

Workshop at the 'Partnerships in Knowledge' Conference, Canberra, 20-23 April 2001 hosted by AusSI ACT Branch and the Canberra Society of Editors. Published in *Australian Society of Indexers Newsletter* 2001, vol. 25, no. 11, p. 105.

Website indexing with HTML Indexer

Colleen Mock

The workshop, presented by Glenda Browne, consisted of two sections. First, Glenda introduced the principles of indexing in a web environment, highlighting ways in which web indexing is different from print indexing, and citing examples of websites that illustrated web indexing principles. The basic principles of indexing in print were reviewed as they are equally applicable to web indexing. The second section consisted of a demonstration of one Web indexing tool, HTML Indexer, in which workshop participants worked on a trial exercise indexing the December 2000 issue of the AusSI Newsletter. A set of course notes summarised the presentation and demonstration.

The workshop was based on material in the book *Website indexing: enhancing access to information within websites*, by Glenda Browne and Jonathan Jermey. Further [information from the book](#) is available. Additional references cited were ' *Information Architecture for the World Wide Web* ', by Louis Rosenfeld and Peter Morville, and a chapter in ' *Beyond Book Indexing* ', by Seth Maislin.

A number of tools provide access to information on the Web, each with their strengths and limitations. Search engines, metadata, and library-style categorisation and classification of websites are useful for searching the whole web, or for constructing subject gateways, while back-of-book style indexes are best suited to individual websites and documents within websites.

Presented in a familiar book-style format, back-of book style indexes have advantages over other information access tools at a website level because they provide immediate access to selected and specific information through multiple entry points, sub-divisions and cross-references, and are browsable, thereby overcoming problems of misspelling and inexact matches. Their disadvantages are the cost of using skilled indexers, the need for updating and for large sites or groups of sites, for coordination and quality control, as well as the time delay between creation of the resource and its addition to the index.

Basic policies of print indexing apply equally to web indexing. Issues of consistency of usage, user audience, purpose of index, depth of indexing, selection of material, format, filing order, whether multiple indexes or single index, type of entries to include, use of cross-references must all be considered in web indexing. Additional policy issues for web indexes include format in which index will be supplied, responsibility for loading, and updating and archiving schedules and arrangements.

Web indexing has some important differences from print indexing. Using the power of Internet technology, website indexes provide links instead of the page references of print indexes, and those links can be to a wide range of material both internal and external to the website, including documents, periodicals, multimedia collections ([Online & Ondisc Conference '99](#)) and databases.

Decisions need to be made on what level to index to, where the link will point to – top of article, section or paragraph, how much material will download with each link, the use of frames, and whether external links will be included or not, considering the issue of checking their currency.

The Web offers the facility for innovative ways of organising indexes – geographic instead of, or as well as, the traditional alphabetical (e.g. Australian Libraries Gateway). There are also various options available for display which need to be considered in index design.

Features of a good website index were covered. The loss of context in a web index compared with print means that extra sign-posting and navigational features are required. In an alphabetical index, the alphabet bar is always displayed, and major headings may also be displayed in a highly structured index. Fonts, colour and explanatory text should be used to distinguish types of links – whether to a major piece of information, just a passing reference, or to an image. Ideally multiple locators should be avoided. The index should be linked to and from the home page, and to other main areas of the website. Introductory and searching notes are important, and a feedback link useful. For a well-designed index see the AusSI indexes.

A number of web indexing tools are available, including plain HTML, WinHelp-style programs, databases (e.g. www.ca.com), CINDEX plus HTML /Prep, and HTML Indexer .

HTML Indexer automates some aspects of website indexing on IBM-compatible machines. The program creates default index entries for all selected files, and for all named anchors within those files. The defaults can be edited or overwritten, and must be saved if required. The index entries are stored in the source HTML files, so the index can be refreshed when files are added or moved. As the index entries are included in metadata tags they can be found by search engines.

The workshop exercise demonstrated the basic features of HTML Indexer, including selection and viewing of source material, blocking non-content files, building the default index, deleting and editing default entries, adding entries, setting filing order, cutting and pasting text sections from the source file, creating subdivisions and cross references, using Preferences to set style settings, and selecting output browser for output.