

Book-style indexes for websites

Heather Hedden, Carlisle, MA (USA)

Site A-Z indexes are back-of-the-book style indexes with hyperlinked entries, and they are a useful method for searching for information in a website. This article examines the types of sites for which an A-Z index is most suitable and then compares website indexes with book indexes with respect to structure, locators, subentries, and cross-references. Certain modifications need to be made to adapt an index to the web. Although not complicated to create, continuous updating is an issue, and site indexes are not applied as much as they could be.

Buchregister als Vorbild für Website-Indexe

Site-Indexe entsprechen Buchregistern mit Hyperlink-Einträgen und sind eine nützliche Methode zur Informationssuche innerhalb einer Website. Dieser Beitrag untersucht die Arten von Websites, für die sich ein Site-Index am besten eignet, und vergleicht Site-Indexe mit Buchregistern bezüglich Struktur, Fundstellenangaben, Untereinträgen und Querverweisen. Bestimmte Modifikationen sind nötig, um einen Index den Erfordernissen von Websites anzupassen. Obwohl es nicht schwierig ist, einen Webindex aufzubauen, ist die fortlaufende Aktualisierung eine Kernfrage. Site-Indexe werden nicht so häufig eingesetzt, wie sie könnten.

To make information easily found in books, there has evolved the standard format of (1) a table of contents to provide a page-ordered list of the book's chapters as a kind of outline, and (2) an alphabetical index, usually at the back of the book, listing all the topics and names, and their synonyms or variant forms, that are worth mentioning, along with their corresponding page numbers. Not only has it become standard to have a table of contents and an index in a nonfiction book, the styles of indexes have become somewhat standardized as well. While there exist different formats, such as hanging versus run-on subentries, a freelance indexer has no difficulty accommodating a limited number of

styles for various publishers, even in different countries.

Making information easily found in websites, however, is a very different matter. As more and more information gets put on websites, and the importance and size of websites has grown, website owners and designers have attempted various methods of making information on the sites more "findable." This fast-growing and changing medium, however, does not have the established standards that exist for printed books. Electronic data and hypertext links can be exploited for various new methods by which users can locate information. Methods include:

- navigational menus, which may include second and third level drop-down menus
- search engines, which may or may not take metadata keywords into consideration
- hierarchical taxonomies, which may or may not make use of a thesaurus or a content management system
- site maps, which might list all pages or only the top two or three levels of pages
- site indexes, which may or may not follow book index styles

Not only do different software technologies allow various search and navigation methods, but the immense variety in the kinds and sizes of websites require different methods for information searching. Finally, as we are aware from books, one method of finding information is never sufficient. Just as books have both a table of contents and an index, a website should also have two or three methods, perhaps more, for its users to locate information. A website index may often be one of these methods.

Definition of website indexes

Although technology can support more elaborate methods of searching, a site index is an option that is definitely worth considering. A site index, as we define it, is an alphabetical arrangement of topics, similar to that of a book's index, where each topic is hyperlinked to the referred web page. A site index is a method of search that is:

- familiar to users from the world of books, and thus very easy to use
- relatively low technology, relying only on HTML code, so that it can be implemented on the most simple website by people without programming/coding expertise
- more accurate in its retrieval results than site maps, navigation menus, or search engines

The term "website indexes" could have different meanings. Thus, for clarification, the browsable alphabetical back-of-the-book style index on a website is often called an "A-Z index." The implication of "A-Z" is that there is an alphabetical browse view or interface. This differs from the ordered view of a site map that looks more like a table of contents. Clarification is needed because some web designers labeled site maps as site indexes. The A-Z index is also different from an index that involves a search box. Even if words entered into the search box are matched against a human-created index, thesaurus, or controlled vocabulary, the index is not visible to the user and therefore not browsable in an A to Z list.

On websites, A-Z indexes are typically called in English (listed in order of popularity):

- Site index
- A-Z index
- Topic index
- Alphabetical index
- Browsable index

Appropriateness of site indexes

Almost all nonfiction books can and probably should have indexes. This is not the case with websites. Some are too small or too large, some change too frequently, and some have content better searched through a taxonomy or a database rather than with a traditional alphabetical index. Website size and changeability are the most important factors, since maintaining an A-Z index can be a lot of work. The type of content and type of users may also be an issue.

Extremely small websites do not require indexes, and indeed a very small index

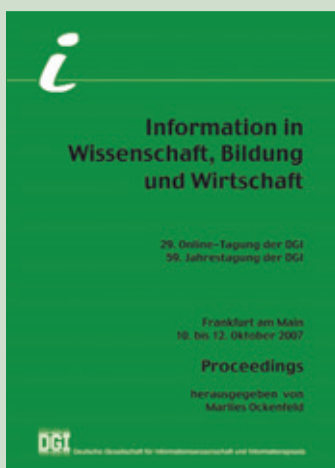
Information in Wissenschaft, Bildung und Wirtschaft

herausgegeben von Marlies Ockenfeld

Das Motto der DGI-Online-Tagung Information in Wissenschaft, Bildung und Wirtschaft ist der Rahmen für ein vielseitiges Programm, das Information Professionals aus vielen unterschiedlichen Aufgaben- und Einsatzfeldern interessante Einblicke in informationswissenschaftliche Forschung, praktische Anwendungen, das Berufsfeld und geplante Projekte bietet: Informationsportale, Informationsarchitektur, Patentometrie, Such- und Antwortmaschinen, Web 2.0, Dokumentation im Gesundheitswesen sowie E-Journals und Open Access, um nur einige zu nennen. Neben den Beiträgen, die aufgrund des Call for Papers eingereicht und nach einer Begutachtung durch das Programmkomitee als Vortrag angenommen worden sind, wurden insbesondere für die Auftaktveranstaltung und die beiden ersten Sitzungen über Trends, Herausforderungen und Perspektiven zusätzliche Fachleute als Rednerinnen und Redner eingeladen. Die begutachteten Beiträge sind als Volltexte enthalten. Von den eingeladenen Vorträgen gibt es Zusammenfassungen. Insgesamt enthält der Tagungsband 31 Langfassungen und elf teilweise ausführliche Zusammenfassungen.

Leserkreis

Informationswissenschaftler, Information Broker, Wissensmanager, Medienfachleute, Bibliothekare, Content Anbieter, Verlagsmitarbeiter, Indexer, Fachangestellte für Medien- und Informationsdienste, Informationsassistenten, Studenten



Marlies Ockenfeld (Hrsg.)

Information in Wissenschaft, Bildung und Wirtschaft. Proceedings.
Frankfurt am Main 2007, 350 Seiten, DGI-Tagungen – Band 9
ISBN 978-3-925474-60-6, EUR 50,- (für DGI-Mitglieder EUR 40,-)

Bestellungen an die DGI-Geschäftsstelle, Hanauer Landstraße 151-153, 60314 Frankfurt am Main, Fax (069) 4 90 90 96, publikation@dgi-info.de

does not work well. A minimum would be about 25 pages, but this also depends on how much information the pages contain. In any website, there will be some pages not suitable for indexing. Extremely large sites also pose a problem. For websites or intranets of over 1000 pages, not only is it a lot of work, but also by the time an index is complete, the site will likely have changed. This does not mean that large sites should not include indexes. Rather than indexing the entire site, individual indexes can be applied to individual sub-sections.

A major issue in website indexing is that, unlike printed materials, websites and intranets can change quite frequently. Content within pages changes and new pages are added and old ones deleted from time to time. Websites that change especially frequently, such as a website dedicated to a special event, should be avoided in indexing. A website with changing sections could still be indexed, but the indexer should omit pages known to be temporary and refer to pages whose content changes by general topic of the pages only and not to specifics.

A website rich in content and with a variety of content is better suited for indexing. Some sites may have a sufficient number of pages, but insufficient indexable content. These would include pages displaying mostly images (for graphic effect rather than as content), components of online games, or short directory entries. Sites that contain a lot of content but all of a similar type, such as a catalog of products or a directory of names or organizations, are also not so suitable for indexes. In some cases, all the information can easily be arranged in categories. A directory-type site might require an alphabetical list of names to look up, but this is not what we would consider a true index. Most sites that sell products do not need indexes if most of the pages comprise a listing of products. Customers tend to look up products by category, not by alphabetical names.

Websites that tend to get repeat visitors make especially good candidates for indexes. They are analogous to reference books, that researchers repeatedly go back to and where an index is especially appreciated. Of course, indexes can be useful on websites that receive many one-time visitors, but often the objective of such websites is to draw in the visitor to explore the site, rather than come in, get information, and then leave. Sites that have high repeat visitors, and thus are good candidates for indexes, include company intranets visited by employees, educational institutions by students, organizations by members, and municipal sites by residents.

Structure of a site A-Z index

The A-Z index is often just a single page within a website. But if the index is very long, it may be broken up into multiple pages, such as one for entries that start with each letter of the alphabet or for a range of letters of the alphabet. When an index page is long, occasional "Back to Top" or "Top of page" jump links are also useful additions to the page. These may be inserted after each letter section or after multiple letter sections if the sections are short.

Compared with a printed book index, which is usually in two columns to save space, there is no need to save space on a scrollable web page, so a site A-Z index can be left in a single column. Furthermore, since at least some scrolling is expected, it's actually preferable to have the website index in a single column, since it is less practical for the user to have to scroll up and down between two or more columns on the same page.

Whether a website index is on one page or divided into multiple pages, the standard method of navigating the entire index, which is too large to fit on one screen, is to present the letters of the alphabet in the initial display and have these letters hyperlinked to the section of the index that begins with that letter of the alphabet. The letters of the alphabet may be displayed only at the top of the index or index page, or also at the bottom of pages and in between letters, or in a separate frame that is always visible.

There is also a choice of style as to whether all the letters of the alphabet are displayed, including those for which there are no index entries, for example Q and X, or to only list letters for which there are index entries. In the Boston-IA topic index example (Figure 1), the letters with no entries (Q and Y) are simply omitted. Alternatively, if there are no index entries for a letter, then the letter is simply not hyperlinked and should display in a different color, or be "grayed out".

Hyperlinked Locators

In printed indexes, entries or subentries are followed by one or more page numbers (locators), referring the reader to the page of the book with the topic of the index entry. In a website index, there are typically no page numbers, rather the text of the entry or subentry itself is hyperlinked to the page within the website where the topic is discussed.

A complication with website indexes is that a hyperlink can go to only one destination. So, when relying on the



Figure 1: Example of an excerpt of a website index.

hyperlink of the text of the entry as the means of jumping to the source content, there cannot be multiple locators, in contrast to a series of page numbers following an index entry when the topic is found on multiple pages. Having hyperlinked page numbers following the entry is a possibility, but the page numbers themselves are rather meaningless in a website. Thus, it is rare to see them in website indexes. There are ways, though, to get around this problem in website indexes:

- Create additional subentries
- Reword the subentries
- Create sub-subentries
- Decide that one mention of the topic does not provide original information, so omit it from the index

As it turns out, the issue of how to handle multiple undifferentiated locators in hypertext is a greater problem in theory than it is in practice. Many websites are written more concisely than books and information is not repeated as often. While the linear nature of print documentation requires a concept to be introduced in an introduction, then discussed in detail, then summarized, web content, by contrast, tends to get to the point. If indexing a book converted to HTML, however, the problem of multiple locators is greater.

Not only do print indexes have multiple page numbers after many entries, but they also have page ranges, such as: pp. 67–70. Page ranges also cannot be indicated in hypertext, but in practice this is not much of an issue in website indexes. If several pages of a website all discuss the same topic, then they are likely to be sub-pages of an intermediate page introducing that section of the website. The index entry then can link to the intermediate page. Also, websites

have no restriction on page length, so if there is a lengthy discussion on one topic, it may be covered in a single, lengthy page. The general design of websites is to create one web page for each topic, no matter the length of the page.

Thus, the nature of traditional websites makes them relatively easy to index with a single locator/link per entry. Websites that are collections of articles, however, are more challenging. But the challenges are the same as in print when

creating a cumulative index of periodical articles.

Hypertext locators also have an advantage over page numbers. Instead of pointing to an entire page, index entries can be more precise by being linked to specific points (sections, paragraphs, etc.) within a page by inserting HTML anchors to link to within the page. The named anchors may already be present in the web pages, but often the indexer may feel the need to create a few more named anchors. Adding named anchors is the only additional task the website indexer has beyond writing the index itself. In most websites, where the pages are not unusually long, creating index entries that point to an entire page, and not a more specific anchor point, is sufficient.

Not every entry in an index is hypertext and linked. If there is a main entry with multiple subentries, the main entry typically is not linked to any pages, but serves as a gathering point for the subentries. This is the same style as in book indexing. In the Boston-IA topic index, the entry “adaptive technologies” is not hyperlinked, since it has no reference of its own, but is rather the gathering term for two subentries. It should be clear which entries are hyperlinked to content and which are not by use of different color/use of underline for the hypertext.

Orphan subentries

In book indexing, a single subentry under a main entry (also called an “orphan” subentry) is considered bad style. If there were a general discussion of indexes on page 4, and a more specific discussion of website indexes on page 6, rather than an index entry of the following:

indexes, 4
on websites, 6

in a book index, it would be entered as:

indexes, 4, 6

Since multiple locators are not supported in a hyperlinked website index, the index entry should be structured as:

indexes
on websites

Here the main entry word “indexes” would be hyperlinked to the general discussion on one page, and the subentry phrase “on websites” would be hyperlinked to the specific discussion of website indexes on a different page.

Cross-references

In a website index, not only are the entries hyperlinked, but so are the cross-references of *See* or *See also*. In website indexes, the term following the word *See* is hypertext, and the term before it is not hypertext. In the Boston-IA topic index excerpt there is a cross-reference, where the hypertext is underlined:

accessibility, web. *See* web accessibility

Clicking on web accessibility jumps to this entry lower down in the index.

See cross-references are used in print indexes, as an alternative to “double-posting,” typically when synonymous main entries have subentries so it would waste space to have all the subentries repeated under each double post. On websites, there is no concern for wasting space, so the *See* cross-references are less often used for this purpose, and equivalent variants, tend to be entered, unless the list of subentries is very long.

Another use of a **See** reference is to educate the user as to the preferred term. If this is the case, then such a *See* reference would also be used in a website index, but, if there are no subentries, there is no need to make the user jump somewhere else within the index before going to the source text. The *See* reference can name the preferred term, but the preferred term can link directly to the source text, rather than to its position within the index.

See also cross-references are just as helpful in website indexes as they are in print indexes, for guiding users to related terms. Whether, the link jumps to the referred term within the index or directly

Das Buchregister

Methodische Grundlagen und praktische Anwendung

von Robert Fugmann

Angesichts der automatisierten Aufbereitung von Buchinhalten in Suchmaschinen und Portalen gewinnen Sach- und Fachbücher wieder mehr Aufmerksamkeit. Sie leiden jedoch häufig unter fehlenden oder mangelhaften Registern. Dort wo es sie gibt, handelt es sich nicht selten um schlichte automatisiert erzeugte Textwortextraktionen.

Fugmann beschreibt die Fallstricke und Herausforderungen beim Registermachen und warnt vor einer Fehleinschätzung des Indexierungsproblems. Mit dem Standpunkt des Physikers Ludwig Boltzmann „Nichts ist praktischer als Theorie“ gibt er eine theoretische Einführung in die Problematik und weist auf Anfängerfehler hin, die man bei der Indexierung von Büchern vermeiden sollte. Gleichzeitig zeigt er am Beispiel des vorgelegten Buchs, wie sich informative Register anfertigen lassen.

Buchregister von Sach- und Fachbüchern müssen, so die Kernforderung des Autors, ihren Lesern ein Instrumentarium bieten, mit dem sie auf Entdeckungsreise gehen können, und nicht zuvor das Buch gelesen haben müssen, um einschlägige Inhalte aus der Erinnerung heraus wieder zu finden. Für diesen Dienst am Leser zeigt er neue Wege und Werkzeuge auf.

Leserkreis

Autoren, Indexierer, Content Anbieter, Verlagsmitarbeiter, Technische Dokumentare, Information Broker, Informationswissenschaftler, Studenten



Robert Fugmann

Das Buchregister. Methodische Grundlagen und praktische Anwendung
Frankfurt am Main 2006, 136 Seiten mit Sachregistern, DGI Schrift (Informationswissenschaft – 10)
ISBN 3-925474-59-5, EUR 25,- (für DGI-Mitglieder EUR 20,-)

Bestellungen an die DGI-Geschäftsstelle, Hanauer Landstraße 151-153, 60314 Frankfurt am Main, Fax: (069) 4 90 90 96, publikation@dgi-info.de

to the source text depends also on the presence of subentries. If there are no subentries under the referred entry, then there is no need to make the user click through another entry in the index.

A well-designed website index will use a combination of intra-index and direct-to-page links, as appropriate, for the **See** and **See also** references. The following table may also be used as a guide.

Table 1:

Main Entries	Subentries	
	None or Few	Many
terms are equal	no cross-references; repeat terms	perhaps cross-reference linked to a preferred term
one term is preferred	cross-references linked to text	cross-reference linked to preferred term

A-Z indexes on the Web

On English language websites, whether in the United States, Canada, the United Kingdom, or Australia, most of the websites with A-Z indexes tend to be either universities or government agencies. Called either “A-Z index” or “site index”, more often than not, these indexes unfortunately lack subentries and sufficient double-posting or cross-references. At a quick glance, it may be difficult to discern whether they are really indexes or merely alphabetically sorted web page titles. True book-style indexes can be found, but they are not as common. Collected examples of good A-Z indexes are listed on the Web Indexing Special Interest Group site: www.web-indexing.org/web-index-examples.htm.

On German sites, “A-Z Register,” “Site Index,” “A-Z Index” are also sometimes used. As with English sites, “Site index” is sometimes used to designate a non-alphabetical site map. The “A-Z” indexes are usually without subentries, double-posts, or cross-references. True back-of-the-book style indexes on German websites are thus still extremely rare. One example is the Deutsche Zentralbibliothek für Medizin www.zbmed.de/site/index.html. In this example, the subentries are not indented as in books, but follow an arrow symbol, and there are also some double-posts in this index. A good example of a German website index with indented subentries as in a book is that on the site of the Deutsches Netzwerk der Indexer www.d-indexer.org/si.html.

In conclusion, despite the development of new search and navigation technologies for websites, the traditional back-of-the-book style index still serves users and certain kinds of websites very well and could still be implemented more than it is. A site index needs to be continually

updated, however, as web pages are added to the site. While there is no standard accepted style for website A-Z indexes, certain modifications should be made to a book-style index to make it more usable on a website. Converting a book index to HTML, however, has the additional problems of dealing with multiple locators. By taking advantage of hypertext, a site A-Z index is even easier to use than a printed index. With the aid of website indexing software [listed in

Glenda Browne's article], indexers who write indexes for books can also write indexes for websites.

References

Hedden, Heather. Indexing Specialties: Websites. Medford NJ: Information Today, 2007.

“Best Practices for Style and Formatting of A-Z Browsable Site Indexes” on the Web Indexing SIG. www.web-indexing.org/practices.htm. [26 August 2007].

Website, Index, Software

Elektronischer Dienst, Sachkatalogisierung, Index, Website

THE AUTHOR

Heather Hedden, MA



an indexer and taxonomist was manager of the Web Indexing Special Interest Group of the American Society of Indexers (2005–2006) and president of the

New England Chapter of the American Society of Indexers (2006).

Hedden Information Management
98 East Riding Drive
Carlisle, MA 01741
USA
heather@hedden.net
www.hedden-information.com