Intranet Indexing
Finding Content from A to Z

HEATHER HEDDEN

Web site designers and information architects take special care to design good navigational structures with hierarchies and site maps to aid users in navigating sites or intranets. However, because intranet users are repeat visitors, they are likely to be less interested in navigation and more interested in searching. The difference? Navigating means getting to know one's way around a site, becoming familiar with the content in general, and knowing approximately where to find what. Searching involves trying to go directly to some specific piece of information. Once intranet users have garnered a basic lay of the land, they are much more likely to be searching for specific information than to be navigating about casually. Thus, a good search tool is essential for an intranet's usefulness.

Yet all too often the most obvious, simplest, and yes, cheapest search tool is applied: a search engine. While your basic search engine provides a practical means for searching the World Wide Web, it is not necessarily so effective for searching a single site or intranet. Searchers can miss appropriate pages if they enter incorrect terms, or can retrieve irrelevant pages due to matching terms taken out of context. A far more effective alternative is a browsable, alphabetically arranged index, like those found at the back of most nonfiction books, or like cumulative indexes to periodical articles. Multiple entries worded differently and cross-references guide searchers to desired concepts and pages.

In implementing a manually created A-Z index on an intranet, consider three different formats: the back-of-the-book style index, the index style used in citations to periodical articles. and the relational database. You can also use a combination of formats, depending on your intranet's size and content type.

BACK-OF-THE-BOOK FORMAT

The traditional back-of-the book index is a browsable alphabetical list of specific topics and names. It typically has two levels: entries and subentries. Occasionally, there is a third—sub-subentry—level. Each entry or subentry is followed by one or more page numbers for pages on which a discussion of the topic can be found. Of course, on an intranet or Web site, there are no page numbers; rather, the text of the entry or subentry is hyperlinked to the Web page featuring a discussion of the topic. Back-of-the-book style indexes also may be used for online editions of periodical indexes. Instead of page numbers following an entry or subentry, issue numbers or dates follow, with Web links to articles mentioned.

One advantage of using this type of index is that it is user-friendly; everyone knows how to use the index at the back of the book. The browsable nature of the index allows users to choose a topic from a list rather than having to guess what to type into a search box. Users can find specific topics with the combination of entries and subentries.

On the other hand, the back-of-the-book style index is not dynamic. Books are published complete, and are somewhat finite, while intranets constantly change. When a new page is added, the intranet's index needs to be updated. No matter what software tools you use to help generate the index, maintenance remains an issue.

The other key disadvantage is that, while multiple page numbers may be listed after an entry in a book index, on the Web, each hyperlinked entry or subentry can link to only one page. If there are multiple pages on the same topic, use one of several work-around solutions:

Site Alphabetical Index

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

A

address of the library
art exhibits
    current exhibit
    past exhibits
author events

B

blog for readers' reviews
Board of Trustees
book club kits
book discussion forum
book donations policy
book groups
book groups for teens
book lists
    children
    teens
books on tape
borrowing policies
borrowing policies, for nonresidents
borrowing records
bulletin board postings
business databases

C

calendar of events
cards, fees
Carlisle history collections
catalog use help

An excerpt of an example of a small A-Z index for a public library Web site.

- *Index only the primary discussion of a topic, and leave lesser discussions of the topic that cannot be differentiated by subentries unindexed.*
- *Add a third level, a sub-subentry.*
- *Create additional subentries to differentiate each reference to a topic.*
- *For periodical articles, create additional subentries with different dates displayed.*
- *Link the entry to an intermediate page or frame that provides a list of pages where topic is discussed.*

PERIODICAL CITATION FORMAT

A citation index style is usually preferable to a strict back-of-the-book style index if there is a tendency for multiple citations on the same topic. This would occur if the index covers a long time period or multiple periodicals or newsletters. The citation index may be in the form of a browsable alphabetical list of topics and names, like a back-of-the-book index. A list of citations, rather than a list of page numbers, follows each entry. This is the style we are familiar with in the printed indexes to periodical literature, such as the *Readers' Guide to Periodical Literature*. Typically, the index is updated on a periodic basis.

The problem is that, over time, as the number of citations grows for each entry, the index no longer remains practical for browsing. Often a separate index is then created for a period of a year or span of years, and, as a result, the user has to repeat a search for each time period. Another way to simplify browsing when the citation list under each entry or subentry gets too long is to move the citation list onto another Web page. Users, however, prefer having direct access to the content, rather than having to navigate through an intermediate citation-list page. Ultimately, even though a citation index includes new articles, the index itself is still static like a book-type index.

RELATIONAL DATABASE INDEXES

A solution to the problem of static indexes, especially where content is frequently being added, is to integrate the index into a relational database. When new content is added, a new database record is all that needs to be created. At a minimum, there is a database field for the article text and for each index term. Usually, there is also a field for the author's name, the newsletter name if the index includes more than one newsletter, and additional attributes, such as article type (e.g., feature, review, product announcement, etc.). A search on an index term will retrieve all articles indexed with that term. This type of dynamic database index is most often found in commercial subscription databases, online library catalogs, and content management systems.

Database indexes usually implement a set of search boxes for the interface instead of an alphabetical browse. This enables users to search for a combination of fields at once—index topic, author, article type, date, etc. Sacrificing the alphabetical browse does not have to be the case, though. The Montague Institute's A-Z index is database-driven, yet has a browsable interface for the topic list.

However, there is a tradeoff. The problem is that the user-interface generally cannot support more than one browsable list in a single view, unless the lists are shorter than a screen or scroll-boxes are used. But the fact is that most A-Z indexes are much longer than one screen or what can fit practically into a scroll box. Thus, you must make a design choice between preserving the user-friendly, browsable A-Z index or adding the sophistication of a simultaneous multiple-criteria search. Furthermore, standard databases also do not support the combination of subentries with main entries in a search, although custom programming at a cost can make it possible. An alternative is to search a combination of two or more topic terms. But then again, you sacrifice the browsable interface and no longer provide users with an A-Z index.

WHICH TO CHOOSE?

A back-of-the-book style index works well for intranets that are not too large or too dynamic and that contain non-repetitive content. In other words, if there are not several pages that discuss the same topic, as a collection of newsletters might.
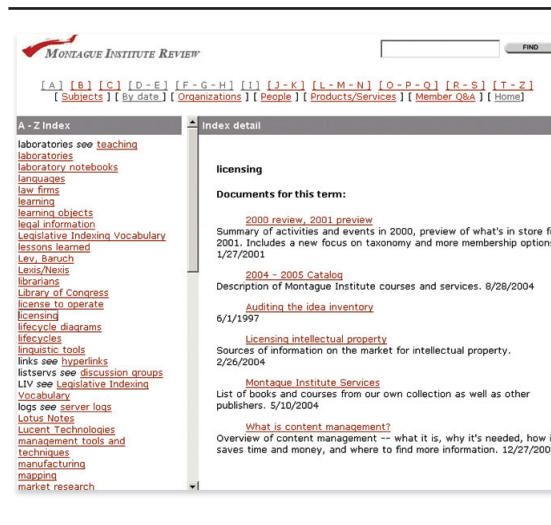
A citation list style index might work for smaller, somewhat static intranets that do contain repetitive content, such as a collection of articles, notices, or images.

With either of these styles, a contract professional indexer can write the index and guidelines for maintaining it.

A relational database index is the best solution for large, dynamic intranets that also contain content that is of a similar format, such as collections of articles, images, or product specifications. This type of index is often best used for subsets of content within intranets, rather than for entire intranets. A combination of a Web developer skilled in databases and an information professional (information architect, librarian, or indexer) is needed to set up a database index.

The choice of an intranet A-Z index format should not be based solely on the skills or software tools readily available. Rather, your intranet's organization and nature of content should be deciding factors. For example, while back-of-the book indexes usually follow strict editorial policies or styles, in the Web interface, there is considerable freedom to modify and adapt the style to what is most suitable for the content and the users. **I**

A page from the Montague Institute Review index, which uses frames and a database-driven index that still offers A-Z ease for users.

HEATHER HEDDEN (heather@hedden.net),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 VP of the New England Chapter of the American Society of Indexers and coordinator of the Web Indexing Special Interest Group.