

New Products

HTML Indexing Freeware: XRefHT32

XRefHT32, pronounced “shrefT,” is a little-known but quite effective tool for indexing web sites. It was created several years ago by Tim Craven, a professor of information science at the University of Western Ontario, and he continues to maintain it. The latest update to the software was made in July of this year.

XRefHT32 is available for free at <http://publisb.uwo.ca/~craven/freeware.htm>. Documentation for the freeware is also available.

XRefHT32 is a Windows-only, stand-alone indexing tool comparable to HTML Indexer that is used to create structured alphabetical indexes with hypertext entries. Rather than having page-number locators, main entries and subentries are linked to the web page, or named anchor within a page, to which they refer. The indented structure of subentries is automatically created by bulleted lists. An example of a web site index created with this program is the index of the site of the Birkbeck College of the University of London: www.bbk.ac.uk/help/index.

The software program’s unglamorous name is derived from cross-referenced hypertext in 32-bit Windows. When used in conjunction with free thesaurus management software by the same developer, XRefHT32 will add hypertext cross-references to indexes automatically. When used alone, cross-reference links in XRefHT32 merely have to be created manually.

How It Works

The interface is similar to that of the indexing tool SKY Index, whereby index entries and locators are entered into rows and columns:

The term “Heading” is used for the main entry, “Subheading” is used for the subentry, and “URL” refers to the locator. Usually the URL is simply the web page file name, such as “contact.htm,” but it could additionally include the folder or directory name if the site has subfolders. URLs can also include an anchor name for a more specific location within a page. A complete URL beginning with “http://” for an external web site link could also be inputted here.

Extracting Web Page URLs with Page Titles, Anchors, and Headings

You can manually type in entries and corresponding URLs.

However, since your pages are already in electronic form, there is no need to retype the URLs or even the headings.

Instead, you simply load the pertinent information from web pages into the program by selecting from the File menu “Extract titles from...” or, for anchors, “Extract targets from....” You may include files other than .htm or .html, such as .php, .jsp, or .asp. All of the selected files will then appear in your XRefHT32 program window list. File names are automatically added to the URL field.

Page titles, where they exist, are automatically added to the Heading field for you to edit (or rather, replace) as the main entries in your index. You can delete lines of index entries or add lines for double-postings and the like.

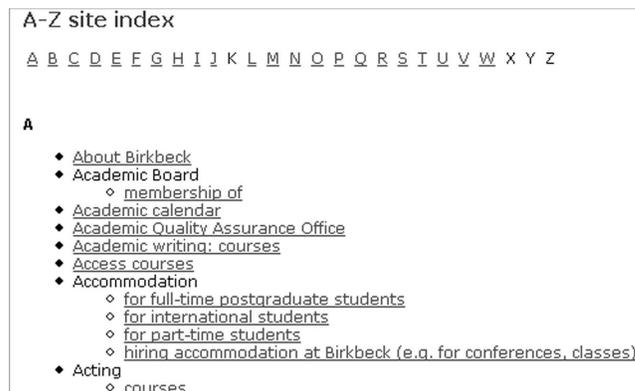


Figure 1. A preliminary screenshot from the Birkbeck College web site index, an example index written with XRefHT32 at www.bbk.ac.uk/help/index.

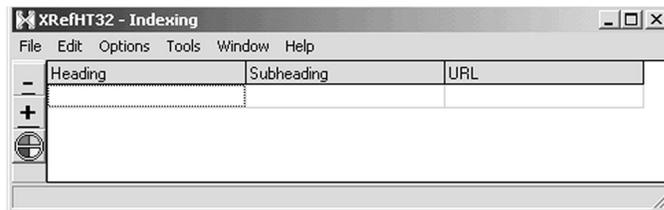


Figure 2. XRefHT32 indexing interface.

Anchors are used in indexing so that hypertext terms can link to sections that may be far from the top of a web page. As anchors in web pages tend to be created inconsistently, insufficiently, or not at all, XRefHT32 also has a feature for adding web page headings to the list of Headings for indexing based on HTML heading codes, such as: `<h1>Main Topic</h1>`.

The indexer can create anchors manually or use yet another feature of XRefHT32, which will automatically create anchors at all the headings.

Advantages

- A feature that automatically inserts named anchors at headings in web pages that lack anchors. Specify the file(s) and the software will add anchors with the names derived from the text of the heading. This allows indexers with no knowledge of HTML to create indexes.
- Find/Replace
- Spell-check
- Options to output the index either as a single web-page index or as separate pages for each letter of the alphabet

Disadvantages

- Poor functionality for creating additional entries. You cannot add a duplicate entry as a template for creating a new term, such as for double-posting, nor can you automatically invert main and subentries. You can merely copy and paste.

New Products

- Subfolders of files to be indexed have to be added individually for extracting. As such, XRefHT32 may be more practical with smaller sites.
- You cannot automatically add linked cross-references without additional software but can only create them manually using HTML or an HTML editor.
- The documentation is limited to brief descriptions of the menu functions and does not answer questions on how to perform the steps of creating an index.
- There are some other minor bugs, and the program is simply not quite as sleek and easy to use as commercial software is.

Conclusions

XRefHT32's ability to extract headings and automatically add anchors is a powerful and unique feature in web-site indexing tools. It should not be misused, however. It is a good idea to extract the headings to see what they

are, but it may not always be a good idea to add anchors to all of them. For one who has no skills in indexing, over-indexing may result. For the indexer who has no skills in creating anchors, 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 development tool and to create cross-references automatically. For most web-site indexes, a thesaurus is not necessary. But for a web-based index of one or more periodicals, a thesaurus could be quite helpful.

Since it lacks the term manipulation input features of commercial indexing tools, XRefHT32 may not be as practical for very large indexes or for the indexer who does multiple web-site indexes. For small, one-time projects, however, this tool is very useful.

All in all, for a free program, XRefHT32 is very impressive and worth trying. Perhaps if

more indexers discover it, we will find more good indexes on the web.

For more information about Craven, the developer of XRefHT32, see the 12-page (printed) index of his site at <http://publish.uwo.ca/~craven/craven.html>.

Other freeware packages available from Craven include:

- XRefHTMW, an HTML indexing utility
- TheW32, thesaurus management software
- NEPHIS32, a compiled program written in Delphi 6 for generating index displays
- NEPHIS (NEsted PHrase Indexing System)
- THSRS (THEsauRuS maintenance)
- and more, including word and phrase extractors, and abstract writing utilities

For more information:

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

— Review by Heather Hedden,
Hedden Information Management