

Changes in website indexing

Glenda Browne, Blaxland (Australia)

Website indexing became important in the 1990s, as indexers, librarians and web managers experimented with different approaches for making the information they were providing on the Internet more accessible. The tools for creating A to Z indexes have changed over time, from simple HTML coding to HTML Indexer and other specialist software. New indexes and software products have been created, but many website indexes have also been lost. This article considers the reasons for creating website indexes and includes examples of website indexes and other access tools which were created in the last 15 years, but which are no longer available. It also provides an assessment of some of the reasons for the changes.

Wandel beim Indexieren von Websites
Website-Indexing wurde in den 1990er Jahren wichtig, als Indexierer, Bibliothekare und Web-Manager mit verschiedenen Ansätzen experimentierten, einen besseren Zugang zu Informationen zu schaffen, die sie über das Internet anboten. Die Tools, mit denen Register erstellt werden, reichen von simpler HTML-Codierung bis zu HTML Indexer und anderer spezieller Software. Neue Indexe und Software-Produkte entstanden, aber viele Website-Register sind auch wieder verschwunden. Es werden die Gründe für die Erstellung von Website-Indexen dargestellt und Beispiele von Website-Indexen und anderen Zugangsoptionen erläutert, die in den letzten 15 Jahren entstanden, die aber nicht mehr zur Verfügung stehen. Es werden einige Vermutungen über die Gründe für diese Veränderungen angestellt.

Website indexing concepts developed as indexers, librarians and web managers experimented with different approaches for making the information they were providing on the Internet more accessible. These approaches included: A to Z indexes; displaying the overall structure of the site (information architecture); site maps; and search

facilities. Search facilities were sometimes enhanced by the creation of subject metadata ("catalogue cards"), which could be organised in different facets, or displayed visually as well as textually.

The tools for creating A to Z indexes have changed over time. Initially, indexers used simple HTML coding to create indexes. Features such as indents and turnaround (wraparound) lines caused difficulties. The development of HTML Indexer was a major breakthrough as it provided a simple, effective way to create indexes.

Alongside the creation of new indexes and new software products, many website indexes have been lost. The reasons for this include changes to the whole company (eg, takeovers) and changes in search policy (eg, provision of a search engine instead of an index). The need for constant maintenance of website indexes when content changes has been a major challenge.

This article considers the reasons for creating website indexes, but does not explain in detail how this should be done – this is covered in our book *Website indexing* (Browne and Jermy, 2004) and in the article by Heather Hedden on page 433 in this issue. The article includes examples of website indexes and other access tools which were cited in the first or second editions of the book (www.webindexing.biz/Webbook2Ed/linkspage.htm), but which are no longer on the web at the same URL, and provides an assessment of some of the reasons for the changes, and commentary on the trends we have observed over the last decade.

Why create an index?

Search engines such as Google provide quick access to much useful information, leading some people to wonder whether any other pathways are needed. People who create website indexes are aware that there are different types of information requests, and different types of information users, and that people are best served by having a range of options available to them. See articles by Brown, Leise (2002) and Bates (1989, 1998) on user needs.

Comparing A to Z indexes with search engine results

An interesting experiment involves comparing a web-based index with the results that a Google search on the same material provides. You can do this by using the 'site:' search facility in Google, through which you can limit a Google search to a single site or part of a site. For example, typing 'site:library.lanl.gov/libinfo/news GeoRef' into the Google search box will find all occurrences of the name 'GeoRef' within the site library.lanl.gov/libinfo/news. This site contains the Los Alamos National Laboratory Research Library Newsletter, which has an index, so we can compare results of an index search with results of a Google search.

If you look up 'GeoRef' in the A to Z index (library.lanl.gov/libinfo/news/newsindx.htm), you find a crafted index heading with six subheadings, clearly showing the different information covered in each section, and the date of publication:

GeoRef
 added to FlashPoint (7/04)
 available through CSA interface (9/02)
 CD version available (6/95)
 funding opportunities section (9/98)
 web version available (1/98)
 web version trial access (7/97)

When you do a Google search, on the other hand, you find a large number of hits, each with a title and brief summary. It is more time consuming to look through these results, but they do provide additional hits that may be of interest. Thus the two information access methods provide different results, each of which might be useful to a specific user at a specific time.

You can do another comparison by searching for 'RDF' (or any other topic of interest) in the W3C site index at www.w3.org/Consortium/siteindex#R, and then doing a Google search of 'site:www.w3.org RDF' (thousands of hits to browse through!).

Usability

Website indexes fulfil many of the usability heuristics (general design principles) mentioned by Jakob Nielsen (2005). For example, they offer:

- Recognition rather than recall (they are browsable, and don't require the user to consider spelling and word formatting such as hyphenation)
- Match between system and the real world (they are a familiar tool for book users)
- User control and freedom (they are easy, quick and direct, and users can move through the index as they wish)
- Help users recognise, diagnose and recover from errors (indexes provide cross-references to guide users from their entry terms to other terms, and glosses (parenthetical qualifiers) to explain the meaning or use of a term)

Provide help and documentation (indexes often provide introductory notes which contain information about the index and its role within the site).

Changes in website indexing

There have been many changes to website indexing in general, and to individual website indexes, over the last decade. This section examines some of the indexes that changed since the publication of the second edition of Website indexing (Browne and Jermeay 2004).

Index removal due to company changes

Website indexes suffer when companies merge or cease certain activities. The Case-in-Point index (previously at www.acxiom.com/caseinpoint/cip-ix-home.htm), which was discussed in Beyond book indexing (1999), and was a place-winner in the AusSI (now ANZSI) Web Indexing Award, no longer exists. This is because the website, the journal, and the index have all been closed down.

The PeopleSoft Products (www.peoplesoft.com/corp/en/indices/prod_index.jsp) index appears to have been lost when PeopleSoft merged with Oracle on June 1, 2005. All PeopleSoft.com content is now on Oracle.com.

Index removal due to policy changes

The website index to the University of Texas Policies and Procedures Manual was a trailblazer (Fetters 1998). By the time we wrote the second edition of our book, however, it had been replaced with a search engine enhanced by metadata.

The Murdoch University Handbook for 2005 had an online index (www.comm.murdoch.edu.au/handbook/search/handbook_index.html), but I can no longer find an index on the site. Similarly, the A

to Z indexes in French and English at the Statistics Canada site seem to have been replaced with a site map, and simple and advanced search.

The Yale Undergraduate Regulations Index used to have an index created using HTML/Prep. When you hovered over an entry it showed 'tips' that provided your context within the index (see an archived copy at <http://web.archive.org/web/20041205153126/http://www.yale.edu/ycpo/undregs/pages/in dexpage.html>). There appears to be no index to the 2006–2007 Undergraduate regulations at <http://www.yale.edu/yalecollege/publications/uregs/index.html>.

Index changes due to website changes

The index to the ANZSI website (www.aussi.org, soon to be changed to www.anzsi.org) has not been updated during the website redesign process, although it may be re-introduced at a later date. Changes to the website creation software including use of a database rather than individual page creation make it more difficult (but not impossible) to use HTML Indexer to create the index.

The British Society of Indexers (SI) redesigned their website. The index is

Neue Erlösmodelle für Zeitungsverlage.

Sandra Huber. – Boizenburg: vvh, 2007. 173 Seiten. (Medienwirtschaft). 27,90 Euro, ISBN 978-3-9802643-9-6.



Wie schon der Titel des Buches suggeriert, geht es hier um das Eingemachte. Wie können wir mit Medien Geld verdienen. Bei der allgemeinen Medien-Euphorie scheint dies

eher trivial zu klingen. Denken wir aber an den harten Wettbewerb, so die vielen Flops, dann kommt uns Sandra Hubers „Neue Erlösmodelle für Zeitungsverlage“ wie gerufen. In diesem Buch steht das Medium Zeitung zwar im Mittelpunkt, aber auch Bibliotheken, Hosts und andere Informationseinrichtungen werden von diesem Buch profitieren, das eigentlich dem Thema Diversifizierungsstrategie einen wichtigen Platz einräumt. Vor die-

sem Hintergrund lassen sich Hubers Vorschläge – mit den entsprechenden Anpassungen – ebenso gut auch für den BID-Bereich anwenden. Was ist nun das Besondere an diesem Buch?

Ausgangspunkt ist die These, dass die Zeitungswirtschaft einen tiefgreifenden Wandel durchläuft. Die Gründe dafür sind offensichtlich: Verändertes Leseverhalten und multimediale Alternativen. Dadurch bedingt, gerät das klassische Anzeigen- und Vertriebsgeschäft ins Wanken. Sandra Hubers These lautet deshalb: „Starke Zeitungsmarken hüten nicht nur die Tradition, sondern beschreiten auch neue Wege“, wie es im Klappentext heißt.

Und hier liegt die Stärke des Buches: Ausgehend von einer Analyse der aktuellen Situation im Zeitungsmarkt (Kapitel 1) und der Beschreibung von „Zeitungen als Wirtschaftsgüter“ (Kapitel 2), beschreibt die Autorin in drei weiteren Kapiteln, wie sich der Zeitungsmarkt den neuen Herausforderungen stellt beziehungsweise stellen sollte.

Das dritte Kapitel stellt „Neue Erlösmodelle für Zeitungen“ vor. In den fünf Unterkapiteln geht es

- um „Erlösmodelle für Zeitungen im Internet“,
- um die „Kernkompetenz: Content“, wo die Autorin thematisiert: Synergie-

effekte zwischen Print und Internet, Erlösquellen (Online-Archiv, E-Paper, Weblogs), Synergien von Print und Mobile (z.B. iPod), verlagsnahe Zusatzprodukte (z.B. CD-ROM, DVD), verlagsferne Produkte (Markendehnung), Crossmedia und schließlich weitere Erlösquellen wie Zeitung als Veranstalter, Vorteilsclubs, Webauktionen.

Im vierten Kapitel „Analyse der neuen Erlösquellen“ unterzieht die Autorin die im dritten Kapitel vorgestellten Potenziale einer SWOT-Analyse und einer Bewertung.

„Die Zukunft der Zeitung“ bildet den Abschluss dieser hochinteressanten Arbeit. Es folgen noch ein Fazit, Literaturverzeichnis und ein Sachregister.

Sandra Hubers Arbeit zeichnet sich durch eine sehr hohe Praxisnähe aus, durch anschauliche Beispiele und einen sehr lesefreundlichen Stil. Dieses ausgezeichnete Buch aus dem neu gegründeten Fachverlag für Medientechnik und -wirtschaft des Ingenieurs für Druckereitechnik und Informationswissenschaftlers Werner Hülsbusch (vvh) sei allen im BID-Bereich Tätigen sehr empfohlen.

Wolfgang Ratzek

now available directly from the home page through provision of an alpha bar (links for all letters of the alphabet to the appropriate part of the index) as well as on the index page at www.indexers.org.uk/index.php?id=217. With a new indexer, a new software package (XRefHT32) has been used.

Long-lived indexes

There are also a number of long-lived indexes on the web. One of the keys to longevity seems to be to have a website indexer who is closely connected to the management of the website; another is to have a site or subsite that changes little or not at all.

The index to the Los Alamos National Laboratory Research Library Newsletter was an early example of the use of HTML Indexer (lib-www.lanl.gov/libinfo/news/newsindx.htm). It was mentioned in the first edition of our book in 2001 and was still going strong in 2005. The index is still online, but the monthly newsletter has been replaced with a blog, so the index is complete and is not being updated.

The index to the book *Orders of Magnitude* (www.informationuniverse.com/ordersmag/orders.htm) has been online for over a decade. As a book index doesn't need maintenance, it is relatively easy to keep such an index available.

Penrith City has for many years provided indexes to its council and library services. These have recently changed URLs, but are still available online:

- A-Z of Council Services at www.penrithcity.nsw.gov.au/index.asp?id=960
- A-Z of Library Services at www.penrithcity.nsw.gov.au/index.asp?id=563
- Quick Index, which covers the whole site, at www.penrithcity.nsw.gov.au/index.asp?id=634.

Montague Institute Review dynamically creates an A to Z index from database-stored metadata at www.montaguelab.com/Public/indexes.htm (go to <http://www.montague.com> then select 'Index'). They say:

'...Even on a relatively modest Web site, standard keyword searching can return too many matches to be meaningful. Furthermore, visitors need cross references and definitions to understand specialized terms, such as 'upstream knowledge management' or 'authority files'. Finally, maintaining indexes and tables of contents for a monthly publication is too time-consuming and error-prone without a standard list of terms in database format.'

Alternatives to website indexing

There have been interesting experiments with access methods apart from A to Z indexes.

Visualisation

Most of the information visualisation examples that were mentioned in Website indexing no longer appear on the web. For example, Inxight now uses text rather than a star tree to display their site map (www.inxight.com/map) and the Antarctica examples are no longer there, although the company itself is (www.antarctica.net). The PubMed example has gone, but you can read an article about it at www.pubmedcentral.nih.gov/article/render.fcgi?tool=pubmed&pubmedid=12556244. Web-wide visualisation tools such as kartoo.com remain.

Classification

In Website indexing (Browne and Jermy 2004) we wrote that few of the projects listed by OCLC as being classification schemes for web resources (orc.rsch.oclc.org:6109/classification) were still current. That website is no longer available. The topic of the use of classification schemes on the web is addressed by Vizine-Goetz (2002).

The main survivor in this area is BUBL LINK (Bulletin Board for Libraries, www.bubl.ac.uk). It groups selected sites according to the Dewey Decimal Classification. Users start with the ten main classes, then browse step-by-step through the classification. The site also provides an A to Z index and search box on home page. Classification schemes do not seem likely to have wide potential on the web because they are time-consuming to implement, and hard to keep up-to-date.

Faceted classifications on the web

Faceted classifications provide effective online searching because they allow users to select the facets (general topic areas) in which they want to limit their search. For example, on a site about wine you can choose to search first by type of wine, or country of origin, or price, depending on your needs at the time (Browne 2003).

Most of the examples of faceted classification from our book have gone, but most of the major programs remain. Those that have disappeared include:

- Annotated Wordnet (www.siderean.com/wordnet17.jsp)
- CMS Faceted Product Directory (www.cmsreview.com/timelines/ShopFeatureDirectory.html)
- Meta Matters (dcanzorg.ozstaging.com/mb.aspx)
- Online proceedings of the DC- 2002 Dublin Core conference (www.siderean.com/dc2002.jsp).

Those that remain include:

- Epicurious (www.epicurious.com/recipes/find/advanced)
- Tower records (www.towerrecords.com). Select Advanced on the home page next to search box. Then the left hand column also has 'Browse by', at which you can search according to specific facets and their values (eg, Price: 'under \$10').

Faceted search at Langemarks Cafe (www.langemark.com/taxonomy_search/blog) has been replaced by a cloud map at www.langemark.com/tagadelic/chunk/1?PHPSESSID=7078c8337f0bba91181735d0c8a14fca.

A new implementation of faceted classification is seen in the NCSU libraries catalogue. You enter terms in the Catalog search box at www.lib.ncsu.edu/search collection, and can gradually narrow your search, while seeing the numbers of hits retrieved.

Topic maps

Topic maps (Browne 2002a) do not seem to have had a major impact on the web, although they are used in some company intranets. The Diffuse site (www.diffuse.org) provided an important example of the way topic maps could be implemented on the web, but it is no longer available on the web.

Subject gateways

There are a number of subject gateways on the web, which lead to quality, selected information resources on a specific topic or, occasionally, on all topics. These depend on human activity and can only deal with a small proportion of the available information.

In October 2003 Infomine stopped taking suggestions for sites to include as they had been plagued by bulk inappropriate commercial submissions. Site submission is now active again at infomine.ucr.edu/contact/suggest.shtml.

Zeal used to take suggestions of non-commercial sites at no charge for inclusion in their directory (www.zeal.com/users/non_profit.jhtml?rpc=49). Its owner LookSmart is now focussing on other objectives and Zeal.com has been shut down. They recommend the use of the online bookmarking service Furl (www.Furl.net).

The AVEL subject gateway (Australasian engineering & IT resources, avel.edu.au/docs.html) operated from 1999–2005. The site has now been decommissioned, and the history of the project can be found at avel.library.uq.edu.au.

New methods of indexing

As well as needing individually-created indexes, the web is being gradually organised through the process of social bookmarking, in which people tag sites of interest to them using keywords, and then share these keywords with other users.

One popular site is Delicious (del.icio.us) – here you can look at your own tags and those of other people as lists or clouds (with topics that have more content presented in a larger font). You can follow links by topic and tag creator, and can add people to your network, thus easily sharing website discoveries. The lists of tags are somewhat like an index, but lack features such as cross-references and subheadings, and often lack consistency. For example, at del.icio.us/url/e6f93dcf17dfd5f1a25e0b8770203ff2 you can see the different ways in which people have tagged the Website Indexing SIG site, depending on their points of view.

A similar approach is taken at sites such as Citeulike (www.citeulike.org/tag/indexing) and Technorati (www.technorati.com/tag).

Software changes

A number of software packages have been written to aid in the construction of website indexes. These are all tools for human construction of indexes, and do not aim to create fully automatic indexes, although some create a default index that can be used as a starting point. The significant change in this area has been the creation of two new programs.

For information on HTML/Prep, which can be used to transform print indexes for use on the web, as well as to create website indexes from the start, see Browne (2002b) and www.levtechinc.com. For information on HTML Indexer, which embeds metadata in webpages to make easily updatable website indexes, see www.html-indexer.com, Browne (1999) and Unwalla (2006). HTML Indexer has a free demo version which makes functional indexes, except that the projects can't be saved for later editing (although the indexes themselves can be edited).

XRefHT32 ('shref') is a free, open source program that can be used to create website indexes. It can be incorporated with a thesaurus created using TheW32, a free thesaurus creation program by Timothy Craven (publish.uwo.ca/~craven/freeware.htm). You can see examples on the web at www.ulb.ac.be/ecoled/eiulb/liens/liensM.htm and publish.uwo.ca/~craven/craven.htm. For more information see Hedden (2005) and Lamb (2006).

Basedex is the latest program, and has just been mentioned on the Webindexing mailing list in March 2007. It can be used to create large-scale indexes through the merging of separate indexes. See the discussion at the Web Indexing Yahoo mailing list – tech.groups.yahoo.com/group/web-indexing – and the site at blazingrails.basedex.com.

Conclusion

Website indexing has been constantly developing since the early days of the web. There have been a range of approaches, supported by software products to make the job easier, although still requiring human input for the best quality results. Although search engines provide one good approach, website indexes can provide complementary approaches to access to important information.

For more information see the ASI Web Indexing SIG (www.webindexing.org/web-index-examples.htm) and 'Indexing resources on the WWW' by Stephenson (2005).

References

- Bates, Marcia J. 1989: The design of browsing and berrypicking techniques for the online search interface. www.gseis.ucla.edu/faculty/bates/berrypicking.html.
- Bates, Marcia J. 1998: Indexing and access for digital libraries and the Internet: human, database and domain factors. In: *Journal of the American Society for Information Science* 49(1998), pp. 1185–1205, www.gseis.ucla.edu/faculty/bates/articles/indexdlib.html.
- Brown, David M.: Why create an index. www.web-indexing.org/article-brown.htm.
- Browne, Glenda 1999: Web site indexing. www.webindexing.biz/Articles/WebSiteIndexing.htm. First published in *Online Currents* 14(1999)10.
- Browne, Glenda 2002a, Topic maps. www.webindexing.biz/articles/TopicMaps.htm. First published in *Online Currents* 17(2002)3.
- Browne, Glenda 2002b, HTML/Prep: transforming indexes for the web, www.webindexing.biz/articles/HTMLPrep.htm. First published in *Online Currents* 17(2002)7.
- Browne, Glenda 2003: Faceted classification. www.webindexing.biz/articles/FacetedClassification.htm. First published in *Online Currents* 18(2003)9.
- Browne, Glenda; Jermy, Jonathan 2004. *Website indexing: enhancing access to information within websites*. 2nd edn. Adelaide: Auslib Press 2004. Now for sale as an e-book through RMIT Informit (www.informit.com.au), as a PDF at www.webindexing.biz and as a PDF or print-on-demand book from Lulu (www.lulu.com).
- Fetters, Linda 1998: A book-style index for the Web: the University of Texas Policies and Procedures Web site. In: *The Indexer*, 21(1998)2.
- Hedden, Heather 2005: HTML indexing freeware: XRefHT32. www.hedden-information.com/XrefHT32_Review_KW2005_10-12.pdf. First published in *Key Words*, 13(2005)4.
- Hedden, Heather 2007: *Indexing specialties: web sites* Medford, NJ, Information Today, for the American Society of Indexers.
- Kamel Boulos, Maged N 2003: The use of interactive graphical maps for browsing medical/

health Internet information resources. In: *International journal of health geographics* 2(2003)1, doi: 10.1186/1476-072X-2-1.

Lamb, James 2006. *Website Indexes: visitors to content in two clicks, or website indexing with XRefHT32 freeware* (available as a print-on-demand book from www.lulu.com).

Leise, Fred 2002. *Improving Usability with a Website Index*. www.web-indexing.org/article-leise.htm. Originally published July 15, 2002 on *Boxes and arrows* (www.boxesandarrows.com).

Nielsen, Jakob 2005: Ten usability heuristics. www.useit.com/papers/heuristic/heuristic_list.html.

Rowland, Marilyn; Brenner, Diane 1999: *Beyond Book Indexing: How to Get Started in Web Indexing, Embedded Indexing, and Other Computer-Based Media*. Medford, NJ, Information Today, for the American Society of Indexers.

Stephenson, Mary Sue 2005: Indexing resources on the WWW. www.slais.ubc.ca/resources/indexing/database1.htm.

Unwalla, Michael 2006: Web indexing: extending the functionality of HTML Indexer. www.techscribe.co.uk/ta/web-indexing.htm. First published in *The Indexer* 25(2006)2.

Vizine-Goetz, Diane 2002: Classification schemes for internet resources revisited. staff.oclc.org/~vizine/JIC/v5n42002/ClassificationSchemesRevisited.doc. Manuscript of paper published in *Journal of Internet Cataloging* 5(2002)4.

Website, Subject Indexing, Trend

Elektronischer Dienst, Website, Sachkatalogisierung, Index, Entwicklungstendenz

THE AUTHOR

Glenda Browne, BSc, MSc, DiplM-Lib



has been an indexer of books, journals, databases, online help and websites since 1988. With her partner, Jon Jermy, she has written two books – *Website indexing* (Auslib Press, 2nd ed, 2004, www.webindexing.biz) and *The Indexing Companion* (Cambridge University Press, 2007). Glenda is the editor of the *Around the World* section of *The Indexer* journal, which involves gathering information from all of the indexing societies. She has taught indexing to indexers and editors, through ANZSI, the NSW Society of Editors, Macquarie University and the University of NSW. She has been involved on ANZSI state and national committees, including roles as newsletter editor, Sydney conference administrator and NSW Branch President (current).

PO Box 307
Blaxland NSW 2774, Australia
webindexing@optusnet.com.au
www.webindexing.biz