

Web indexing prize 1997

Dwight Walker

Briefly describes entries to the 1997 AusSI web indexing prize, with a discussion of trends in web and online indexing.

In January this year, Alan Wilson and I judged the 1997 Web indexing prize. It was even more challenging than last year's competition! We had many more technically oriented entries plus many high quality manually generated ones. Many were so entrancing they took hours to assess.

The winner for 1997 is Christabel Wescombe from the University of Sydney. She produced an amazing annotated Web index based around the curriculum of the Faculty of Education. Her in-depth annotations gave many insights to readers about the value of the information on Web sites. She had a graphic designer and Webmaster to assist her. She was awarded the prize at the AGM of the Australian Society of Indexers in March 1998.

Ann Treacy's *Minnesota Web* came second. This was an index of links from all over Minnesota, USA. It was database driven and had broad categories.

Web indexing comes in all manner of forms such as Web databases, annotated Web indexes, online book indexes and Web site indexes. They are usually either very large — in the order of thousands of entries — or contain hand-picked material that gets to the heart of a subject area. On some of the bigger public sites, the indexing teams were using databases to sort and store the growing numbers of links and categories that were being demanded. On the other hand, some indexing teams would manually maintain a limited number of quality links.

The sizes of the teams varied from single people working from home to multi-person teams in universities. The Internet has opened up self-publishing to all comers.

Another contrast was in the content of the indexes. Some were purely jumping-off points to a particular subject. Others were pointers to information stored offline. Some had extensive annotations to links outside their sites while others were detailed links into a particular site or online book. The diversity was remarkable.

The range of subjects again was very interesting — from crime records to hacker research to fan fiction to genealogy to education to recreation networks for disabled people to medical research. The indexes were giving out direct information either through a link or actual information stored in a Web database for easy access. Some indexed documents were only found on the Web such as the fan fiction from an Internet newsgroup.

Let's look at some specific examples. Graham Greenleaf's AustLII and Project DIAL used spiders pointed to law sites to garner the best links for their institute's Asian law database. These search results were then organized into a hierarchical index. Others, like the FBI files, can be used to present actual data to readers to help them find a particular offline document. An online book such as the *Western Australian Aboriginal Language Guide* lets non-literate tribesmen look up their

language group by clicking on a map.

So Web indexing continues to mature and evolve at a great rate. The accessibility of the Internet has accelerated more and more ways of indexing its content. Following are the entrants. See the AusSI Web site to browse them: <http://www.zeta.org.au/~aussi/>. Enjoy!

Indexes judged first, second and third are listed in that order, followed by remaining entrants in order of receipt. Indexes are all described in entrants' own words.

Faculty of Education Internet Guide — FIRST

Christabel Wescombe (c.wescombe@library.usyd.edu.au)

Fisher Library, University of Sydney

<http://www.library.usyd.edu.au/Guides/Education>

type: annotated index

role: as Faculty of Education librarian, chose the sites and arranged them to support the teaching and research profile of the Faculty of Education at the University of Sydney.

audience: education administrators, policy makers and researchers from University of Sydney, but also has relevance to people in the private and public sector related to education.

how created: graphic and technical assistance from Steve Ryan (Fisher) and John Harvey (Faculty of Education)

hardware/ software: editing: Mac, BBEdit, Photoshop; Web server: Silicon Graphics workstation running a version of UNIX (AIX), Web server Netscape Commerce server.

comments: Kerry Webb mentioned it in the Dec 97 issue of *inCite*; it won the Argus Clearinghouse Digital Librarian award (U Mich) for Oct 97; the seal of approval and maximum score.

Index for Minnesota web sites — SECOND

Ann Treacy (atreacy@mr.net), MRNet, Minneapolis, USA

<http://www.mnonline.org/uffda>

type: hierarchical list

role: I developed the organization scheme and continue to work registering new sites, promoting the site, and providing customer support.

purpose: We wanted to develop an index for Minnesota web sites. We index only Minnesota sites, but do not confine ourselves to the metropolitan areas. Our goal was to create an index that was accessible and useful to users with a wide range of competencies.

techniques: We developed a category scheme to suit the types of Minnesota web sites. We also developed a database so that users could search via directory and keyword. The liaison between the database and the user interface is Perl scripting to accommodate most computers.

software: Informix, hosted on an Enterprise server

AustLII's World Law Index (including Project DIAL) — THIRD

Graham Greenleaf (graham@austlii.edu.au), Geoffrey King, Daniel Austin

AustLII, UTS and UNSW Faculties of Law, Sydney

<http://www.austlii.edu.au/links/World/>

type: hierarchical list with spider searches

roles: G. Greenleaf: intellectual index, structure and entries; targeting of intellectual index and search engine web spider; embedded search links to integrate. G. King: Developed 'feathers' intellectual indexing software; 'gromit' web spider, and 'shaun' search interface. D. Austin: further development of 'gromit' and 'wallace' spider harness

purpose: To maximise the advantages of both intellectual and automated web indexing by integrating them. To create a maintainable world-wide law index (with a statutory emphasis through Project DIAL)

audience: For the World Index, any legal researchers needing a genuinely world-wide (i.e. not US-centric) starting point. For the DIAL component, legislative personnel in developing countries.

techniques: For full details see 'Future proofing a global Internet index with a targeted web spider and embedded searches', AusSI Conference 1997

<http://www2.austlii.edu.au/~graham/Futureproof/indexers.html>.

Since the Conference, the web spider now indexes 40,000 documents, about 250mb of legislative texts; there are embedded searches for most countries; searches find other legal indexes for those countries. It is still a prototype, but we believe it is becoming very useful, maintainable, and genuinely global.

software: All software created for project: 'feathers', 'gromit', 'wallace', 'shaun' interface

Asian Resources Directory

Susan MacDougall (smacdoug@alianet.alia.org.au)

ALIA, Canberra, <http://www.alia.org.au/~apsig/directory.html>

type: hierarchical list

This is the URL to my Asian Resources Directory set of web pages: <http://www.alia.org.au/~apsig/directory.html>. It has a main index, then indexes for geographic area, subject, database name, virtual libraries, online education/ training courses, organisations and people, library catalogues.

Index of women ...

Cathy Dunn (cathyd@shoalhaven.net.au), Milton, NSW

<http://www.shoalhaven.net.au/~cathyd/raladies.html>

type: genealogical index

role: Index of Women — convicts, settlers' wives, convicts' wives and infants who arrived in Australia on the Royal Admiral in 1792. Personal research. Index part of private publication.

purpose: to promote the Internet as a tool for family history research.

Primary records are made available to researchers without their having to travel great distances.

techniques: Index from convict indents, log entries and other primary records. This site is attached to the *History Home of Milton Ulladulla* at <http://www.shoalhaven.net.au/~cathyd/>

software: Claris Home Page and Netscape Gold

Australian Churches

Kevin Keep (kevin_keep@vision.net.au), Tasmania

http://www.vision.net.au/~kevin_keep/churches.htm

type: hierarchical list

Apparently this page is one of the best listings of Australian Churches world wide.

Australian Defence Force Academy Page

Ros Boyce (r-boyce@adfa.oz.au), ADFA, Canberra

<http://www.adfa.oz.au>

type: Web site index

role: Chairperson of the committee to design and implement the Academy Page. Have done most of the web authoring personally. Responsible for organising, designing and maintaining the Library site along with a small group of staff responsible for some individual pages.

purpose: To provide information about the joint military and academic roles of the Academy. To provide access to information about postgraduate, undergraduate and distance education programmes, to recruitment and enrolment details, to the centres, schools, military, student pages and the library. The library site has been designed to provide a subject oriented client service to allow the users access to services twenty-four hours a day, seven days a week.

techniques: All web authoring has been done manually without any help from a web authoring editor.

software: Netscape communicator, Notepad, Photoshop
Hardware: PC

Transformers fan written fiction (fanfic) titles

Charlotte Brogen (cbrogden@s054.aone.net.au), Newport, NSW

<http://www.home.aone.net.au/charl/charl.htm>

type: newsgroup index

role: Suzanne Ferree has been listing Transformers fan-written fiction (fanfic) titles that have appeared on the newsgroup alt.toys.transformers for years now. With her permission, I've turned her list into HTML, made an author list and added many more titles and authors.

purpose: Transformer fanfic is often surprisingly good but it is scattered over many pages. My purpose was to create a central index of links to Transformers fanfic. The audience for Transformers fanfic is people in their 20s (they were children in the 80s when TFs were at their height) and is surprisingly large. I sometimes get more than 500 hits a day.

how created: Originally from Suzanne Ferree's list. Then I indexed the fanfics by author, and went around checking WWW pages and adding the fanfics that weren't already listed. The indexing is very much an ongoing process. Unlisted authors have seen my page and sent me their details to index and new fanfic is being written all the time. I send Suz a list of what I've found each month so she can index it too.

software: Originally Netscape Gold freeware and a 486. Now FrontPage and a 686. FrontPage has a tendency to muck up my GIFs — I prefer Netscape Gold!

Health & Medicine bookmarks

Saroj Bhatia (sxb553@anu.edu.au), Canberra Hospital Library

<http://pride.anu.edu.au/library/bookmark.htm>

type: bookmark list

role: I have used the Internet for the last 6 years and collected 4,000 bookmarks related to health and medicine. Over the years as Reference Librarian, I have evaluated and updated the bookmarks according to content, currency, and ease of use. I indexed web sites according to medical subjects and subheadings under discussion lists, journals and resources.

purpose: This index is on disk and on my home page <http://pride.anu.edu.au/library>. The disk has been sold all over the world on a cost recovery basis. I have used the disk for training purposes and people who bought the disk can import the file into their bookmarks and customise it as their training tool. The audience for this index is health professionals, medical specialists, medical students, medical libraries and health organizations all over the world. We have sold 1500 copies all over the world. A company which deals with library automated systems is using it for its cataloguing module and University of Sydney Graduate Medical Program is using it for GMP home page links.

techniques: Medical Subject Headings (MeSH) and evaluated bookmarks and Netscape Software.

Hark/ internet-help

David J. Hark (dhark@fred.net), Shepherdstown, WV, USA
<http://www.fred.net/dhark/theone.html>

type: search engine list

purpose: To help students in Internet/WEB classes approach finding information with some order.

techniques: By teaching, learning, searching, networking, listening, collecting and sharing bookmarks.

software: notepad, HomeSite Compaq Presario 9240

Information security index

Paul Ashley (ashley@fit.qut.edu.au), Information Security Research Centre, School of Data Communications, Queensland University of Technology, Brisbane

<http://www.fit.qut.edu.au/~ashley/crisis/index.html>

type: bibliographical database

role: Original concept, database design, some programming and data entry. Students performed some programming and data entry

purpose: To allow researchers new to an area of information security, for example postgraduate students or researchers, to get started. The database is used internally by students of the ISRC, and also receives numerous hits from the web.

techniques: The index is a Mini-SQL database with HTML/CGI Interface. The database is relational and code has been written to allow easy data entry and data retrieval.

software: Mini-SQL from Hughes Technologies, HTML/CGI code, Linux on PC for development, Solaris on Sparc for live version

comments: The index is for searching for information relating to computer security. We have found the index so far has been very popular with students and external researchers to QUT. We hope to have over 2000 records in the database by February 1998.

Recreation and leisure index

Grant Malcolm & Andrew Geach (gmalcolm@earthling.net)
Recreation Network Inc., WEMBLEY, WA

<http://kandinsky.fste.ac.cowan.edu.au/ceedsite/>

type: resource database

roles: Grant Malcolm, Project Manager; Andrew Geach, site design and programming

purpose: To improve the access of the community to comprehensive information on recreation and leisure. A free index to 8,000+ records of recreation & leisure information for the community of metropolitan Perth, especially people with a disability. To encourage direct community involvement in information provision, by allowing community input.

techniques: The original database was acquired and maintained by Recreation Network Inc. through 11 years of facilitating the access

of people with a disability to community recreation and leisure. In recognition of the immense value of the resource to the Perth community, avenues for making the resource available via the Internet were investigated. Enquiry forms and procedures are based on years of expertise in community information provision. Web pages, forms, CGI scripts and a copy of the database are hosted on a server at Edith Cowan University.

software: Filemaker Pro 3.0, BBEdit, Lasso Lite

Orders of Magnitude: a history of the US space program

Linda Kenny Sloan (indexer@ix.netcom.com)

Information Universe, Carson, CA, USA

<http://pw1.netcom.com/~indexer/infoundiv.htm>

type: online book index

role: The index is entirely my own effort based on an online text downloaded from the NASA online historical publications site.

purpose: To provide easier access to the online document. Previously there was chapter access only which entailed a lot of scrolling to find information.

how created: I created a new index for the book using Cindex. I then used HTML/Prep and HotDog Pro for developing and converting the index to HTML.

software: 486 computer with Cindex, HTML/Prep, Hot Dog Pro, & Netscape Navigator (upgraded to Communicator)

Australian studies network

Frank B Poyas (frank@austudies.org), Washington, DC, USA

<http://austudies.org> (also at <http://sunsite.anu.edu.au/austudies> and <http://www.lamp.ac.uk/ASN>)

type: Virtual Library

role: Created and maintain entire network

purpose: To provide a central access to Australian material for academics located off-shore. To develop and serve web sites for academic organisations dealing with Australian studies. Includes the official WWW Virtual Library for Australia.

techniques: Basic manual HTML! A consolidation of numerous existing web sites I had maintained separately.

Western Australian Aboriginal languages

Nicholas Thieberger (nickt@cheops.anu.edu.au)

Vanuatu Cultural Centre in Port Vila

<http://coombs.anu.edu.au/WWWVLPages/AborigPages/LA/NG/WA/contents.htm>

type: online book index

purpose: It is an annotated bibliography with dense cross-referencing between each chapter and the bibliography of Western Australian Aboriginal languages. Maps at the beginning of each section are hot linked to the appropriate language name. A document describing the construction of the site is also located at the same address. I am currently working at the Vanuatu Cultural Centre in Port Vila where Web access is still too expensive for everyday use.

North Queensland index

Jean Dartnall (Jean.Dartnall@jcu.edu.au)

James Cook University, Townsville, Qld

<http://www.jcu.edu.au/gen/Library/Tropicat/tropicat.html>

type: online catalogue annotations

role: I have done about 70% of the indexing for this and managed the project throughout.

purpose: to provide enhanced catalogue access to materials held

in the closed collections of this Library. It is of particular interest to those working on NQ topics. The main audience is staff and students of this University and researchers elsewhere with an interest in North Queensland topics.

techniques: All monographs, including theses, held in the JCU Library Special Collections were scanned and given free text subject headings in addition to standard LCSH headings. The aim has been to emphasise NQ headings such as place names, family names and species names.

software: The index is an additional field in the standard Dynix catalogue structure.

QldWeb

Margie Barram (m.barram@slq.qld.gov.au)
State Library of Queensland, South Brisbane, Qld
<http://www.slq.qld.gov.au/qldweb/welcome.htm>

type: hierarchical index

role: Creator of the index, but now one among several others who work on QldWeb.

purpose: QldWeb is an index of World Wide Web (WWW) resources that are about Queensland, or which are Australian resources of interest to people in Queensland. QldWeb includes sites where a significant portion of the site is about Queensland or is on a subject of social, political, cultural, historical, scientific or economic significance and relevance to Queensland

techniques: Collecting URLs of relevant sites, by searching ISP servers' member lists, search engines, subject directories, and those emailed to QldWeb by their authors. Web sites are arranged topically and by subcategories within a topic. The entire index is searchable using keywords in a search engine. A check for broken links is done monthly and correct URLs found, or broken ones deleted.

software: Mostly done manually! (looking for tools to automate parts of it). Use Netscape for browsing, Sun server running freeware to host information, freeware link checker.

Online policy and procedure manuals

Linda Fetters (lfetters@caller.infi.net) Fetters InfoManagement Co., Corpus Christi, TX, USA

Online policy and procedure manuals, University of Texas at Austin, <http://www.utexas.edu/policies/hoppm/>

type: online book index

role: I indexed the two policy and procedure manuals involved. This project was a carryover from an online version on an IBM mainframe at the University of Texas at Austin. The original online index included five manuals, but the administration decided to put only two at this particular Web site. In addition to formatting the index, I also formatted all the text files and created all the links between the table of contents and index to the text files, and the links between the text files themselves.

purpose: to find relevant policies for student behavior and student grievances. Employees of the University of Texas can also find complete employee regulations and benefits.

techniques: The index was created using Cindex. Most of the HTML codes were inserted using Cindex code tables and function keys.

software: I used HTML Assistant Pro 97 software for formatting the text files and the index and for inserting links.

FBI archival list index

Michael J Ravnitzky (mikerav@ix.netcom.com) / Paul Skolnick (skolnick@crunch.com)

Independent/ Thunder & Lightning News Service
Belleair Bluffs, Florida, USA

archival list of what the FBI calls "exceptional" cases
<http://www.crunch.com/01secret/01secret.htm>

type: criminal record database

role: Michael J Ravnitzky obtained a handwritten archival list from the US National Archives and Records Administration of what the FBI calls "exceptional" cases — thousands of names and case numbers of investigations over 7 decades involving notable Americans and others: actors, activists, union organizers, politicians, athletes, academics. Paul Skolnick posted the list on his Web site, <http://www.crunch.com>, devoted to investigative reporting.

purpose: The purpose of the index is to make available to scholars, journalists, and other serious researchers information (case names and numbers) to help them obtain FBI documents under the Freedom of Information Act (FOIA) and other US disclosure laws.

techniques: Michael J. Ravnitzky keyed the individual case names and numbers (literally thousands of them) from handwritten documents he obtained. Paul Skolnick designed the index, created tables, and sorted the data into easy-to-read and easy-to-search indexes.

software: MS Word; MS Internet Explorer, Corel Draw 7.0, Netscape Communicator 4.00-4.04

ChemoReception Web

Frederic A. Spangler (fred@csa.com)
Cambridge Scientific Abstracts, Bethesda, Maryland, USA
<http://www.csa.com/crw/prehome.html>

type: hierarchical list

role: I conceptualized the ChemoReception Web and then built it from scratch including the generation of all the images.

purpose: The ChemoReception Web covers the sensory and biological aspects of taste, smell, pheromones, perfumes, soaps, cosmetics, essential oils, flavorings, and spices. It explores how tastes, smells, and pheromones are received, perceived, and then translated into responses and actions.

techniques: The index is based on the structure of an abstracts database that I manage. This abstracts database is for sale. I simply wanted to bring interesting portions of it to the scientific community and to the general public for free. Additionally, I am augmenting the abstract citations with key resources from the Internet. The site has been widely accepted. I constantly receive queries about smell and taste from around the world which I do my best to answer immediately. The site has always received very high rankings from the major Web search engines.

software: I did not use any special Web editing tools; rather I did it all the old way — manually, using the vi editor. For the images and animation, I used a combination of Gifcon, PowerPoint, PaintShop Pro, and the online service, PIXELSITE located at: <http://beta.pixelsight.com/PS/pixelsite/pixelsite.html>

Dwight Walker, former AusSI webmaster, runs WWWalker Web Development: <http://www.wwwalker.com.au>