

Web Indexing – An Overview

ASI-Rocky Mountain Chapter
Fall Workshop 2011
October 22, 2011
Sarah E. Myers, MLIS

Introductions

- About me...
 - About you...
 - Anyone with web authoring experience?
 - HTML knowledge? And Comfort level?
 - Content Creation? Web Editing?
 - Interested in learning more?
 - Who has web indexing experience?
-

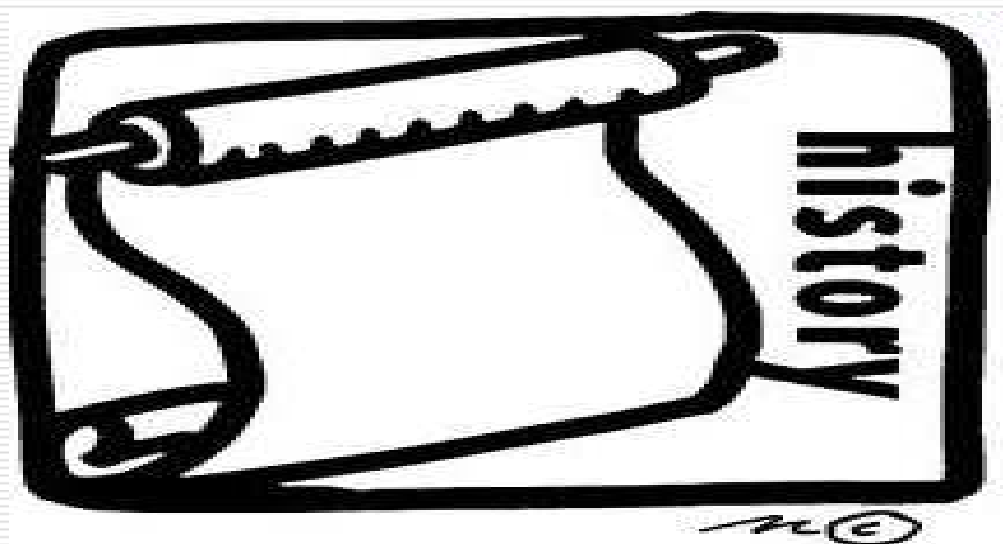
Presentation Goals

1. Overview of Web Indexing:
 - a. Answering the “W” Questions: Who, What, When, Where, Why, How
 2. Where Web Indexing is at today
 - a. Differences between search engines and A-Z Indexes
 - b. Basic overview of Web technologies in concert with A-Z Indexes
 3. Basics in creating a web index
 - a. Tools to get started
 4. Future of Web Indexing
 5. Provide resources to learn more
 6. ASIRM Site – Create A-Z Index
-

Quotes to Get Us Thinking...

- ❑ "Anyone who has lost track of time when using a computer knows the propensity to dream, the urge to make dreams come true and the tendency to miss lunch." ~ Tim Berners-Lee (Inventor of World Wide Web, Computer Scientist, Director of W3C)
 - ❑ "The real problem is not whether machines think but whether men do." ~ B. F. Skinner (Psychologist)
 - ❑ "Getting information off the Internet is like taking a drink from a fire hydrant." ~ Mitchell Kapor (American Businessman)
 - ❑ "If human-crafted indexes have been effective in helping readers find information in millions of non-fiction books and manuals, then they ought to be useful for many content-rich Web sites." ~ Heather Hedden
-

History



Setting the Stage...

A Very Brief History of the Web

- ❑ 1945: Vannevar Bush writes an article in Atlantic Monthly, "As We May Think", about a hypothetical photo-electrical-mechanical device called a Memex, for memory extension, which could make and follow links between documents on microfiche. The concept of the memex influenced the development of early hypertext systems, eventually leading to the creation of the World Wide Web.
- ❑ 1958-1966: President Eisenhower requests funds to create ARPANET for the Air Force and the project moves forward.
- ❑ 1960s: Doug Engelbart prototypes an "oNLine System" (NLS) which does hypertext browsing editing, email, and so on. He invents the mouse for this purpose.
- ❑ Ted Nelson coins the word Hypertext in A File Structure for the Complex, the Changing, and the Indeterminate. 20th National Conference, New York, Association for Computing Machinery, 1965.
- ❑ 1976: Apple Computer founded by Steve Jobs and Steve Wozniak.
- ❑ 1980: Tim Berners-Lee writes program called "Enquire Within," predecessor to the World Wide Web.

Sources:

A Little History of the World Wide Web from 1945 to 1995, W3C, <http://www.w3.org/History.html>

Nerds 2.0.1, PBS, <http://www.pbs.org/opb/nerds2.0.1/timeline/>

Brief Timeline of the Internet, Webopedia, http://www.webopedia.com/quick_ref/timeline.asp

Setting the Stage...

A Very Brief History of the Web

- ❑ 1981: IBM announces its first Personal Computer. Microsoft creates DOS.
 - ❑ 1983: Domain Name System (DNS) designed by Jon Postel, Paul Mockapetris, and Craig Partridge. .edu, .gov, .com, .mil, .org, .net, and .int created.
 - ❑ 1986: 5000 hosts on ARPAnet/Internet.
 - ❑ 1989: 100,000 hosts on Internet.
 - ❑ 1989: Berners-Lee wrote a memo proposing an Internet-based hypertext system, essentially creating the World Wide Web.
 - ❑ 1992: "Surfing the Internet" is coined by Jean Armour Polly.
 - ❑ Late 1990s: Berners-Lee specified HyperText Markup Language (HTML) and wrote the browser and server software.
-

Definitions of Indexing

- "An index is a systematic arrangement of entries designed to enable users to locate information in a document." (British indexing standard BS3700:1988)
 - "A systematic guide designed to indicate topics or features of documents in order to facilitate retrieval of documents or parts of documents." (National Information Standards Organization NISO TR-02-1997)
-

Definition of a Web Site Index

- A Web site index – often called an “A-Z site index” – is a finding aid for a Web site, intranet, or sub-site, organized in the same way as a traditional, alphabetical back-of-the-book index. In addition to the alphabetical arrangement of index entries, the following conventions can be found:
 - Use of one or more alternate or variant forms for entries to provide multiple entry points for user access to the topic, including:
 - synonyms (e.g., “careers” and “jobs”)
 - terms and their acronyms (e.g., “DVDs” and “digital versatile disks”)
 - phrase inversions (e.g., “web site indexes” and “indexes on Web sites”)
- Wording of entries so that the most important part of the term or phrase appears first when alphabetized. (e.g., “information, general,” not “general information”)
- The presence of subentries, indicated through the use of bullets, dashes, indentation, etc.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

Definition of a Web Site Index Continued

- In addition, cross-references (See or See Also references) may be used, but are not required.

- An index can be limited to named entities (such as the names of departments, people, etc.), but – when compared with an index of topical terms – is usually not sufficient for searching a site.
 - On a Web site, a named-entity index is more accurately called a “directory,” like a telephone directory.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

1.2.3. Add, Divide, and Multiply!

- There are multiple ways and tools to create an A-Z Index but the indexing content analysis and process is still the same.



Who



Answering the W's - WHO

Who creates web site indexes?

- Site indexes are best done by individuals skilled in indexing who also have skills in HTML and/or in using HTML indexing tools.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

Answering the W's – WHO

Continued

- Full-time or part-time web position
 - Web team in a larger organization
 - Contractor or Freelancer
 - See Also: [Web Indexing SIG Indexer Search database](#)
 - Volunteer, Student Intern
-

Answering the W's - WHAT

What is a Web Site Index?

- ❑ An alphabetical back-of-the-book style index, in which each index entry is hyperlinked to the relevant place on a web page within a site.
- ❑ An index based on the assignment of controlled keywords (metadata) for each web page, for retrieval by a search tool or within a database.
- ❑ A web index can be an index of an entire site, subsite, intranet, collection of online articles, online course, print-to-web documents, or an e-book.

Answering the W's - WHAT does a Web Site Index Contain?

- ❑ An index contains all significant mentions of people, places, things, concepts, and ideas.
- ❑ An index should lead users to relevant material and to significant content that provide useful information.

Source: Improving Usability with a Website Index by Fred Leise (2002):
<http://www.webcitation.org/5vJwZDVkj>

What Web Indexing is NOT...

- Concordance – list of words
 - But A-to-Z ‘indexes’ often show up under the guises of a directory or ‘glossary index’
 - Example: The Contented Website - Glossary Index:
<http://www.thecontentedwebsite.com/contented-website-glossary.html#GlossaryIndex>
 - Table of Contents
 - Example: Access Web Table of Contents:
<http://access.mvps.org/access/toc.htm>
 - Site Map: a navigation aid that provides an overview of a site's content at a single glance
 - Example: Apple’s Site Map:
<http://www.apple.com/sitemap/>
-

What Web Indexing is NOT...

A Site Map

- ❑ A site map is a list of pages of a web site accessible to crawlers or web users. It can be either a document in any form used as a planning tool for web design, or a web page that lists the pages on a web site, typically organized in hierarchical fashion. This helps visitors and search engine bots find pages within a site.
- ❑ A site index is often used to mean an A-Z index that provides access to particular content, while a site map provides a general top-down view of the overall site contents.

Source: Site Map: http://en.wikipedia.org/wiki/Site_map

How is an index different from Site Map?

... and why wouldn't a site map be sufficient?

- ❑ A site map is a finding aid organized in the same way as a table of contents; it follows the structure of the site, section by section, instead of being alphabetical and is based on major page headings/titles.
- ❑ A site map provides a helpful overview of a site but does not always enable users to quickly find a specific topic.
- ❑ A site map lists each web page only once and by its given name (for example, "Work For Us"), without cross-references or variants to make it more easily found.
- ❑ Site maps tend to include only major web page sections, and usually does not include all pages within a site.
 - Presents challenges for finding non-HTML documents, PDFs, images, etc.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

A to Z Web Indexing is NOT...

Google Indexing – What is it?

- When you sit down at your computer and do a Google search, you're presented with a list of results from all over the web. How does Google find web pages matching your query, and determine the order of search results?
- In the simplest terms, you could think of searching the web as looking in a very large book with an impressive index telling you exactly where everything is located. When you perform a Google search, Google's programs check their master index to determine the most relevant search results to be returned ("served") to users.

The three key processes in delivering Google search results are:

- 1. Crawling:** Crawling is the process by which Googlebot discovers new and updated pages to be added to the Google index.
- 2. Indexing:** Googlebot processes each of the pages it crawls in order to compile a massive index of all the words it sees and their location on each page. In addition, Google processes information included in key content tags and attributes, such as Title tags and ALT attributes. Googlebot can process many, but not all, content types. For example, they cannot process the content of some rich media files or dynamic pages.
- 3. Serving:** When a user enters a query, Google machines search the index for matching pages and return the results believed to be the most relevant to the user. Relevancy is determined by over 200 factors, one of which is the PageRank for a given page.

Source: Webmaster Tools Help - Google Basics:

<http://www.google.com/support/webmasters/bin/answer.py?answer=70897>

Questions Yet?



Answering the W's - WHEN

When can a web index be created?

- The short answer: Anytime!
 - New website development/design
 - Redesign
 - Adding and expanding content
 - Maintenance
 - User feedback
 - Organization requested
-

Where



Answering the W's - WHERE

- ❑ Strong industries & fields for web indexing...
 - ❑ For profit companies with content-rich sites
 - ❑ Education and research institutions
 - ❑ Medical and consumer health
 - ❑ Museums and libraries
 - ❑ Digital (or fully online) libraries
 - ❑ Associations
 - ❑ Non-profits
 - ❑ Governmental (Non-governmental organizations - NGO's and Intergovernmental organizations - IGO's) and Municipalities
 - ❑ Information-based organizations and companies
 - ❑ User guides, technical manuals, software companies, other instruction-based environments
-

Answering the W's - WHERE Continued

Sites Best Suited for A-Z Web Indexes Medium-sized (20-500 pages)

- Static sections of large sites
- Content-rich sites
- Sites with repeat visitors (membership organizations, education institutions, libraries, financial services, etc.)
- Intranet subsites
- Online periodicals or e-zines
- E-learning / online courses
- Parts of sites that have a large quantity of relatively unchanging content, such as help documentation (for example, eBay's help topics) or policy handbooks, can benefit from indexes.
- Collections of articles, although constantly being added to, should also be indexed.

Sources:

ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

ASI Web Indexing SIG Brochure: <http://www.web-indexing.org/brochure.pdf>

Answering the W's - WHERE

Continued -- Not so suitable...

- ❑ Very small or very large sites.
 - ❑ Sites with constantly changing content or products.
 - ❑ Some sites may have a sufficient number of pages, but the content is not suitable for an index, such as online games or short descriptions of directory entries.
 - ❑ A directory-type site might require an alphabetical list of names to look up, but this would not be a structured, topical index.
 - ❑ Even most sites that sell products do not need indexes if most of the pages of the site are a listing of products. Potential customers tend to look up products by category, not by alphabetical names.
-

WHERE – Continued

Examples of A-Z Indexes

- ❑ Colorado State University: <http://www.colostate.edu/>
 - ❑ CSU Libraries: <http://lib.colostate.edu/>
 - ❑ Healthier Communities Coalition – Snapshot: <http://www.healthylarimer.org/snapshot/>
 - ❑ Centers for Disease Control: <http://www.cdc.gov/>
 - ❑ MedLine Plus (Health Topics / Drugs & Supplements): <http://www.nlm.nih.gov/medlineplus/>
 - ❑ US Department of Labor: <http://www.dol.gov/dol/siteindex.htm>
 - ❑ United Nations: <http://www.un.org/en/siteindex/>
 - ❑ City and County of Denver: <http://www.denvergov.org/>
 - ❑ American Society for Indexing: <http://www.asindexing.org/>
-

WHERE – Continued

Examples of A-Z Indexes

Sites **WITHOUT** an Index... and perhaps need one?

- ❑ UNICEF: <http://www.unicef.org/>
 - ❑ U.S. Department of Transportation: <http://www.dot.gov/>
 - ❑ Federal Transit Administration: <http://www.fta.dot.gov/>
 - ❑ World Bank: <http://www.worldbank.org/>
 - But take a look at Topics: <http://www.worldbank.org/html/extdr/thematic.htm>
 - ❑ City of Fort Collins: <http://www.fcgov.com/>
 - ❑ Colorado.gov: <http://www.colorado.gov/> (Watch this one =web portal! See under State Agencies)
 - ❑ Community Resource Center: <http://www.crcamerica.org/>
 - ❑ And our own Rocky Mountain Chapter: <http://asirm.org/>
-

Questions?



Answering the W's - WHY

Why is this Important?

Benefits of a Web Index:

- User-friendly back-of-the-book style, familiar to all users
- Synonyms and variant spellings provided
- More precise indexing with entries linked to anchored content
- Increased search engine optimization due to internal links
- Single-click access to content, compared to using hierarchical categories (taxonomies)
- Overall improved web site usability

Source: ASI Web Indexing SIG Brochure: <http://www.web-indexing.org/brochure.pdf>

Why is this Important?

Continued

- Enhance website usability:
 - A web index offers easy scanning for finding known items, they provide entry points to content using the users' own vocabulary and they provide access to concepts discussed, but not named, in the text.
 - Provides direct access to granular chunks of information without the need for traversing multiple links in a hierarchy or depending on search engine results.

Source: Improving Usability with a Website Index by Fred Leise (2002):

<http://www.webcitation.org/5vJwZDVkj>

Why is this Important?

How does a Web Index Increase Usability?

How do indexes increase usability?

1. Easily browsable.
2. Translate the vocabulary of the users to that of a text.
3. Reveal other topics of interest to the user.
4. "See also" references can lead users to additional or more specific information that might more closely meet their information needs.
5. Can link to precise points within a Web page
6. Useful in "know-item finding," -- where users know specifically what they are looking for (or what information they saw previously and want to get back to).
7. Acts as a complement to the site map, or table of contents, search within feature, and enhances the search engine optimization ranking of the site.

Sources:

Improving Usability with a Website Index by Fred Leise (2002):

<http://www.webcitation.org/5vJwZDVkj>

A-Z Indexes to Enhance Site Searching by Heather Hedden (2005) http://www.digital-web.com/articles/a_z_indexes_site_searching/

Why is this Important?

About Search Engine Optimization

- ❑ An A-Z index can enhance the search engine optimization ranking of the site. A large number of new internal links created and the words within the linked text are high-quality keywords.
- ❑ Search engine optimization (SEO) is the process of improving the visibility of a website or a web page in search engines via the "natural" or "algorithmic" search results.
- ❑ SEO may target different kinds of search, including image search, local search, video search, academic search, news search and industry-specific search engines.
- ❑ As an Internet marketing strategy, SEO considers how search engines work, what people search for, the actual search terms typed into search engines and which search engines are preferred by their targeted audience.
 - Optimizing a website may involve editing its content and HTML and associated coding to both increase its relevance to specific keywords and to remove barriers to the crawling (indexing) activities of search engines.
 - Crawling: Other terms for Web crawlers are ants, automatic indexers, bots or robots, and Web spiders.
- ❑ SEO tactics are often incorporated into website development and design. The term "search engine friendly" may be used to describe website designs, menus, content management systems, images, videos, shopping carts, and other elements that have been optimized for the purpose of search engine exposure.

Use a Search Engine or an A-Z Index?

What are some differences search engine and an A-Z Index?

Off the bat:

- ❑ An A-Z index can take user directly to the point in text where topic is discussed - saves time.
 - ❑ Search engine results may or may not deliver user to direct point in text - spends time.
-

Use a Search Engine or an A-Z Index?

- ❑ In many web sites, a user can search hundreds or even thousands of files with a Search Within feature (search engine).
- ❑ The user may get back hundreds of links, without helpful or useful identification. The user struggles through the list, trying to find a link or two that leads to the desired information.
- ❑ Some results make the user wonder what the link has to do with the word or phrase originally searched for resulting in frustration and less than desired search results.

Source: Why Create an Index? By David M. Brown: <http://www.web-indexing.org/article-brown.htm>

Search Engines vs. Web Index

Drawbacks of search engines:

- ❑ Search engines only pick up exact words or phrases.
- ❑ If a user enters a synonym, singular instead of plural, a spelled-out form instead of an acronym, a misspelled word, or merely a concept with words that never appear in the text, appropriate pages may be missed.
- ❑ Searching the entire Web, missed pages are usually not a problem since so many results are retrieved.
- ❑ But on an individual Web site, it is essential that all relevant pages be returned.
- ❑ Of course, a site search engine can be customized to search only keyword meta tags as long as keywords are carefully created for each page.

Source: A-Z Indexes to Enhance Site Searching by Heather Hedden (2005): http://www.digital-web.com/articles/a_z_indexes_site_searching/

Search Engines vs. Web Index

Benefits of a Real Index

Advantages of an A-Z Index:

- All the entries are visible.
- The quality of results can be controlled by the indexed entries.
- Well-labeled internal links in an A-Z index increase the SEO rating of the linked pages and of the entire site.
- The absence of irrelevant page retrieval makes an index more efficient, enhancing usability.

Sources:

A-Z Website Indexes Explained by Heather Hedden (2004):

<http://www.sitepoint.com/a-z-indexes/>

Why Create an Index? by David M. Brown: <http://www.web-indexing.org/article-brown.htm>

Search Engines vs. Web Index

Benefits of a Real Index

Advantages of an A-Z Index (continued):

- ❑ The ability to browse the index enables users to digress and explore other topics that catch their attention, keeping them on the site longer, improving "stickiness".
- ❑ A-Z indexes can be effectively implemented on sites that are too small to work with site search engines (ie. 20-50 pages).
- ❑ Using a good index, the user can make discoveries: material related to what is being sought, including terms the user might not have thought to look under, and concepts the author and indexer considered important.
- ❑ Creating the index gives the author and indexer a chance to focus on the content, to see the document as a whole, to identify inconsistencies, and improve the content of the web site.

Sources:

A-Z Website Indexes Explained by Heather Hedden (2004):

<http://www.sitepoint.com/a-z-indexes/>

Why Create an Index? by David M. Brown: <http://www.web-indexing.org/article-brown.htm>

Search Engines vs. Web Index

Ideally, a site will have:

- Search Engine/Search Within feature
 - A-Z Index
 - Site Map
-
- The goal for all web sites should be usability, functionality and accuracy with organized information presented in a clear and readable manner.
 - An A-Z Index acts as a **complement** to the site map or table of contents and enhances the search engine optimization ranking of the site.
-

Questions Now?



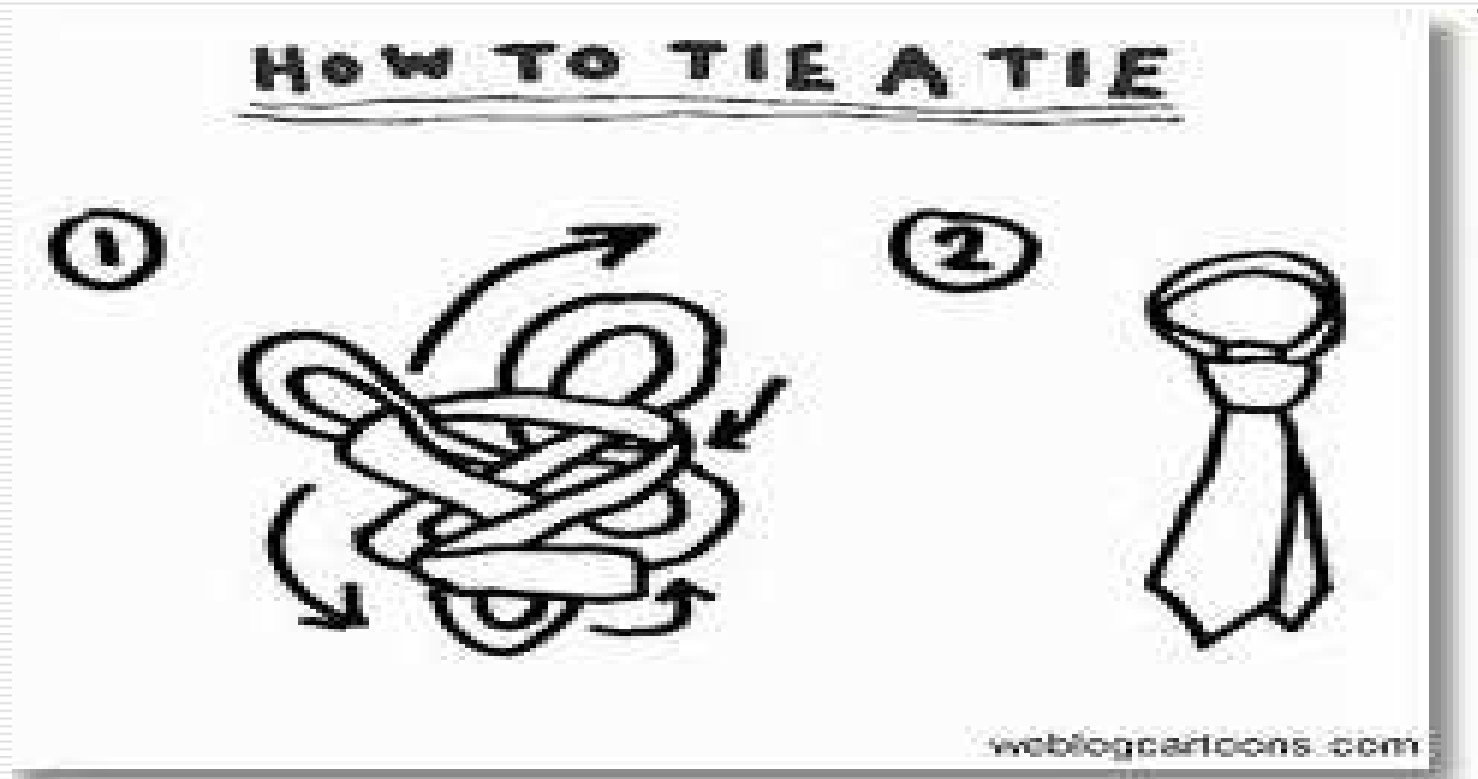
Answering the W's – HOW MUCH

How much does a Web site index cost?

- ❑ Freelance indexers with a background in book indexing are accustomed to being paid per page, such as US\$3.50 - \$5.00 per page.
- ❑ Web pages, however, vary greatly in length and in the amount of indexable content, so the number of pages is not an ideal way in which to determine cost.
- ❑ Site indexing is therefore charged by the index entry or by the hour.
- ❑ Alternatively, once the exact scope of what is to be indexed has been defined, the indexer may be happy to quote a flat fee for the job.
- ❑ Indexers tend to charge more per hour than copy editors, but less than Web developers or information architects.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

How To... do what?!



Answering the W's – HOW

How does a Web site index work?

- ❑ Each of the alphabetically arranged entries or subentries is hyperlinked to the page or to an anchor within a page to where the topic is discussed.
- ❑ Divided into sections for each letter of the alphabet.
- ❑ Typically, each letter is linked at the top of the page and allows a jump to the start of that letter's section of the index.
- ❑ Cross-references (See and See Also), while indicating a preferred term, may link directly to the appropriate Web pages.

Sources:

ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>)

A-Z Indexes to Enhance Site Searching by Heather Hedden (2005): http://www.digital-web.com/articles/a_z_indexes_site_searching/

How to create a web index...

How to create a web index without a utility tool:

1. Create a new HTML file.
2. Create list and type in all the index entries from the website or project
 - Sort terms alphabetically and indent subentries
3. Create anchors for each indexed entry within the website pages
4. Carefully type in or copy/paste the correct URL with named anchors for each entry by hand.
5. Format the entries.
6. Add the headings for symbols, numbers, and letters of the alphabet.
7. Add navigation links to help users "move around" in the finished index.
8. Test every navigation link and every index entry to make sure they work the way you want.
9. Make sure all links work – no broken links or dead-ends.
10. Find and correct any typographical errors that make your index entries unusable.
11. And when anything changes—if you add files, remove files, or reorganize the directory structure—do the process again.

Adapted from "How hard is it to make an index without HTML Indexer?" [HTML Indexer™ Frequently Asked Questions] at <http://www.html-indexer.com/>

So, You Ask...



Isn't there software for automatically creating indexes?

- ❑ Software can automatically extract page titles or headings, retain their page URL links, and sort them alphabetically. But only a human indexer can edit (or, more precisely, rewrite) such a list of titles and headings into a useful and meaningful set of index entries, add variant terms and cross-references, and decide where and how to structure subentries.
- ❑ Software can also automatically extract metadata, such as keywords, to create an index. But in order to be truly useful, an indexer will have had to create the keywords for each page in a systematic way, using a controlled vocabulary for consistency.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

Purposes of Web Indexing Software

- Aid the indexer in creating a browsable, alphabetical-style web index, where index entries are hypertext and linked to the page, section, or paragraph that is referred to by the index entry.
- Aid the indexer in structure and formatting:
 - back-of-the-book style,
 - indented subentries under main entries, and
 - cross-references (See and See Also)
- Aid in navigability with an easy way to navigate through the index
 - Using hypertext 'jump' letters of the alphabet or other design

Source: Software for HTML Indexing by Heather Hedden: http://www.hedden-information.com/Indexer_Apr_06_Hedden.pdf

Tools



Web Indexing Tools

- Any HTML editing software
 - Dreamweaver, FrontPage, etc
 - Requires HTML knowledge and comfort with WYSIWYG editing
- Word or other text editing software
- A-Z Site Indexing Tools include:
 - HTML/Prep by [Leverage](#), maker of CINDEX
 - This utility is often used in conjunction with index preparation software, such as CINDEX, Sky Index, or Macrex. The input format is very straightforward to produce so it can also be created using Word, WordPerfect, or a text editor.
 - HTML Indexer, by [Brown Inc.](#)
 - HTML Indexer embeds the index entries right in your HTML files
 - XRefHT32, pronounced 'shreft' a freeware utility by [Tim Craven](#)
 - Produces index displays in HTML form, including links to Web documents. Allows for automatic target extraction, target and link insertion, and production of tables of contents in HTML files and insertion of certain other HTML codes in plain text files.

Sources:

Software for HTML indexing by Heather Hedden at http://www.hedden-information.com/Indexer_Apr_06_Hedden.pdf

Web indexing tools by Kevin Broccoli at <http://www.stcsig.org/idx/articles/webindexing.pdf>

Web & Electronic Indexing SIG: Resources: Software: <http://www.web-indexing.org/web-indexing-software.htm>

Using Indexing Tools

Index editing software plus HTML-conversion tool:

- ❑ Most professional freelance indexers use one of three commercial indexing software packages: Cindex, Macrex, or SKY Index.
- ❑ To convert an index created in one of these programs to an HTML document, use a utility called HTML/Prep.
- ❑ An index must first be created, and URLs for links instead of page numbers need to be manually pasted.
- ❑ HTML/Prep then converts the compiled index file, which has been saved as text, into an HTML file, preserving the index format style and adding hyperlinked letters of the alphabet along the top of the page.

Source: A-Z Indexes to Enhance Site Searching by Heather Hedden (2005):

http://www.digitalweb.com/articles/a_z_indexes_site_searching/

Using Indexing Tools Continued

HTML index editing software:

- ❑ If you feel comfortable with the techniques of indexing but do not own a dedicated indexing program such as Cindex, Macrex, or SKY Index, the most comprehensive tool for back-of-the-book style indexes is HTML Indexer.
- ❑ With the site's files on the local computer, HTML Indexer automatically creates an editable entry for each page or page plus named anchor, with the URL already retained.
- ❑ Indented subentries and cross-references are supported, but it lacks more sophisticated editing features.

Source: A-Z Indexes to Enhance Site Searching by Heather Hedden (2005):
http://www.digitalweb.com/articles/a_z_indexes_site_searching/

Pros and Cons



Pros of Utilities vs Hand Coding

Software Pros

- May save time
- Assists in error catching
- Quality control

HTML Hand Coding Pros

- More developer control
 - For smaller sites, may be just as easy
 - Quality control
-

Cons of Utilities vs Hand Coding

□ Software Cons

- Costs – software and time
- Learning curve
- Not automatic, still requires massaging

□ HTML Hand Coding Cons

- Requires comfort and skill with HTML coding
 - Costs still exist
 - Learning curve
-

Boils down to...

□ Personal Preference

- Similar to choosing between indexing tools like Cindex, Sky, Macrex or basic editing tool.



How I Create a Web Index...

- ☐ With A LOT of Coffee!



How I Create a Web Index...

1. Review existing A-Z Web Indexes in similar fields.
 2. Read and analyze the content of site
 3. Create list of terms to index in Word or Notepad, copy and paste into Dreamweaver
 4. Use Dreamweaver to draft the index
 5. Create new HTML page for A-Z index
 6. Must have access to the main site and code to insert named anchors. Save pages. Always have a backup!
 7. Record named anchors and url (web address) and insert directly to hyperlink in my Dreamweaver index draft for each entry name.
 8. Constantly save.
 9. Can only test when pages with named anchors and site index are uploaded and live.
 10. Work on formatting using style guide best practices and preference for look and feel.
 11. Refine, test, refine, test.
 12. Upload. Go Live.
-

Questions?



Style and Best Practices

Is there a standard style for Web site indexes?

- There are no standard styles for Web site indexes like there are for book indexes.
- When it comes to format, the guiding factor should be the usability of an index.
- However, there are some style conventions, known as Best Practices.
- Also note that guidelines and standards should be applied in web design and development for all web projects.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

ASI Web Indexing SIG

A-Z Best Practices

- ❑ **Site Navigation:** A-Z index should be clearly labeled and linked.
- ❑ **Number of Pages:** The index can be kept on one page or broken into multiple pages if it is long.
- ❑ **Internal Index Navigation:** should be a means of navigating the index either by hyperlinked "jumps" or other format.
- ❑ **Columns:** Only a single column should be used. Scrolling is easier when there is only a single column.
- ❑ **Entry Spacing:** The entire index should be single-spaced. Entry Case: Lower-case should be used for entry topics to better distinguish.
- ❑ **Entry Font Color:** Different colors for links, visited and unvisited, should be used.
- ❑ **Entry Font Size:** A standard font size should be used.
- ❑ **Entries to Nonstandard Pages:** A convention (different color, parenthetical notes, added small graphic icon, etc.) should be used to indicate links to nonstandard pages, such as external links, password restricted pages, and non-HTML pages such as pdfs or Word documents.
- ❑ **Indenting:** Indenting of subentries is the preferred style, with subentries as hanging indents.
- ❑ **Subentry Levels:** Up to three levels is acceptable.
- ❑ **Subentries, Number of:** Contrary to book-indexing style, single subentries are acceptable.
- ❑ **Cross-references:** If a See or See Also reference points to an index entry that has no subentries, then the link should be to the source page and not the other referred term in the index, to eliminate the need for an extra click within the index. If there are subentries under the referred term, then the See or See also reference should link to the preferred term within the index.

Source: ASI Web Indexing: Resources: Best Practices: <http://www.web-indexing.org/practices.htm>

Challenges?



Challenges in Web Indexing

- Updating issues:
 - A-Z indexes are more practical on sites that are not constantly changing.
 - Specific content within pages can be updated without affecting the index, as long as the topic of the page or page section remains the same.
 - If pages are added and deleted in an unsystematic manner, then more work is required to keep the index maintained.
 - If pages that get added tend to be of a consistent type, then the original indexer can write indexing policy guidelines that the Webmaster can follow.

Source: A-Z Indexes to Enhance Site Searching by Heather Hedden (2005):
http://www.digital-web.com/articles/a_z_indexes_site_searching/

Challenges in Web Indexing

- ❑ Web sites change and too many updates can be difficult to hand code or rebuild the index even with utilities.
- ❑ Requires time and money: staff, contractor
- ❑ Requires pre-coordinated work and maintenance.
- ❑ Search engines may be easier to implement with time and money considerations.

Source: The Future of Indexing? Jan Wright, WrightInformation, Writers UA,
http://www.writersua.com/articles/indexing_future/

Getting Started in Web Indexing

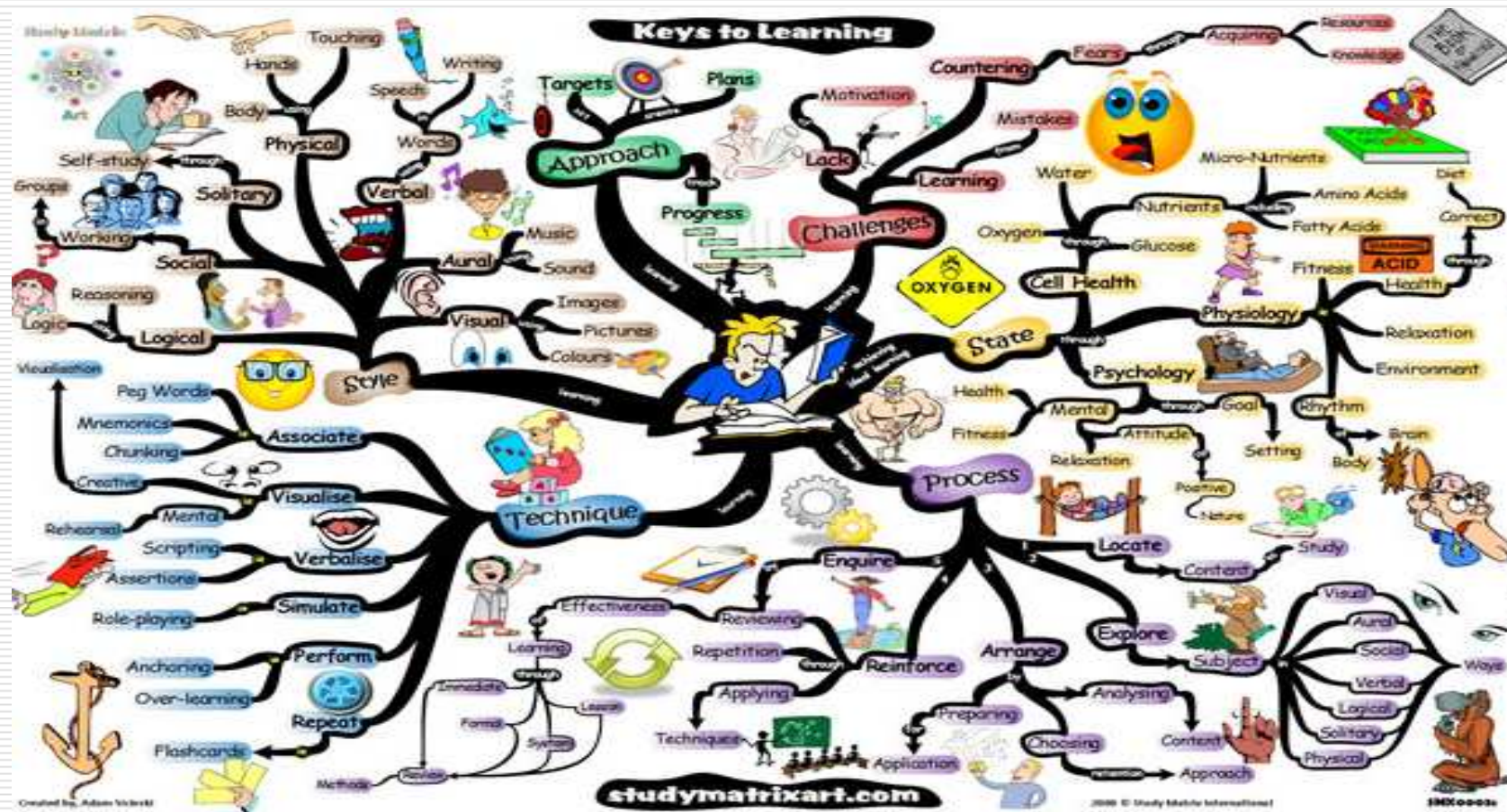
Learning HTML

Web indexing tools

Finding projects



Learning



Training Options – Web Indexing

- ❑ Classes, resources, ASI Web Indexing SIG (<http://www.web-indexing.org/>)
 - ❑ Hedden Information Management (Heather Hedden), Web: <http://www.hedden-information.com/workshops.htm>
 - ❑ Middlesex Community College, offers its Indexing Books and Websites Certificate Course over the Web (Seth Maislin), Web: <http://www.middlesex.mass.edu/CareerTraining/WritingIndexesforBooksandWebsites.htm>
 - ❑ University of California Berkeley—Indexing: Theory and Application (online learning course with Sylvia Coates, Fred Leise, Jan C. Wright and others), Web: <http://unex.berkeley.edu/cat/course394.html> / Now has a Module on Embedded Indexing: Indexing a Web Site
 - ❑ More training options viewed at ASI Web Indexing Courses page at <http://www.web-indexing.org/courses.htm>
-

Training Options - HTML

Basic HTML Resources

- Although one can create a Web site index with a Web indexing tool and no knowledge of HTML, a basic skill in HTML is highly desirable to tweak or fix an index and to meet client style format requirements. The basics of HTML can be learned from any of various online tutorials. The following are but a few of the online HTML tutorials available:
 - HTML Tutorial on the W3 Schools Web developers' site: <http://www.w3schools.com/html/default.asp>
 - "Getting Started with HTML" by Jukka Korpela: <http://www.cs.tut.fi/~jkorpela/html-primer.html>
 - HTML Cheatsheet on the Webmonkey site: http://www.webmonkey.com/2010/02/html_cheatsheet/
 - Web Design Group: <http://www.htmlhelp.com/>
 - Ed2go: <http://www.ed2go.com/>
 - More advanced certificates and degrees are offered in web design, web development, and across the IT field.
-

Food for Thought...

Special Considerations - 1

Why wouldn't a site search engine serve just as well or better than an index?

- ❑ A search engine query will not always provide you with the information you're looking for. Compared with the entire Web, the number of pages within a site is relatively small, so a simple search engine query might not yield enough or any results, even if there are good pages on the subject. This is most likely to occur because the search phrase you type is worded differently than references to that topic within the page text.
- ❑ Whole-Web search engines usually produce "satisfactory" results in the quality of articles, since the major search engine companies have developed complicated criteria and algorithms for the retrieval and ranking of pages. Off-the-shelf search engines to be used within a site are not so sophisticated. They often retrieve pages that include a mere passing mention of the search term, but do not really focus on the subject at all.
- ❑ Jared Spool's article "[Why On-Site Searching Stinks](#)" further points out the limitations of on-site search utilities.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

Food For Thought...

Special Considerations – 2

How do you keep a Web site index up to date as the site changes?

- ❑ An index should be written so that the most dynamically changing parts of a site are not indexed to the specifics. For example, a page on events should be indexed for the topic of “Events,” but not for the specific events themselves, which would be constantly added and deleted.
- ❑ If the index is created by a contracted indexer, an agreement needs to be reached regarding how the index will be maintained. Either the indexer can be retained for future updates, or the indexer can provide written guidelines.
- ❑ If an indexer is not maintaining the index, it might be a good idea to contract an indexer to review it periodically. An indexer who updates an index need not be the same indexer who created the index, but anyone who updates an index needs to be informed of what major content changes have occurred since the last index update.
- ❑ Indexing review and updates should also catch broken links, HTML problems, typos, and any other issues that affect the Web site as a whole, and be reported for correction.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>

Food For Thought...

Special Considerations - 3

Web Accessibility/ADA Compliance Considerations

- ❑ Section 508 is a regulation for government websites and many consider it a good reference (best practice) and use it as a guideline.
 - ❑ W3 Accessibility Evaluators:
<http://www.w3.org/WAI/RC/tools/complete> and has a list of web accessibility evaluators like BOBBY and WaveAIM (WAVE).
 - ❑ For specifics, see the Usability.gov Guidelines at <http://www.usability.gov/guidelines/>
-

Food For Thought...

Special Considerations - 4

Is an alphabetical list of Web pages a form of an index?

- ❑ An alphabetical list of web page titles, or even edited titles, is not a Web site index, just as an alphabetical list of chapter titles does not constitute a book index.
- ❑ Wording of page titles can be quite different from a typical index entry.
- ❑ A page title "Work For Us," while catchy as a heading, is not likely to be quickly found by a user searching for a job in an alphabetical list of Web site pages.
- ❑ A true index entry to the "Work For Us" page, on the other hand, could be posted as "careers," "employment," or "jobs" – or indeed all three.
- ❑ A simple alphabetical sorting of Web pages would lack variant or cross-reference terms and second-level entries, all of which would be found in a true index.

Source: ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>)

Examples

Case
Examples

Web Indexes – A to Z Examples

- More A-Z index examples:

- A - Z Indexes on the web: An annotated sampler:

- <http://www.montague.com/review/AtoZ.htm>

- ASI Web Indexing SIG – Examples:

- <http://www.web-indexing.org/web-index-examples.htm>

Let's Revisit a Few Examples of A-Z Indexes

Sites **WITHOUT** an Index... and perhaps need one?

- ❑ UNICEF: <http://www.unicef.org/>
 - ❑ U.S. Department of Transportation: <http://www.dot.gov/>
 - ❑ Federal Transit Administration: <http://www.fta.dot.gov/>
 - ❑ World Bank: <http://www.worldbank.org/>
 - But take a look at Topics: <http://www.worldbank.org/html/extdr/thematic.htm>
 - ❑ City of Fort Collins: <http://www.fcgov.com/>
 - ❑ Colorado.gov: <http://www.colorado.gov/> (Watch this one =web portal! See under State Agencies)
 - ❑ Community Resource Center: <http://www.crcamerica.org/>
 - ❑ And our own Rocky Mountain Chapter: <http://asirm.org/>
-

Into the Future...



Into the Future...

- ❑ The future of web indexing:
 - ❑ E-books
 - ❑ E-libraries & Digital libraries
 - ❑ E-zines and periodicals
 - ❑ Print to web docs
 - ❑ Blogs and Social Media
 - ❑ More crossovers: A to Z Glossary Index, Directories, Site Maps
 - ❑ Growing content and larger sites with more information needing access points
-

Into the Future...

Web Indexers as:

- Web architects
- Web designers

Creating:

- Metadata
- Controlled Vocabularies
- Search Engine improvements & work in collaboration with metadata and keywords

What Is A Controlled Vocabulary? by Fred Leise, et al (2002):

http://www.boxesandarrows.com/view/what_is_a_controlled_vocabulary

Into the Future...

- Social Media and Web Indexing?
 - Blogs? Maybe, Yes.
 - But what about:
 - Twitter??
 - Facebook??

The fact is:

Content-rich web sites will remain viable, useful and necessary, as will their web indices AND Social Media sites and tools will co-exist.

Into the Future...

Quotes by Tim Berners-Lee

- “The Web as I envisaged it, we have not seen it yet. The future is still so much bigger than the past.”
- “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

About Tim Berners-Lee:

- A graduate of Oxford University, Tim Berners-Lee invented the World Wide Web, an internet-based hypermedia initiative for global information sharing while at CERN, the European Particle Physics Laboratory, in 1989. He wrote the first web client and server in 1990. His specifications of URIs, HTTP and HTML were refined as Web technology spread.

Source: <http://www.w3.org/People/Berners-Lee/>

Questions?

- What questions do you have?
- Are you interested in learning more?



Web Indexing Activity

- ASIRM Web Site: <http://asirm.org/>
 - First Review: American Society for Indexing A-Z Index:
<http://www.asindexing.org/>

 - Also Needed: ASIRM Web site feedback, usability, content
-

REMEMBER: ASI Web Indexing SIG A-Z Best Practices

- ❑ **Site Navigation:** A-Z index should be clearly labeled and linked.
- ❑ **Number of Pages:** The index can be kept on one page or broken into multiple pages if it is long.
- ❑ **Internal Index Navigation:** should be a means of navigating the index either by hyperlinked "jumps" or other format.
- ❑ **Columns:** Only a single column should be used. Scrolling is easier when there is only a single column.
- ❑ **Entry Spacing:** The entire index should be single-spaced. Entry Case: Lower-case should be used for entry topics to better distinguish.
- ❑ **Entry Font Color:** Different colors for links, visited and unvisited, should be used.
- ❑ **Entry Font Size:** A standard font size should be used.
- ❑ **Entries to Nonstandard Pages:** A convention (different color, parenthetical notes, added small graphic icon, etc.) should be used to indicate links to nonstandard pages, such as external links, password restricted pages, and non-HTML pages such as pdfs or Word documents.
- ❑ **Indenting:** Indenting of subentries is the preferred style, with subentries as hanging indents.
- ❑ **Subentry Levels:** Up to three levels is acceptable.
- ❑ **Subentries, Number of:** Contrary to book-indexing style, single subentries are acceptable.
- ❑ **Cross-references:** If a See or See Also reference points to an index entry that has no subentries, then the link should be to the source page and not the other referred term in the index, to eliminate the need for an extra click within the index. If there are subentries under the referred term, then the See or See also reference should link to the preferred term within the index.

Source: ASI Web Indexing: Resources: Best Practices: <http://www.web-indexing.org/practices.htm>

Bibliography – Online Resources

- ❑ ASI Indexing the Web: <http://www.asindexing.org/i4a/pages/index.cfm?pageid=3418>
 - ❑ ASI Web & Electronic Indexing SIG: <http://www.web-indexing.org/>
 - ❑ ASI Web Indexing SIG Brochure: <http://www.web-indexing.org/brochure.pdf>
 - ❑ ASI Web Indexing Resources: Articles: <http://www.web-indexing.org/articles.htm>
 - ❑ ASI Web Indexing Resources: Books: <http://www.web-indexing.org/books.htm>
 - ❑ ASI Web Indexing Resources: Software: <http://www.web-indexing.org/web-indexing-software.htm>
 - ❑ ASI Web Indexing FAQ: <http://www.web-indexing.org/faq.htm>
 - ❑ ASI Digital Trends Task Force:
<http://www.asindexing.org/i4a/pages/index.cfm?pageid=3647>
 - ❑ Also has a Linked In discussion group
 - ❑ Improving Usability with a Website Index by Fred Leise (2002): <http://www.web-indexing.org/article-leise.htm>
 - ❑ Embedded indexing by James Lamb (2005):
<http://www.jalamb.com/2005%2010%2001%20Embedded%20indexing%20The%20Indexer.pdf>
 - ❑ Web Indexing: <http://www.allegrotechindexing.com/news026.htm>
 - ❑ Why Create an [web] Index? By David Brown: <http://www.web-indexing.org/article-brown.htm>
 - Note: Brown is the designer of HTML Indexer software, <http://www.html-indexer.com/>
 - ❑ Web indexing tools by Kevin Broccoli:
<http://www.stcsig.org/idx/articles/webindexing.pdf>
-

Bibliography - Books

- ❑ Henninger, Maureen. Hidden Web: quality information on the Net. Sydney, UNSW Press, 2nd ed, 2007. ISBN: 978-0868408552, 352 pages.
 - ❑ Hedden, Heather. Indexing Specialties: Web Sites. Medford, New Jersey: Information Today, Inc., 2007. ISBN: 978-1-57387-302-4. 165 pages. (book review by Richard Shrout)
 - ❑ Lamb, James. Website Indexes: Visitors to Content in Two Clicks, or website indexing with XRefHT32 freeware. Ardleigh, Essex, England: Jalamb.com Ltd., 2006. ISBN: 978-1-4116-7937-5. 144 pages.
 - ❑ Browne, Glenda and Jonathan Jermey. Website Indexing: Enhancing Access to Information within Websites, 2nd edition. Adelaide, South Australia: Auslib Press, 2004. 148 pages.
 - ❑ Schroeder, Sandi. Software for Indexing. Medford, New Jersey: Information Today, Inc., 2003. ISBN: 1-57387-166-4. 220 pages. [Includes a chapter on Web and online indexing software.]
 - ❑ Meisheid, William. Successful Indexing With RoboHELP HTML Edition. Ellicott City, Maryland: Sageline Publishing, 2001. ISBN: 096-725-704-2. 216 pages.
 - ❑ [Currently out of print and a little outdated for the current version (RoboHelp X5), but the author has turned it into a PDF that he may distribute or even break out as a section of his Web site.]
 - ❑ Brenner, Diane and Marilyn Rowland. Beyond Book Indexing: How to Get Started in Web Indexing, Embedded Indexing, and Other Computer-Based Media. Medford, New Jersey: Information Today, Inc., 2000. ISBN: 1-57387-081-1. 150 pages.
-