

Software for HTML indexing: a comparative review

Heather Hedden

This article reviews and compares the features of various software programs for the creation of website or HTML indexes. The different categories of software programs are discussed, and specific attention is given to the programs HTML/Prep, XRefHT32, and HTML Indexer.

Introduction

Ever since the Web became popular in the mid-1990s, there have been indexers who have attempted to create indexes on websites. Few indexers then, and only a minority now, have skills in HTML coding however, and even with these skills the procedure is quite laborious. HTML editing or authoring software helps only minimally, because such software cannot sort terms alphabetically nor indent subentries, and links still need to be created individually. Over the years, various software options have emerged, so now an indexer with minimal knowledge of HTML can efficiently create an HTML index. Information about such software has been scarce. This article seeks to provide a comprehensive overview and comparison of the software options.

Researching the literature on web indexing software can be tricky, because the terms 'web indexing,' 'html indexing,' or 'site indexing' can mean different things to different people. Indexers are interested in software tools that automatically create the hyperlinks and take care of sorting, but leave the creation of the terms up to the indexer. To web developers, webmasters, and other Internet professionals, however, these terms tend to imply simply an automatic listing of all the HTML files in a site for inventory purposes or so that the pages can be accessed, or 'spidered,' by a search engine. A literature search on 'web indexing,' 'html indexing,' or 'site indexing' software will more likely bring up the software for the inventory or 'spidering' uses.

The purpose of editorial HTML indexing software is to aid the indexer in creating a browsable, alphabetical-style index, where index entries are hypertext and linked to the page, section, or paragraph that is referred to by the index entry. Ideally the index should be structured in back-of-the-book style, with indented subentries under main entries and cross-references (*See* and *See also*) also supported. Finally, there should be a way to navigate quickly through the index, either by hypertext 'jump' letters of the alphabet or by type-ahead scrolling.

Even when focusing on the narrow field of editorial web indexing tools, the products available are quite varied. Software tools for creating HTML indexes are of three kinds:



Figure 1 An excerpt example of a website index from a public library

- indexing features of HTML online Help authoring tools
- utility programs that help convert a sorted, formatted index into HTML with hypertext links
- dedicated, stand-alone software that combines both the term-sorting/formatting and hypertext link creation features.

Online Help indexing software

Help authoring tools (often abbreviated as HAT) have long included the feature for creating an alphabetical index for the documentation which is also created in the program, and the index topics are hyperlinked to the relevant page within the Help documentation file. Online Help files are traditionally not HTML files, though. Thus not all Help authoring tools support HTML output. More recently, however, online Help in HTML has become common.

Online Help indexes are often displayed in a vertical frame, and are an option, typically selected from a tab that also gives access to the table of contents and text search. The index is typically scrollable and often searchable with a type-ahead feature. While type-ahead searching is an attractive feature, it takes time to load. The index itself is not an HTML file, but rather a special file with the extension .hkh that needs to be called up with a java applet program. The delayed time in running the applet makes indexes created with such an online Help tool somewhat slow in running on remotely accessed websites. Such indexes are quite feasible, though, for intranets that act over a high-speed network, or HTML documents stored and accessed on a single desktop computer, such as ebooks.

The indentation of subentries is supported by many, but not all, Help authoring tools. A minor disadvantage with many Help-style indexes is that the entries hyperlinked to topics are not indicated by being displayed in a different color and underlined. (Subentries and unsubdivided main headings will link to the text, but main headings that have subentries often have no links.)

There are a number of tools for creating HTML-based online Help with indexes:

- AuthorIT V4 Desktop Edition, <http://www.authorit.com>, US\$459
- Deva Tools for Dreamweaver, <http://www.devahelp.com>, US\$249
- FAR HTML freeware from The Helpware Group, <http://helpware.net/FAR>, free
- HDK (Hypertext Development Kit), from Microway, http://www.microway.com.au/virtual_media/hdk.stm, US\$549
- HTML Help Workshop from Microsoft, <http://www.microsoft.com/downloads>, free
- Hyper Text Studio Professional Edition from Olson Software, <http://www.olsonsoft.com>, US\$199
- RoboHelp X5 from Macromedia, <http://www.macromedia.com/software/robohelp>, US\$999
- WebWorks ePublisher Pro for Word from Quadralay Corp., http://www.webworks.com/products/wwep_w, US\$1,000.

The index is designed to work on the documentation texts that are created as part of the Help project. Therefore, not only does one have to learn how to use the indexing tool component of the software, but also how the Help-authoring software works overall. Although the software is designed to aid in indexing texts that are created within the project, rather than existing websites, it is possible to import files. The ability to easily manipulate terms and subentries (duplicating, rotating, etc.) is often somewhat limited in comparison with dedicated indexing software, though.

Macromedia RoboHelp is probably the preferred Help authoring tool among professional writers of online Help. Some freelance indexers also use it for certain clients. It is the tool with the most support, documentation, online discussion, and literature, including information about its indexing features. While other tools might provide comparable indexing features, one is not as likely to hear about them.

FAR HTML is one of the less-known Help authoring tools, and it is worth looking into, since it is free. Furthermore, an example of an index created in the program is available to see and try out on its website, http://helpware.net/FAR/help/hh_start.htm, which is the index to the online Help for FAR HTML itself. With the full Help documentation on the website, one can get an idea of how this software works before downloading it. The indexing features are somewhat limited and it is a little bit cumbersome, especially compared to the professional dedicated

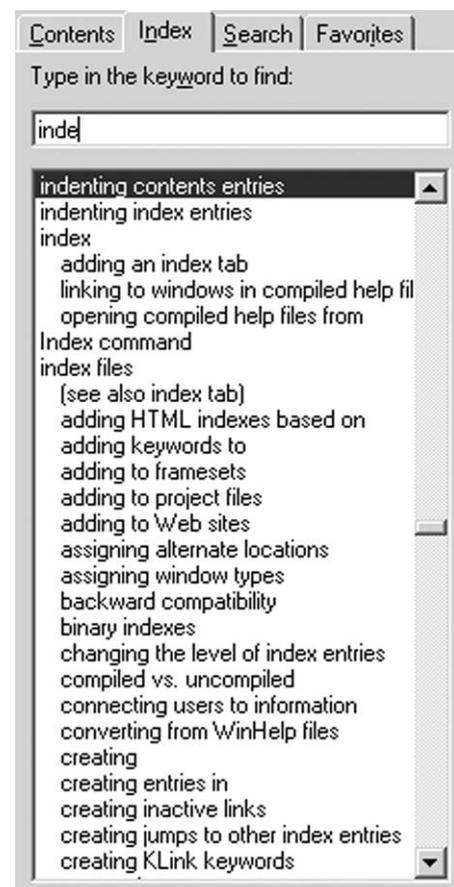


Figure 2 An example of an index with type-ahead scrolling, created in HTML Help Workshop

indexing tools. If the requirement, however, is to create a scrollable index in a frame that is on every page of the website, some compromises have to be made in selecting the indexing tool.

Deva Tools is worth considering for the indexer who is already using Macromedia Dreamweaver to maintain the content of the website to be indexed. The current version of this add-on utility is for Dreamweaver MX 2004. The output is an index (and table of contents) in a scrollable frame in a similar display to that created by other online Help authoring tools. The software comes bundled with another application, Deva Search, which creates full-text search indexes for web pages, and this combined package for \$249 is more than what the indexer needs. Deva Tools' indexing features, as with other Help authoring tools, are somewhat limited. It stores added index entries in an XML file, not in the source file, and several steps and manual interventions are required to generate the final HTML index page.

Conversion tools

Back-of-the-book indexers use one of various dedicated indexing tools, such as Cindex, Macrex, or SKY Index, which have many powerful features for index creation. For indexers skilled in using one of these tools, the best option might be a tool that converts an index created in one of these programs.

Actually, Macrex now has features for HTML output that are sophisticated enough to preclude any additional conversion tool. The procedure is rather complicated to implement, but the North American Macrex Support is helpful in guiding users through the steps. In addition to Macrex, Cindex and SKY Index both support tagged HTML output, but considerable editing of the HTML index is still required.

The most basic tool to convert any text, not merely a sorted index, is any HTML editing software. Common HTML editors are Mozilla Composer, Microsoft FrontPage, Homesite, Macromedia Dreamweaver, CoffeeCup HTML Editor, and HotDog. While HTML editors cannot convert an entire formatted index into HTML with a single press of a button, they do allow the indentation of individual subentry lines and the addition of hypertext links without requiring the typing of HTML code.

The first tool created as an add-on specifically to convert indexes into HTML was WEBIX, made available in 1995 by Australian Dwight Walker. WEBIX was a bundle of two earlier programs, Weblinkr, a Windows tool developed by Australian indexer and PC trainer Jonathan Jerney, and INDTOHTM, a DOS tool developed by Canadian information science professor Timothy Craven. WEBIX was offered free to get indexers started in web indexing and to participate in the newly created Web Indexing Prize of the Australian Society of Indexers. WEBIX is still available from WWWalker Web Development (www.wwwalker.com.au) on an 'as is' basis with no support available. It is rather simple, though, with its command-line interface and text editor for creating the input index file. The Windows version relies on an out-dated DLL file, VBRUN200.DLL, which is

available free from DLL-files.com. Walker later designed his own version of WEBIX but it was never released due to competition from HTML Indexer.

HTML/Prep

<http://www.levtechinc.com/ProdServ/LTUtils/HTMLPrep.htm>, \$125

Today the most common tool for converting indexes to HTML is HTML/Prep from Leverage Technologies (LevTech). LevTech is also the authorized agent for North American corporate and government accounts for the indexing tool Cindex. As such, LevTech has developed several software utilities to be used on indexes created in Cindex or similar indexing programs. HTML/Prep is one of these. It converts a file, with or without web links as the locators, into an HTML-coded web page.

HTML/Prep is a DOS utility for converting a tagged index into an HTML file with functioning links. As a utility, it has no interface of its own. The user merely issues the commands at the DOS command prompt. For those not familiar with command-line DOS, this might seem intimidating, but it is not too difficult. The HTML/Prep documentation needs to be read thoroughly, though, since there are no intuitive menus. Unfortunately, there is no free demo version.

HTML/Prep automatically does the following to an index:

- indents subentries
- makes the entries hyperlinked to the web page file name that has been entered in the page field
- adds hyperlinked letters for navigation at the top of the page, top and bottom, or in a separate page/file, linked to letter headings within the index
- makes cross-references hyperlinked.

Thus HTML/Prep can eliminate quite a bit of tedious work that is involved in manually creating an index. HTML/Prep also allows for different output options, such as the placement of navigation letters and choosing whether or not the index entries are bulleted.

HTML/Prep is easiest to use with Cindex. Among the file types that a Cindex index can be saved as, one is .htp for HTML/Prep. The HTML/Prep package comes with a style sheet for Cindex (although it is not necessary to understand how it works). Before executing HTML/Prep, one merely needs to issue a command indicating what version of Cindex is being used. When converting a SKY index with HTML/Prep, the correct output format tags need to be set first; this is not one of the Preset choices, but rather needs to be created as a custom set of tags.

An interesting feature of the generated HTML index is that when the mouse is placed over a subentry, the main entry appears in a pop-up. This is useful to the user if there is a very long list of subentries, with the main entry scrolled off the top of the window display, as it enables the user to keep track of what the main entry was. This works for sub-entries as well.

Dedicated, stand-alone web indexing tools

One drawback of using software that merely converts an index to HTML is that, as the index is created, each file name or URL (uniform resource locator, the address of web pages) needs to be individually copied and pasted into the locator field. Dedicated, stand-alone web indexing software eliminates this task by automatically extracting the URLs in association with the web page title, heading, or other information.

There are two good options of software for creating website indexes that provide the combined capabilities of automatically sorting the index entries, creating associated indented subentries, and capturing the target URLs to hyperlink the entries automatically. One is the commercial software HTML Indexer, which is developed and sold by David M. Brown of Brown Inc. The other is freeware called XRefHT32, developed and distributed by Tim Craven, one of the original WEBIX developers.

XRefHT32

<http://publish.uwo.ca/~craven/freeware.htm>, free

XRefHT32, commonly pronounced 'shrefht,' is a little-known but quite effective freeware indexing tool. Its name derives from the phrase 'cross-referenced hypertext for 32-bit Windows.' An example of a website index created with this program is found on the web pages of the developer, Tim Craven, who is a professor of information science of the University of Western Ontario.

XRefHT32's ability to extract headings from within a page and automatically add anchors (locations tagged anywhere within a page for a link destination) is a powerful and unique feature in website-indexing tools. It should not

be misused, though. It is a good idea to extract the headings, to see what they are, but it might not always be a good idea to add anchors to all of them. If only a few headings need anchors added, it might be more practical to add them manually so as to avoid adding numerous, strange, lengthy anchor names to the file. For one who has no skills in creating anchors, though, this automated feature could prove quite helpful.

The other unique feature of XRefHT32 is its ability to incorporate a thesaurus created in a related thesaurus-creation tool, also developed by Craven, 'TheW32,' and to automatically make the cross-references. In addition to *See* and *See also* cross-references, the thesaurus also includes Broader Term, Narrower Term, and Related Term references. For most website indexes, a thesaurus is not necessary. But for a web-based index of one or more periodicals, with the continual addition of new articles to be indexed over time, and possibly by more than one indexer, a thesaurus could be quite helpful. Without 'TheW32,' the process of creating cross-references in XRefHT32 merely takes one additional step for each cross-reference.

Additional nice features are spell-checking, find/replace, and options to output the index either as a single Web page index or with separate pages created for each letter of the alphabet.

As an indexing tool, however, XRefHT32 lacks some of the term-editing features of the commercial indexing tools of Cindex, Macrex, SKY, or even HTML Indexer. This might be considered the program's major weakness. There is also the drawback that subfolders of files to be indexed have to be added individually for URL extracting. As such, XRefHT32 may be more practical with smaller, rather than with larger, sites. The program also lacks output options, and the only choice is an index with bulleted entries.

Heading	Subheading	URL
About Web Indexing		about-web-indexing.htm
A-Z Indexes to Enhance Site Searching		article-broccoli-contentious.htm
Why Create an Index		article-brown.htm
A-Z Indexes to Enhance Site Searching		article-hedden-digitalweb.htm
Why Create an Index		article-leise.htm
Articles on Web Indexing		articles.htm
Contact Information		contact.htm
Web Index Examples by SIG Members		faq.htm
Web Indexing SIG		index.htm
Links to Articles and Resources Links We		links.htm
Membership in the Web Indexing SIG		membership.htm
Other Examples of Indexes on the Web		other-web-index-examples.htm
Resources for Indexers on Web Indexing		resources-for-indexers.htm
Web Site Indexing by Marilyn Rowland		resources-rowland-webindexing.htm
HTML Indexing by L. Pilar Wyman		resources-wyman-HTML-indexing.htm
Types of Web Indexing and Related Serv		types-of-web-indexing.htm

Figure 3 XRefHT32 interface with extracted web page titles, to be edited into index entries, and associated URLs

Since it lacks the term manipulation input features of the commercial indexing tools, XRefHT32 may not be as user-friendly for large projects or for the indexer who compiles multiple website indexes. For a small, one-time project, however, this tool is very useful, and for a free program it is quite impressive and worth trying.

HTML Indexer

<http://www.html-indexer.com>, \$239.95

HTML Indexer is the only commercial stand-alone, dedicated website-indexing tool on the market. It was developed by and is marketed by David M. Brown, a former technical writer and documentation manager, through his own consultancy, Brown Inc., in Portland, Oregon. Examples of indexes created with HTML Indexer are listed on the HTML Indexer site. A demo of the software can be downloaded, which has no time restriction, but does not allow saving. Nevertheless, an HTML index can be generated, and the resulting index can be saved. In other words, the HTML Indexer demo can be tried out on very small projects that do not require intermediate saving.

HTML Indexer works in a somewhat similar way to XRefHT32, and has similar capabilities. It extracts web page file names and also anchors within pages along with their URLs. It does not extract headings, though, nor can it automatically add anchors to headings. This may or may not be important for a project. Creating additional index entries is easier and more sophisticated than in XRefHT32. HTML Indexer does not require a separate thesaurus tool to automatically create cross-references. Finally, there are many more options in HTML Indexer to control the formatting

and output of the index without having to manually change the code. As it is a commercial product, unlike XRefHT32, one can also count on getting technical support to answer any question. In addition to standard online Help documentation, HTML Indexer also comes with a tutorial that includes sample files.

Unlike XRefHT32, URLs of both web page file names and anchor targets are extracted into the indexing program in a single step. Subfolders are also extracted, so the entire site can be extracted at once. Additional unanchored headings, however, cannot be extracted.

One of the nice things about HTML Indexer, is that one can easily call up and view in a web browser the pages being indexed, without having to copy and paste the URLs. Similarly, one can call up web pages to edit (namely add additional named anchors) in a chosen HTML editor in another window. If, when viewing a web page in the browser, the indexer decides that an additional anchor is desired for indexing, such as at a heading, where no anchor exists, the indexer can easily call up that page in the specified HTML editor to create the anchor and then add that page back to the indexing project to complete the indexing.

Another unique feature of HTML Indexer is that it inserts code into each indexed page, linking the page to the index in such a way that, if the indexed page ever gets removed from the website, the index terms linked to that page will automatically disappear from the index. Similarly, the index is automatically updated in the event of a file name change. As such, there is a limited built-in maintenance feature. To take advantage of this feature, all the indexed pages must then be reloaded to the site.

HTML Indexer has several advantages over XRefHT32: the ease with which one can add multiple index entries to

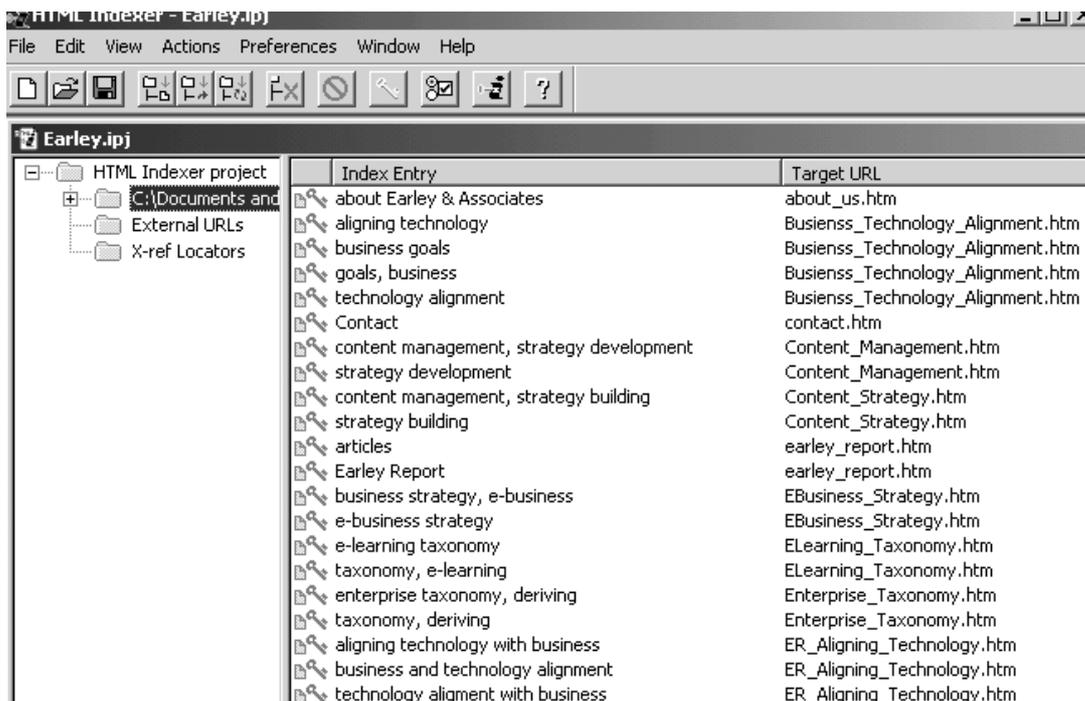


Figure 4 HTML Indexer interface with edited index entries and associated URLs. Subentries are created by typing commas in the entries

the same web page or anchor, a simpler method of adding cross-references, and an efficient way to add additional anchors manually to a web page and then import the page data back into the index file. It also has better documentation. There are several output options, such as the grouping of the navigational letters and whether subentries will be indented by means of repeated blank spaces or if a specific style sheet is to be used. In either case, the generated index does not use bullets. The main drawback of HTML Indexer is its lack of a spell-checker and of any find/replace feature. Unfortunately, there are no plans by its developer to add these or any other additional features in the future.

Conclusions

Indexers owning a professional indexing tool, such as Cindex, SKY, or Macrex, can create HTML indexes, which will then require further editing and copying of links in an HTML editor. This approach makes sense if a website index is a one-time or rare project and is not too large (perhaps no more than 200 entries). If, however, an indexer expects to create more than one or two website indexes, or the index is of a large size, it is probably cost-effective to purchase HTML/Prep, and to take the time needed to learn it. HTML/Prep is also the best option if you need to create an index that will be used both in print and online.

Using an online Help authoring tool purely for its indexing feature is probably not practical, especially for the more expensive tools. Online Help authoring tools are not designed to index general websites, but rather HTML documents kept on a single computer or on an intranet. An online Help authoring tool is not the preferred choice of freelance website indexers. If, however, as an employee or under contract, an indexer is told to create a scrollable Help-style index, then an online Help authoring tool is the way to do it.

Stand-alone tools, such as HTML Indexer or the freeware tool XRefHT32, are probably the best option for website indexing, especially for those not dedicated to using only their favorite book indexing software or those who are intimidated by the DOS-based utility of HTML/Prep. XRefHT32 does a fairly good job, considering it is freeware, but for indexing large or multiple websites, the HTML Indexer would probably be a better choice.

References

- Broccoli, K. (1999) Web indexing tools: a comparative review. *A to Z: The Newsletter of the Indexing SIG* 2(3), 3-7.
- Hedden, H. (2005) HTML indexing freeware: XRefHT32. *Key Words* 13(4): 141-2.
- Rowland, M. and D. Brenner (2000) *Beyond book indexing: how to get started in web indexing, embedded indexing, and other computer-based media*. New Jersey: American Society of Indexers, Info Today, Inc.
- Wyman, L. Pilar. (2001) HTML indexing: a few hints on what it's all about. <http://www.web-indexing.org/resources-wyman-HTML-indexing.htm>

Heather Hedden, principal of Hedden Information Management, is a freelance indexer of books, periodicals and websites. She also teaches an online continuing education course in website indexing through Simmons College Graduate School of Library and Information. Heather leads the Web Indexing Special Interest Group of the American Society of Indexers and is president of the New England Chapter of ASI for 2006. Prior to freelancing, she worked at Thomson Gale (previously Information Access Company) for over ten years as a periodical database indexer and controlled vocabulary editor. Heather has a BA in government from Cornell University and an MA in Near Eastern Studies from Princeton University. Email: heather@hedden.net

Cambridge Mozart Encyclopedia

Like many gold mines, the *Cambridge Mozart Encyclopedia* sometimes yields up its nuggets reluctantly. Who, instinctively, will look up an alphabetical entry called 'Shorter Piano Works', let alone 'Smaller Church Works'? But there are welcome surprises: while looking for an article on Melodrama (which I eventually found under Monodrama, not cross-referenced), I came across Neal Zaslaw quoting an inspiring account of the Fantasia K608 by Ignaz von Seyfried, which was new to me. If you want a particular piece, look up the summary index and you may or many not be lucky ...

Review by Nicholas Kenyon of the *Cambridge Mozart Encyclopedia* (2006), reproduced by permission of the TLS.

SI T-shirts



You can never have too many T-shirts, so why not buy a Society of Indexers T-shirt, featuring the SI logo, and advertise your profession during the summer months. Excellent quality, they wash well and come in an attractive range of colours and a wide range of sizes, to fit indexers large and small.

For more information, visit
www.indexers.org.uk