywords

Search engines generally guess the keywords for your site based on the titles, headings, link text, META-keywords, and the home page text. Write all of these deliberately to make your site show up in the appropriate global searches. The first paragraph and headings on your home page need to contain all the keywords you can pack in. Consider including the region you are in. Often, I only want hits from companies in my region. Try to prevent your site from appearing as irrelevant noise in many searches.

The Powerful Visual Layout of the Traditional Index

A traditional index enables you to visually scan and compare nearby entries. An index is a device by which the reader attempts to guess what keyword entries the author assigned to a topic, section, or page. The term "index" has been weakened through careless use, taking advantage of its broader meaning. Some sites use the term 'Index' to refer to a single-level alphabetical list of main subject areas. A real index is highly structured for rapid visual scannning. There are 2 or 3 levels of visual indenting, and items are alphabetized.

cars

- musical group -- see *The Complete Guide to Rock*
- new
- roller coaster -- see *roller coaster*
- used

felines

- cats
- lions

The conventional index is topical, alphabetical, indented, and visually scannable. This is different than the database keyword type of index, which requires you to blindly guess at available search terms.

Overuse of the 'Overview' Structure and Links within Sentences

Example of overuse of the Overview structure: The site for No Dogs or Philosophers Allowed. I wanted them to "just show me the schedule, *now*!" But they offered up paragraph after windy paragraph of *full sentences* -- when a ToC or detailed Site Outline would have been a thousand times easier. I was almost *unable* to find the page I was looking for! I could have found it in 5 seconds, if they used the ToC structure I am asking for.

An Outlined List of 1-Line Headings Is Faster than Links in Sentences

Most overview sections merely express the clean contents headings in the inflated form of a sentence: "In the Configuring Widgets chapter, we give you a nice introduction that we wrote. We tell you all about the configuration of widgets, so that you know how to start preparing to learn all about Widget Configuration."

Traditional book structures are so superior to the many online systems I've seen. The only feature that most online systems have gotten right so far is Search. But Search is no substitute for hierarchical, complete contents and indexes. Neither is the "home page" -- a truly *feeble* info-access device. I think that overview sections in books are absurd and unjustified -- if you want an overview, just look at the contents.

The "overview" section or wordy home page could be replaced by a sleek *cover*, or elegantly simple welcome screen, combined with a good ToC that has one-line headings that are so clearly worded, they don't *need* to be shown in the context of a paragraph. The link headings should be self-explanatory.

The Overuse of Prose Paragraphs; Use Lots of Visual Cues and Visual Structures such as Lists and Headings

Philosophy is a good example of a highly inefficient structure of written information. Prose paragraphs without headings, diagrams, tables, and lists are the worst-case information transfer rate. You can see this especially in transcriptions of university lectures prior to mid-century. Prose paragraphs are most appropriate for verbal communication. But written material should use lists, headings, and outlines. Written

communication lacks the immediacy of verbal communication, so it should compensate by taking full advantage of visual information layout by breaking topics into several levels of headings. Each heading level should be obviously above or below the other levels. For example, the level 1 or 2 headings should always have a line breaking the page, and should use a very large font extending to the left of the body text.

Even this article has too many prose paragraphs. I would like to convert much of it to add more headings, examples, lists, tables, and diagrams, and organize them. In the process, the file size could probably be cut to a tenth the size with no loss of information. Visual richness does not necessarily require graphics. Even a plain text page can be much more visually rich and scannable than prose paragraphs. The goal: maximum information transfer rate from author to reader. Lists and headings are closer to pictures than are sentences.

The Home Page is Limited to a High-Level Overview

The Home Page is functionally equivalent to the traditional Cover plus Overview. The first link on the Home page, and the second button in the navigation bar, should be Map or Contents.

For example, see the example Map link at the bottom of this page. No explanation is convincing without a clear example or model.

I hope that everyone else copies my ToC structure. And I hope we can soon use string-search and a well-designed topical index on *every* site.

Use Visual Formatting: Lists, Headings, Diagrams, Spacing, Rules, Fonts, Indents, and Icons

With visually monotonous paragraphs, the eye gets lost. Provide pictures and distinctive elements so that your eye says **"Hey, I recognize that section. I know where I am."** Visual landmarks, or visual changes, act as handles for the memory and eyes. These visual grip-points support skimming, moving through the document, and returning to familiar locations.

Forcing Spacing with HTML Tags

To force blank lines, use <p>
<p>
 tags. <dl></dl>also works. There are many creative workarounds to force HTML to produce good visual layout.

To force spaces between letters, use   tags.
To indent, you can use nested <dl> tags.

An award-winning HTML hack example site: Severe Tire Damage on the Information Superhighway.

The usual trouble with workarounds is that they have side-effects. Using a 1-pixel graphic may work invisible wonders for spacing when viewed with "load images", but this turns into a generic picture-icon when viewing the page with graphics turned off. You might try to implement indenting in a multi-level list by using several   characters to force an indent, but if the line wraps, it might continue all the way over on the left margin. Tables give you mighty page layout power, but you have to consider the browsers without tables. Make the best visual formatting you can while using the simplest, most reliable tagging you can.

Visually group related paragraphs together, using a degree more or less spacing. The eye and brain then mentally group and associate the paragraphs to the appropriate degree. If two groups of paragraphs have a very large spacing separating them, you need to create a separate heading. This variable spacing technique works well in email and newsgroup postings.

A style sheet a company had defined for help systems resulted in headings, paragraphs, and indented lists all jammed up together, rigidly single-spaced. Tweaking the 'space above' and 'space below' values enabled me to sneak in the visual formatting that is essential for rapid comprehensibility.

In a long document like this one, with many h2's and h3's, consistently use a horizontal line above the h2's. Sometimes the line follows a paragraph of the preceding section too closely, and the eyes get confused. <p>
<p>
 or <dl></dl><hr> forces more space above the line.

Provide Full-Text Search

3rd-party computer books have pretty decent ToCs and Indexes. No electronic search, though... even though these computer books include a CD-ROM. One hardcopy book comes with full-text search on CD-ROM: *Electronic Publishing Unleashed*.

Many of the "Search" tools on the web are not a substitute for indexes because they are not full-text search. On the other hand, printed indexes often omit references which would be helpful. Writers should make full text versions of their document available. There is plenty of room on the CD-ROMs that come with 3rd party computer books. You can often find your information in a fraction of the time it takes to use the Index and

Contents.

Reduce Mouseclicks and Cognitive Overhead: Why Shorter, Highly Granular Pages Aren't So Great

If a little hypertext is good, then extreme hypertext must be perfect, right?! No, overly granular bits of information lose structure. It takes too long to jump from one page to the next. The information becomes completely dis-integrated. Information must have a **visually obvious, scannable structure**. I can very rapidly scroll down a page, comparing several sections and headings all at once. When you break out a section as a separate page, you lose the ability to view it inline with the neighboring sections. Linearity is powerful -- it puts blocks of information together firmly. Is linear sequence oppressively rigid? As a reader, it's easiest for me if a document has a firm, reliable structure that holds still so that I can scan it and remember it. Once you click from one topic to another, the first one is gone... that's bad. Out of sight, out of mind. Contiguous topics should be visible simultaneously, and should be placed next to each other. Two separate hypertext pages may be linked, but they are still farther away from each other than two contiguous sections of a scrolling topic. With hypertext granularity, more is not better. Extreme hypertext is extreme *fragmentation*, and the mind finds extreme fragmentation hard to map out and remember.

Long Topics Are Easier to Skim, Print, and Do a 'Find' Within

The connect-time to retrieve a page is a significant cognitive overhead. I sent an email to the encyclopedia on CompuServe explaining that they must not break each encyclopedia entry into a separate page, because it takes longer to retrieve the page than it does to read it! I can't comprehend an article when I have to stop, click, and wait, just to read the next couple paragraphs. They were breaking the articles into tiny fragments of 2 or 3 paragraphs! It takes me longer to retrieve 2 paragraphs than it does to read it, so this insanely granular hypertext chunking made me take *2 or 3 times as long to read articles*.

In the end, *balance* is what counts, not extremist guidelines like "smaller chunking is better". Have these designers never tried to scan large amounts of text? The concensus is that the best HTML guide is A Beginner's Guide to HTML. It is one long document: easy to scan, 'Find', and print.

Bad advice from Matterform Media:

> Keep your pages small. This is only common courtesy. No one wants to

> scroll through a gigantic page. A little scrolling is an unfortunate and occasional necessity. A lot of scrolling is pure thoughtlessness. Particularly abhorred is the custom of creating links that scroll you to a far-away point down on the same page. Why not break that monster up into a number of smaller pages? It's really not that much work and your users will appreciate it.

What I "appreciate" with smaller pages:

- I can't print easily.
- I can't do a 'Find' easily.
- I can't skim easily.
- I have to spend more time navigating than reading.
- I can't view related sections contiguously and handle them as a single entity.

Vertical Scrolling is Far Faster than Link-Traversing

Vertical scrolling actually has the lowest cognitive overhead of all online navigation methods. This is because of the following factors:

- Scrolling is much faster than page retrieval or link traversal time.
- Longer pages with multiple topics support obvious visual contiguity.
- It takes fewer aiming, mouse, and keyboard actions to scroll than to select a link.
- Your eyes skim much faster than your fingers can aim and point. You can take in multiple topics almost simultaneously.

Different levels of your document's hierarchy can have different degree of granularity; some types of pages or sections should be short, and some should be longer. In general, pages or sections at level 1 and 2 should be short, navigation pages, and levels 3 and 4 can be longer sections. Within a web, there are levels of pages. Within a page, there are levels of headings. While some pages should be long, nonfiction pages should be broken up using headings -- but not broken out into separate pages.

Long pages are appropriate for scanning through a great deal of substantial information. Breaking a freight-car of information into cute little fragments each on their own page seems like a good idea to hypertext enthusiasts who aren't used to speed-scanning through books, slowing down when an especially interesting passage comes along. Screen-sized information chunking seems like a good idea until you have to use it for **serious** heavy reading and research. Very soon, traversing links becomes an irritating labor. The dominance of small pages is a fetish for flimsy gadgetry at the expense of substance. They work fine for boutique sites... we seem to have been taken over by a global mall of information boutiques.

You may like or hate this long page, but I know that you are able to almost instantly ascertain its scope and pick up its main points, by reading the structured Contents list and quickly scrolling up and down through the page using only the up- and down-arrow keys

or "paging" through it by pressing the page-up and page-down keys *as fast as you want*. That's my goal with info-access devices such as detailed Tables of Contents and reasonably long pages: to enable extremely rapid scoping and transfer of information. You clicked a link, waited only a few seconds, and were rewarded with some 20 topics full of information. If I had broken this page into 20 separate chunks, you would have had to "explore" and "navigate" using many keystrokes and mouseclicks. "Exploring" is nice to do on a lazy Sunday, but annoying when you just want the information, now. This way, as soon as you load this page, all you have to do is lean on the arrow keys to rapidly scan 20 "separate" topics. A single long page saves you a lot of thrashing around through a rat's nest of spaghetti links, at 15 seconds or more per link. You might have given up reading the document as soon as you hit a lame topic, to save time. But with this long page, if you don't like a topic, you can *very* rapidly scroll to another section, using your eyes to scan the headings as they flash by as fast as you can scroll. The eyes can survey scrolling information much faster than the hands can aim and click. Long pages do not waste the reader's time; highly granular pages do.

Don't be uncritically persuaded by the cheap novelty of hypertext chunking. Traditional linearity is powerful. This debate reminds me of the controversial report claiming that character-mode word processors encouraged more substantial content than graphical word processor/page layout programs. Similarly, the glitz factor of hypertext can detract from the common-sense, practical needs of fast results for information lookup and knowledge transfer.

Static, Stationary, Linear Pages Hold Still Long Enough to Be Comprehended

The exciting novelty of hypertext has caused a reactionary demonization of staticness. This is yet another unfounded, uncritical dogma put forth by the hypertext enthusiasts who try to elevate hypertext by condemning every aspect of hardcopy.

My mind shifts too often as it is; I don't want my information devices to do the same: static structures are good because they are stable and comprehensible. Extreme flexibility is as bad as extreme inertness. Authors need a sense of balance and judgement.

The Limitations of Imagemaps for Navigation

Graphic imagemaps for site navigation have several limitations and costs compared to text-based outlines.

Imagemaps will always be limited to the very highest levels. They are poor information lookup tools as far as speed and detail.

Imagemaps Transmit Slower than Text Outlines

Even a text-based ToC page that is long and detailed will load much faster than any graphical map can.

The Blind Can Handle Text Outlines

People who are blind can handle a text outline easily, but not a graphic map. This is not a compromise of the capabilities of the web, because the text outline is a supplement to graphics, not a replacement. You could also provide a high-resolution virtual-reality map of the site, but to speed up access for sighted and blind alike, you should offer a text outline as well.

Text-Based Browsers Can Handle Text Outlines

Text-based browsers can easily handle text outlines, but not graphic maps.

Imagemaps are Limited to Levels 1 and 2

Imagemaps are limited to levels 1 and 2 of a site; text-based outlines can go down to level 4 or 5 and remain usable.

Large imagemaps showing branches of a web site down to level 3 or 4 will only be feasible when everyone has 10 M bit/sec pipes. For now, you can only include levels 1 and 2 in a graphical site map. More levels would require a huge full-page graphic that would take too long to load.

Imagemaps are Harder to Maintain than Text Outlines

Imagemaps are harder to maintain than a text-based outline. The imagemap must be limited to very high level pages, so it doesn't go out of date immediately.

Many People Turn Off the Display of Images

Some figures claim that the majority of surfing is done with the graphics turned off. For routine information-gathering, people want speed more than graphics. Speed is the worst bottleneck of Web usability. Most people surfing the public web currently have 14.4 K bits/second modems.

Including Both an Imagemap and Detailed Outline

I include a plain text outline of this whole site, as a "site map" that's not a graphic. That outline page just has a link to the graphical imagemap -- that way, people can choose whether to load the graphical map.

Keep the GIF file down to a 16-color palette and keep the filesize around 20-40 KB or it will take way too long to load. You *could* do a full-page graphic, if it is very simple. There are workarounds such as arranging a bunch of smaller graphics on a colored background.

The level of detail acceptable in a graphical map depends on your audience assumptions: what's reasonable at 128 K bits per second won't fly at 14.4 K.

For more about reducing graphics filesize, see the Bandwidth Conservation Society.

I recommend that sites provide both an imagemap showing levels 1 and 2 and a detailed text outline showing pages at levels 1 through 3 or 4. Combined with keyword searching, this gives the user a choice of several powerful methods for navigating your site.

I hope that this article has given you some good ideas about increasing the ability of your visitors to quickly scope out your site or document, compare and identify the page titles that most interest them, and jump directly to those pages with minimal clicking-and-waiting.

Michael Hoffman is a senior technical writer specializing in document structuring and information design. He has written several large help systems and has designed sets of books and help systems that are based on a single set of source files.

See also 7 Keys to Single-Source Documentation.

For more guidelines, see http://union.ncsa.uiuc.edu/HyperNews/get/www/html/guides.html.
