Web Indexing Tools
A Comparative Review
by Kevin Broccoli

Excerpted from

Web indexing tools

by Kevin Broccoli

It wasn't too long ago that the concept of creating an index for hypertext documents was completely foreign. However, many webmasters, corporations, and new media publishers are seeing the clear benefits of using human indexers to design and write indexes to their Web sites and intranets, hoping for superior results in information access and retrieval. Technical writers may also be called on to provide indexing for Web sites.

What type of software is available for indexing of hypertext documents? What are the benefits of using one tool over another? This article will address these questions. We won't be giving detailed instructions on the use of each program. Rather, we will highlight the pros and cons of each tool, and the best situations for their use.

In the beginning

As I mentioned, hypertext indexing is rather new. It stands to reason, then, that the software used to produce such indexes has not been around for very long. Many pioneers in this new method of indexing lacked tools designed specifically for producing such indexes. Many indexers used HTML editors such as Homesite (http://www.allaire.com/products/homesite/40/hstour/index.cfm) or FrontPage (http:/ /www.microsoft. com/catalog/display.asp?site =768&subid=22&pg=1).

The problem with these tools was the complexity involved in the indexing process. Instead of simply indenting as you would in a word processor, for instance, it was necessary to use HTML coding to produce the indents. When the content of the index was completed, what lay before you was a visual nightmare of HTML coding and the text of the index itself. Finding and editing mistakes in the HTML tagging was like finding a needle in a haystack. It is quite easy to make an error in the tagging if you are doing the index "by hand" in an HTML editor. The fees that professional indexers needed to charge for such tedious, laborious indexing were often beyond the budget of many companies. And, of course, it wasn't very fun.

Let there be tools

Finally, indexing tools for Web sites have appeared. In this article, I am going to discuss three of them. Some of these tools are used exclusively for index creation, others are used for multiple purposes. The chief benefit of these tools is that the indexer does not have to bother with all of the tedious tagging that would otherwise need to be done using HTML. The tools we will discuss do the tagging for you, or create an online HTML help interface for the index. They are discussed in the order in which I discovered them, not necessarily the order in which they were released.

HTML/Prep by Leverage Technologies

(http://www.levtechinc.com/htmlprep.htm),
Price: $65.00

HTML/Prep is a DOS utility that basically converts a tagged ASCII index file to a set of HTML documents. The index is not created directly in the software, but rather in a word processor, or dedicated indexing software such as Sky Index, Cindex, or Macrex. HTML/Prep is a conversion utility that runs in DOS, and screen shots aren't really helpful to show its use.

Here's a sample output of a Web site (the American Society of Indexers) indexed using HTML/Prep:

Index

Numbers following the entries signify the number of times the subject is referred to in the text; NOT to a page number.

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | Y | Z

A

AAP. *see* Association of American Publishers
AcqWeb's Directory of Publishers and Vendors, 1
Acronym Directory, 1
address, ASI, 1
ADE. *see* Association for Documentary Editing
administrative office, ASI, 1
Adobe Web site index, 1
AEJMC. *see* Association for Education in Journalism and Mass Communication
AESE. *see* Association of Earth Science Editors
African American Writers Guild, 1
Alcohol Industry and Policy Database, 1
aliaINDEXERS (discussion group), 1
ALIWEB (database of Web sites), 1
Allied Health Database (Cinahl) thesaurus, The, 1

Here's the basic process:

1. Write and edit the index in whatever program is preferred. Instead of numbers, you use URLs as reference locators. (Yes, they're rather lengthy and cryptic, but try copying them from your browser and pasting them into your indexing software.)

2. Save the index as a .TXT file. (If you are using Cindex, it will also use the .HTP file that Cindex creates.)

3. In DOS, type the appropriate command along with the name of the file that you saved and any options that you may choose.

4. The file is then converted into HTML.

5. Edit the HTML for cosmetic reasons. (It's advantageous to have some knowledge of HTML. For instance, you might want a colored background to match the rest of the site, a company logo may be needed, etc.)

What it can do:

1. The letters of the alphabet are automatically generated, appearing above the index. Each letter links to its appropriate letter group below. Without this feature, the indexer would need to create these letters and each of the links by hand. Additionally, there is the option to have the alphabet appear between each letter group, so that the user will not have to scroll to the top of the screen each time a new word is sought. You can also provide links for automatic return to the top of the page. Another option is to put the alphabet in a separate frame above the index. This way, the user can click on whichever letter he wants, and the index scrolls in the frame below while the alphabet remains stationary. See the FICUS Web site index below for an example of this.

2. A list of main headings can also be generated, appearing in a third frame. This allows users to quickly scan only the main headings. When the user clicks on a heading, the index in the right frame scrolls to that main heading, where it appears in the index along with its sub-entries. This main-heading list can also be on a separate HTML page.

3. Cross-references are automatically linked to the main headings to which they refer.

4. Linked text can be displayed either by underlining and colorizing the heading's text (as most links are normally displayed in HTML documents) or underlining and colorizing *part* of the heading's text, so that the link is a letter, number, or graphic following the text. Notice, for example, that in the ASI Web site in screen shot #1, numbers are used to refer to the various occurrences of the index entry within the Web pages.

5. Since you are not creating the index directly in HTML/Prep, you have the freedom to use your favorite indexing software. Editing the index is not a problem since it would be done in your regular indexing software before converting the index with HTML/Prep.

6. It is very easy to use.

What it *cannot* do:

Although, as I mentioned, you do not have to worry about tagging mistakes using HTML/Prep, you do have to be concerned about bad hyperlinks. Remember that you either hand-typed or copied and pasted these URLs into your indexing software. Therefore, if you made typos, or you only copied and pasted *part* of the URL, the link will not work. Even if you are off by one letter, the link is bad. It is surprising how easy it is to make such errors.

1. If the site's webmaster decides to move pages around, putting some of them in a different directory, again your links will not work. You would have to edit each of the bad links in your indexing program and then re-

convert the index.

2. HTML/Prep will only convert your index file to a standard, back-of-the-book style index. It does not support the production of type-ahead scrolling indexes, such as those used in online help indexes for software.

Conclusion: HTML/Prep is an extremely useful tool. By eliminating tagging by hand, HTML saves oodles of time. Considering its utility, this tool is quite inexpensive.

HTML Indexer, by Brown Inc.

(http://www.html-indexer.com/index.htm), Price: $229.95

HTML Indexer is a stand-alone indexing program. You create the index entries directly in the software. There is no need to use Sky Index, Cindex, Macrex, or any indexing software. Below is a screen-shot of the interface:
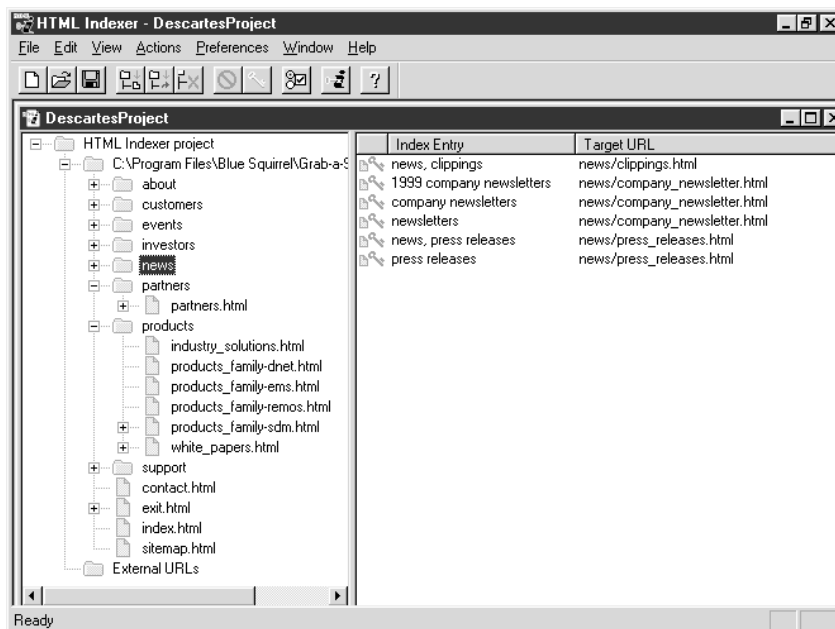
Basic process:

1. Specify the top-level directory (where your index file will go) and the files you want to include in your index project. HTML Indexer scans those files for the targets of your index entries, and displays them as a hierarchy of files and named anchors. The appearance is somewhat like Windows Explorer.

2. If you wish the index entries to refer to specific paragraphs on a page, create additional named anchors to which your index entries point. You would need to add such anchors in an HTML editor. If you want the index entries to simply refer to the pages as a whole, apply the index entries at the file level.

3. Evaluate the content of the HTML files to determine the best index entries:

· Where appropriate, accept the default entries created by HTML Indexer.

· As needed, add and edit index entries for selected targets.

· Exclude targets that don't require index entries.

4. Add or remove files as the project matures, creating entries for them and adjusting existing entries to maintain the quality and consistency of the overall index.

What it can do:

1. HTML Indexer, like HTML/Prep, produces the letters of the alphabet above the index. Also like HTML/Prep, you can provide links for automatic return to the top of the page.

2. Instead of just a hanging list (indented sub-entries), you also have the option to produce a run-in style index.

3. The index can be formatted for display in two columns, as in the back of a book.

4. Using the Shift and Control keys, you can select multiple files and anchors anywhere in the project, and then create index entries for all of them at once.

5. HTML Indexer embeds the index entries right in your HTML files. This is the greatest benefit of HTML Indexer. If the webmaster of a site decides to move the site's files to different directories, you simply open the site in HTML Indexer, re-compile the index, and you're on your way. There is no need to edit each and every URL that refers to the file which has been moved. Also, since the index entries are embedded in the HTML files, you are not typing in or copying and pasting URLs. The result is fewer mistakes and links that work *every* time. This aspect of the program is an incredible time saver. I've personally spent hours correcting such mistakes before acquiring this tool.

6. HTML Indexer 3.0 includes options for creating HTML Help and JavaHelp index files and (for JavaHelp) mapID files. This is truly a bonus! Not only can you create a back-of-the-book style index, but also the index files necessary to incorporate into an online help project. You would need additional software, though, to build the project itself, such as HTML Help Workshop, which isfree from Microsoft.

What it *cannot* do:

1. It can be difficult to edit the index because you cannot see the index until you compile it. Then you need to edit the individual entries that you want to change and re-compile the index in order for the changes to be visible within the index. Anyone familiar with any of the existing embedded indexing software available for print publications, such as FrameMaker and Quark, will recognize this problem.

2. There is no option for dividing the alphabet and the index itself by means of frames. However, you could do this yourself in an HTML editor.

Conclusion: Using HTML Indexer saves lots of time not only when creating the index, but also when there is a need to update it. Since the content of Web sites and intranets change quite frequently, this ability is invaluable. Additionally, the indexer can be assured that all of the hyperlinks are accurate. It costs a considerable amount more than HTML/Prep, but you get a lot more for those additional dollars.

RoboHelp HTML Edition, by Blue Sky Software
(http://www.blue-sky.com)
Price: $750

This software is one of the leading authoring packages that online help authors use to write help systems for software, the Web, or for intranets. The interface used to create the index is shownat the bottom of this page:

The process for index creation is:

1. Import the HTML files into your project.

2. Read the pages' content in the window to the right and decide on appropriate index entries.

3. In the Index Designer window, enter the index entry in the Keyword Text Box and click "Add" to add the entry to the index. (Actually, it's more effective to add all the index entries for a particular topic using the properties dialog box, and then go back and edit the index after adding everything.)

4. Compile your HTML Help project.

What it can do:

1. Create an index like the one often used with online help for software. See the example directly below.

The user of this index does not need to click on a letter of the alphabet. As soon as the Keyword Text Box is typed into, the index below automatically scrolls to the appropriate letter group and eventually the exact word as the user continues to type. This can be quite useful if,
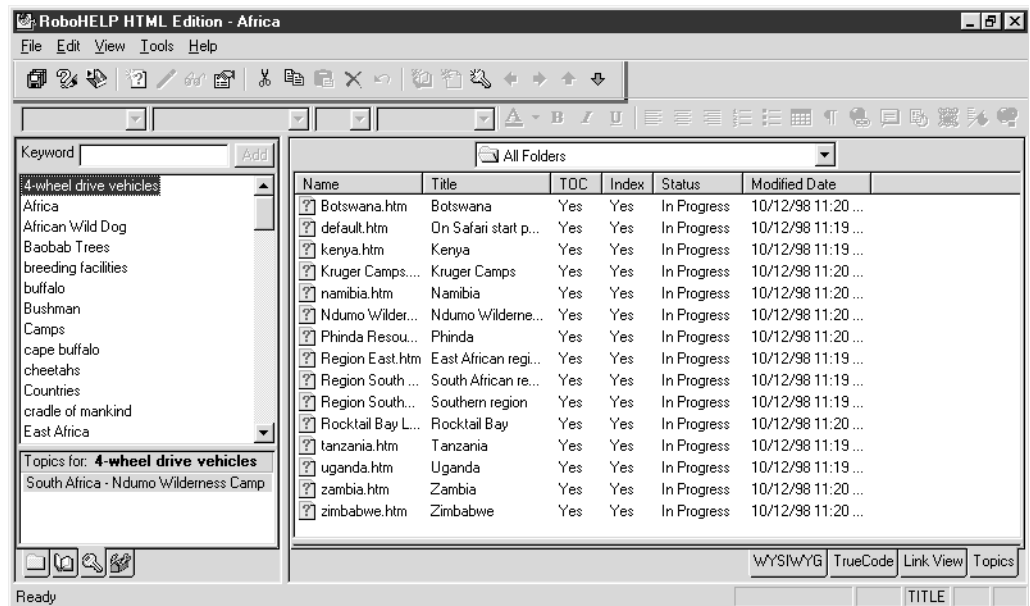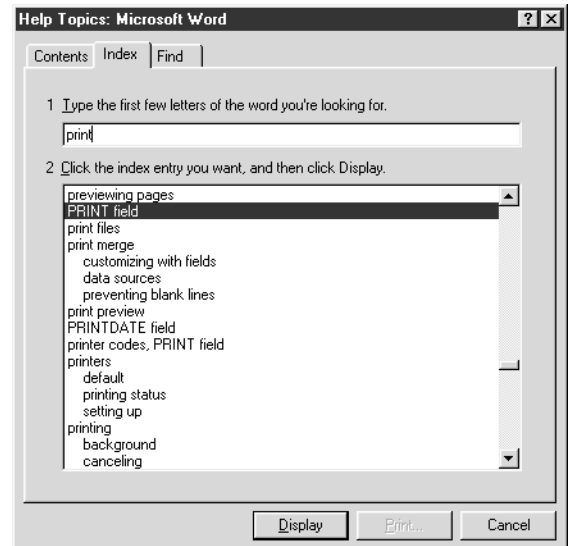
for example, the index is very lengthy and just one letter group may be many pages long. It also looks snazzy. :-)

2. Keywords can be added by dragging and dropping, which is nice and quick.

What it *cannot* do:

1. RoboHelp cannot produce back-of-the-book style indexes for the Web. Although the online help indexes it produces work well, they can take a long time to load, since they use a Java applet. Users may be frustrated with the wait and possibly even decide not to use the index. For this reason, it may be best to also supply a back-of-the-book style alternative. Since RoboHelp is not

designed for this, you would need to re-create the index using another tool.

Conclusion: RoboHelp is good at what it does. However, spending $750 and then only being able to create an online help index doesn't make much sense. It seems more logical to use a tool such as HTML Indexer so that you can create files for both types of indexes, and then use HTML Help Workshop to compile the help index.

The future:

Certainly, as indexing of HTML documents becomes more common, new tools will be developed. No doubt the above-mentioned tools also will be enhanced to better address the needs of information access and retrieval.

By becoming familiar with HTML, indexers and technical writers can find ways to custom-design their HTML indexes to fit the style of the site or intranet that they are indexing. By such intelligent construction, users will experience quick and accurate results to their search for information.

Kevin Broccoli is the owner of Broccoli Information Management (www.bim.net). He creates indexes for web sites, intranets, online help, image archives and publications. He can be reached at broccoli@bim.net.