centred design projects for a diverse range of clients, such as Microsoft, Land Rover, Diageo, Amway, and the DTI's Small Business Service. He has lectured in the UK on HCI, internet applications, e-commerce and digital media. John has over ten years experience in usability and web development in the public and private sector, both in the UK and USA. With a love of document structure and web standards, he enjoys specialising in web accessibility and, sadly, takes pleasure in reading W3C Recommendations. He can be contacted via: **john@optimum-web.co.uk**

Feature Article:

"Enhanced Site Searching with an A-Z Index"
By Heather Hedden

As web sites have grown, so has the challenge to help users find information within a site. There are two ways that a user may look for information within a site: navigation and searching. Navigation is the exploration of a site to find out what information is available. Searching is the function of trying to find information on specific topic. In recent years the specialty of information architecture has emerged, which has significantly improved site navigation. Improvements in search capabilities, however, are still needed for many sites. Because the web is a relatively new medium, new techniques and technologies, such as search engines, tend to be applied. But traditional methods of searching, with slight modifications, such as the A-Z index, can also work very well in serving the search needs on web sites.

Drawbacks of Search Engines
The most common method to enable user searching of site is by adding a site search engine. Search engines do have significant drawbacks, though.

Site search engines may not retrieve enough or any pages. Search engines for the entire web usually produce satisfactory results in the quantity of pages, as users generally want "some information about" a subject, and this can typically be found on some of the numerous pages retrieved. If many good pages are missed by the search engine, the user usually does not notice or care. Within a web site, however, the number of pages is relatively small, so a simple search engine search 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 subject that the user enters is worded differently from the references to that topic within page text.

Site search engines may retrieve too many irrelevant pages. 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. The 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. In the end, the quality of the search engine results reflects the sophistication of the search string entered by the user, which cannot be controlled. In the A-Z index, on the other hand, the quality of the results reflects the sophistication of the indexer.

Site A-Z Indexes
A-Z indexes are created not by machines, but by humans who take care to add index entries only to pages on which good information about the topic appears. In this way, the indexing of topic words mentioned in passing or out of context is avoided, boosting the overall relevance and quality of the index itself.

An A-Z Index offers an alphabetical list of "entry point" topics through which the user may browse and select. In an index at the back of a book or manual, the entries are followed by page numbers. On a web site, the entry points are hyperlinked to the appropriate pages, and often to named anchors within web

pages for an even greater level of detail in indexing.

As with book indexes, a site index may contain multiple entries, each worded differently, that point to the same page, or page and anchor. This approach is used to cover all the different ways a user may think a topic is named. Indexers call this feature "double posting." It covers synonyms, such as "cars" and "automobiles," and the different word order of a phrase, such as "automobile engines" and "engines, automobile." The browsable nature of the index solves the problems that might arise from incorrect or variant spellings, and singular vs. plural usages that the site user might choose. In addition, there is often a second level of terms, called "sub-entries," that are listed and indented under some of the main entries.

A web A-Z index is typically a single, long HTML page, although it could be broken into separate pages for each letter of the alphabet if it were long. At the top of the page, a horizontal list of the letters of the alphabet usually appears. The user makes a selection from this list, and jumps to the appropriate section of the alphabetical index.

A list of some examples of A-Z indexes can be found on the Web Index Examples page of the Web Indexing Special Interest Group: http://www.web-indexing.org/web-index-examples.htm

Site Indexes versus Site Maps
Site maps are not an alternative to site indexes, since they act as the table of contents and serve a somewhat different purpose. But some web site owners and designers are unaware of the difference, and might even mislabel a site map as a "site index."

A site map tends to reflect the hierarchical structure of the web pages of a site with categorized web page titles. A purpose of a site map is to have a list of a site's web pages that can be quickly scanned in one screen view with minimal scrolling, without having to go through each menu and submenu one by one. A site map might not include all the pages of a web site. If the site is large, only the top few levels of the hierarchy would be displayed. The entries in the site map tend to be the page titles, but they could be modified slightly. Each page in the site map is listed in only one place. Like the navigation menu, the site map is to aid navigation rather than searching, and thus should not be seen as an alternative to an index.

There are software tools to aid in the creation of site maps, by extracting web page titles along with their hierarchical links within a site. The problem is that some of these tools also offer the feature of alphabetically sorting web page titles to create an "index," a feature that probably should not have been created. An alphabetically sorted table of contents in a book does not create the book's index, so this method should not be used for web site indexes. An alphabetized list of topics or names is useful; but a list of alphabetized page titles is not.

Site Indexes versus Taxonomies
Taxonomies are hierarchical classifications of terms, concepts, or topics, in a tree-like structure. On a web site taxonomy, the user typically clicks from one level to the next most specific level. An example of a very broad taxonomy is on the Amazon.com site http://www.amazon.com in the left-hand bar of categories under Browse.

A taxonomy's function and purpose falls somewhere between that of a navigation menu and that of an index. As with an index, the terms or labels of a taxonomy tend to be carefully selected, and a narrower concept can be placed in more than one place in the taxonomy. The purpose of a taxonomy is more that of guided search, than for site navigation.

Certain types of sites or parts of sites are better served by taxonomies, and others are better served by indexes. If most of the content is dealing with a narrow subject area, such as a web site devoted to information on a product line, heart disease, or historic films, a taxonomy might work better than an index. An index, on the other hand, serves best a site with varied types of content. There has been a lot of interest recently in taxonomies for aiding the organization and retrieval of information in large web sites and intranets. As a result,

taxonomies might end up being implemented where an index would actually serve better.

Site Suitability for an Index

The most suitable sites for A-Z indexes are those with repeat visitors (intranets, government sites, organizations, periodicals, and companies offering recurring services), of a medium size (perhaps 30 to 600 pages), with not too many pages changing too frequently, and with a rich and varied content.

Who does Indexing?

Creating an index is more complicated than creating a hierarchy of categories. To become competent at indexing requires appropriate training. Information architects with backgrounds in library science and a good sense of labelling, however, could probably pick up indexing from reading a good book on the subject. In addition, a tool for automatically embedding the index is recommended.

If you don't want to invest the time and energy in learning indexing yourself, it's probably best to contract a freelance indexer. Most of the professional associations of indexers, listed below, maintain searchable databases of freelance indexers. Limit your search to HTML or web indexers. The nice thing about web site indexes is that samples of an indexer's work are usually accessible online, so you can easily evaluate a potential indexer's work.

- Society of Indexers:
  http://www.indexers.org.uk

- American Society of Indexers:
  http://www.asindexing.org

- Indexing and Abstracting Society of Canada:
  http://www.indexingsociety.ca

- Australian and New Zealand Society of Indexers: http://www.aussi.org

There is also a database exclusively of web site indexers on the Web Indexing Special Interest Group of the American Society of Indexers:

http://www.web-indexing.org/contract-indexer-search.php

Conclusions

Search engines are often indispensable on a very large and changing site. Yet a site certainly can have both an A-Z index and a search engine. This would make sense if certain pages of the web site are skipped in the index or not indexed in detail because they are constantly changing.

As web sites, and especially intranets, tend to be large and complex, different kinds of indexes on different parts of the same site may be the best solution for the site. An A-Z index or a directory could be applied to just the top few levels of pages of the site for overall site navigation, while a database could be applied to a section of numerous frequently updated articles. A thorough A-Z index could be applied to a book-like section, such as a policy manual. Finally, a taxonomy could be applied to collection of similar pages dealing with various entities, such as a product directory.

Just because A-Z indexes follow an older style, that of book indexes, does not mean that they are too outdated to serve web sites well.

Heather Hedden, principal of Hedden Information Management (www.hedden-information.com), is an information architect and indexer with a specialty in web A-Z indexes. Previously she worked on thesauri and taxonomies as senior vocabulary editor at the Gale Group. She is vice president of the New England Chapter of the American Society of Indexers and coordinator of the Web Indexing Special Interest Group. She can be contacted at heather@hedden.net.