

Memo #475981 (40803) Received: from godzilla.zeta.org.au (godzilla.zeta.org.au [203.2.228.19]) by tom.compulink.co.uk (8.6.9/8.6.9) with ESMTTP id JAA26425 for shuter@cix.compulink.co.uk; Message-Id: 1160856.TAA11411@godzilla.zeta.org.au. To: shuter@cix.compulink.co.uk. From: aussie@zeta.org.au (Australian Society of Indexers) Subject:

AusSI Web Indexing Prize

1: a simple index

The prize

I am writing to give you the background to a new Web indexing prize, open to indexers worldwide. The prize will be a year's free membership of AusSI, or a copy of the Proceedings of AusSI's First International Conference in Marysville, 1995.

This article is to help you get up to speed by creating a Web index of a document or family of documents by the same author. This could be developed later to the stage where I would mount it on the AusSI Web site for example.

The prize is meant to allure some of you, our erstwhile indexers, onto the electronic medium. Currently I think 80% or more of the work being done is paper based. There are about 20% doing database indexing which involves computer-based work.

The main tools you will need to construct the final product are:

1. your normal indexing software, e.g. CINDEKX or MACREX;
2. a specially created tool (WEBIX) to convert a standard index into HTML (Hypertext Markup Language)
3. a Web browser such as Netscape or Mosaic to view the Web document

WEBIX produces documents ready for the Web. No HTML knowledge will be required. Feed in the index and out comes the HTML document ready to go. This will be available from the AusSI Web site or on disk by mail from me at a nominal fee.

Later, after creating some successful mini Web indexes, we will venture into editing the HTML ourselves and searching the Web for more interesting information on a chosen topic. At this point a more advanced form of Web index will be introduced. It is known as a bibliographic index because of its complexity and is created using bibliographic searches, similar to a research project.

If you have access to the Internet, the prize is described on the AusSI Web site under 'the art of indexing the Internet'.

What will a web index look like?

A traditional hierarchical index will be made using CINDEKX, MACREX or other indexing software. Instead of page numbers you will have URLs (Universal Resource Locators - WWW jargon for addresses of pages). For our purposes, we will be creating links to pages in the one area so we will not have to put in complex addresses as on the Web at large, e.g.

P	
subject	URL
publications	publicat.htm

This will be run through the WEBIX (index to HTML converter) to create a Web page. In a Web browser like Netscape, choose File | Open File. Punch in the Web page's name and up will come:

P
publications

'Publications' is underlined, which means there is a hypertext link to the publicat.htm page. That means if you click on publications you will be transported to the publicat.htm page—akin to turning to the exact page as with thumb tabs on a large text. You can see why the URL is the same as a page number in a traditional index—it tells where in the family of documents the subject lies.

So you would edit the original index, print it to disk then run it through WEBIX to produce the HTML. You then have an up to date Web index.

Setting up CINDEKX or MACREX to produce the output for WEBIX:

WEBIX expects a simple format:

header URL.htm
subject level 1: URL.htm — indented 2 spaces
subject level 2: URL.htm — indented 4 spaces

Check your defaults to make sure the indents are the right number of spaces and that the index is saved as an ASCII file.

2: a bibliographic index

This is the sequel to the simple Web index above. It is what the main form of Web indexing will be.

The background to bibliographic web indexes

As Webmaster, I was approached by an academic in Melbourne University, Steve Hunt, who has suggested indexers could make money indexing small portions of the Internet. They need to be computer literate and have access to the Internet. As a service for their clientele, they could update the index on a regular basis for a set fee. This is becoming more and more important as even the best search engines often turn up spurious information. Dozens of false hits are generated as the search engines often do not have adjacency operators and return hits say only when two words occur in the same document, let alone side by side or in the right order.

It may seem a pretty arid and untapped area but the returns are high. With a much higher hit rate, people will be more likely to use the Internet. Indexers will become sought after. Currently I am losing my patience with the endless hours of searching finding information is producing. It is a jungle.

As support for this, I have created an area on the AusSI Web site called 'the art of indexing the Internet'. This holds links to sample indexes of sites e.g. *The whole Internet catalog* by O'Reilly & Associates publishers.

How to get on top of HTML

The main tools you will need to construct the final product are as before plus:

HTML Writer (Windows) or a similar HTML editing program to touch up indexes.

HTML is a markup language. It consists mainly of tags. Most of them come in pairs, one to switch on the feature and one to switch it off. Many tools can automate creating these tags.

When creating links in HTML Writer, there are buttons which will prompt you for a URL. You just highlight the subject heading with the mouse, click on the 'create a hypertext link' button then type in the URL when prompted. HTML Writer produces all the correct HTML details—no worries. You do not have to be a HTML whiz to do this. Keep one heading per line and it will be tidy and readable. Using the 'create a hypertext link' button you can also create hypertext links for see references as well.

Creating the bibliography

Well, how can you find material to insert into a Web index? Here are a few sources:

1. there is a source of electronic journals at the National Library of Australia and the University of California San Diego Library (with ARL, Yale, CCAT-UPenn);
2. another source is the gangling list of Usenet newsgroups e.g. comp.infosystems.www.announce;
3. some would be obtained from work—e.g. cataloguers using thesauri on the Internet;
4. finally, from surfing the Internet.

3: building a bibliography

Sifting and searching to construct the index

In this stage, we are building an index of URLs from all round the world. Create a hierarchical bibliography. Pick a subject area to index, e.g. music. See what already exists. Have a look in our 'indexing the Internet' page to sample possible subject areas.

Established virtual libraries as examples

W3 organisation has a virtual library including a whole area on astronomy. This cross links to all the Australian CSIRO astronomy sites out here, although it resides in Switzerland. Instead of duplicating, create links to an existing hierarchy in your index. In most of these large indexes, the first thing you will notice is that the subject headings are very broad—astronomy, education, music, engineering, science.

Using people's home pages as sources of index entries

Being a librarian, I often find material on the Internet which is heavily biased towards academics in a post-graduate vein. This needs to be filtered out. Rarely will you find a narrower term. The only times you will find really in depth data is on the home page of someone doing their thesis. We are moving on untapped territory here. You could create a niche.

When talking about possible classification schemes of the Internet, Mike Middleton from QUT mentioned at the Marysville International Conference that users' home pages were a simple index. People have chosen links to pages relevant to their interests. I have visited some home pages found via Webcrawler and they go on for pages. They can include pages of people's buddies; Microsoft, Intel and the US Federal Government.

The paper chase: building a bibliography of organisations

Using meta-indexes I traced several organisations' Web pages including OCLC, BIOSIS and Defence Information. This formed a mini bibliography of organisations relevant to

indexing and abstracting organisations. A meta-index is an index of indexes, a master list. See the NFAIS link below. Finding these used the paper chase mentality as when researching a paper or finding a book for a client in a public library. Using bibliographies from seminal papers you can fabricate an entire area within a few hours.

The search engine I used for this was not Webcrawler—too much sporadic information. Instead I used the Telstra 'server' search engine. This only carries site names, e.g. Australian Bureau of Statistics, not sites with say just 'Bureau' in their name like Webcrawler and other WAIS search engines devolve into—not too good!

Conclusion

So from these searches and browsing sessions, you will find some valuable leads. You start to create a bibliography of URLs which are relevant to your topic. It is an information base. As a tip, in order to keep track of the sites you want to remember, use Netscape's rudimentary bookmark facility.

It may be even worth your while getting a piece of paper to start mapping the hypertext world of the Web. Later when you start creating these kind of bibliographies for clients, you can tailor and improve them with feedback from your users. Specialised indexes can be created by dropping or demoting useless links. Added cross-references can be made using your new skills with HTML Writer.

You have started to produce a growing index which can be added to create a pathway through the Internet's information soup. There has definitely been an information explosion this century.

So, decide on an interesting topic to hunt, start using your existing indexing software to build an index, WEBIX to create a Web page from your index, an HTML editor to touch it up, a relatively fast low cost link to the Internet (or switch the graphics off on your browser) to hone your Web surfing, and, voila! a Web index emerges. Best of luck!

Surf's up!

Dwight Walker, Webmaster, Australian Society of Indexers, URL: <http://www.zeta.org.au/~aussi>

Web Bibliography:

Indexing the Internet by Michael Middleton, AusSI 1st International Conference, Mar-Apr 1995:
<ftp://ftp.fit.qut.edu.au/InfoSys/papers/asindex.wps>

AusSI Internet Indexing Page:
<http://www.zeta.org.au/~aussi/inetindx.htm>

NFAIS indexing organisations:
<http://www.zeta.org.au/~aussi/nfaisinx.htm>

Webcrawler search engine:
<http://webcrawler.com>

Telstra's Meta-Index of search engines:
<http://www.telstra.com.au/metaindex.html>

Australian Electronic Journals. National Library of Australia:
<http://www.nla.gov.au/oz/ausejour.html>

NewJour electronic journal mailing list archive (ARL, Yale, CCAT-UPenn, UCSD):
<http://gort.ucsd.edu:80/newjour>