Doing It the Aussie Way : Comparison of Anzwers and Web Wombat Search Engines

Sam Berner
May 23, 2001

1. So the Aussies are gonna a'searchin?

There is a number of search engines and search directories which concentrate on Australian resources. Some of them provide both a search engine and a directory listing. They are usually faster to search than the big ones - ALTA VISTA, INFOSEEK etc. - because of the restricted coverage. I have provided a list, by no means comprehensive, below:

Anzwers (http://www.anzwers.com.au/)
Aussie Index (http://www.aussie.com.au/index/ai.htm)
Australia Announce Archive: Matilda (http://www.aaa.com.au/matilda/)
Cowley Online (http://www.cowleys.com.au)
EdNA (http://www.edna.edu.au)
Kapow (http://kapow.bam.com.au)
Web Wombat http://www2.webwombat.com.au/
http://www.sofcom.com.au/ (http://www.sofcom.com.au)
Looksmart Australia (http://www.looksmart.com.au/)

As search tools become increasingly indispensable to searchers, both professional and lay, users will need to be very discerning as to which search engine to select for best results. It is, therefore, appropriate to examine, compare and evaluate these tools.

In this Opinion Paper, I propose to evaluate and compare two of the Australian search engines, namely Anzwers and Web Wombat, since both have proven to be tenacious enough to hold on to their guns, unlike other (much better funded) Australian members of the species, which now sleep the sleep of the righteous.

Evaluation is an ongoing challenge (**Webster & Paul, 1996**). Not only is the rate of publishing expanding the indexes and subject hierarchies at an exponential rate, but enhancements and refinements to the search algorithms and the search interfaces change the performance of these tools from one day to the next. New alliances between organizations building and sponsoring these tools, such as Yahoo and Google, change the functionality and, thus, their effectiveness for finding information. Search engines are here one day, and gone the next.

Over recent years there has been a number of attempts at providing some sort of objective evaluation of search engines (see **Chu & Rosenthal, 1996**; **Leighton, 1996; Schlichting & Nilsen, 1996**; **Peterson, 1997**; **Zorn, 1997 and Stojanovski 2000**) . These attempts came mainly from academics and concerned librarians, and have petered out in the last three years as more exciting subjects come up (robots, intelligent agents, etc.) and as the escalating number of new search engines, both specialised and outright crappy, make any comprehensive study impossible.

One main problem with evaluating search engines is the secretiveness with which originators surround their brain-children. It is rather difficult to "get into" the newer search engines as cyber-economics push these tools to become more and more profit making oriented. Secretiveness has always been a major enemy of open research; it is rather difficult to generate any objective judgement based solely on the points advertised by the manufacturing company.

Australian Internet service provider **Ozemail,** one of Australia's largest Internet Service Providers, launched ANZWERS search engine on 27th November 1996 in a bid to go head-to-head with the Australian Alta Vista mirror site jointly operated by **Telstra** and **Digital.** ANZWERS - a joint venture between Ozemail and Inktomi - provides a wide array of search options (including the ability to search by keyword or phrase) as well facilities to search for cross-links and media types such as .gif and .jpg images, shockwave files, java and javascript. Searches can also be localised by region or even down to specific domains (eg: .com.au) or

specific web sites. The new engine uses spider technology to locate pages and recognises the META tag for page descriptions and keywords. It offers speeds closely comparable to Alta Vista. Anzwers is powered by Inktomi search technology. It is kept reasonably up to date.

According to a report from WebMaster World.com, Inktomi technology allows for the following: it is sensitive to spam, will bury pages that show different content per visit, since it's spiders visit often. It will tolerate some degree of higher keyword repetition and overall keyword density, but not irrelevant content and will accept cloaked pages through paid placement as well as free submission. Inktomi will generally only spider submitted pages, rarely following links. Multiple spiders will usually arrive requesting the same page, with each spider using a different IP Address and a mix of User Agents, the index page tends to be requested most.

Web Wombat is a locally developed Australian search portal. The search software is similar in functionality to that of Alta Vista and Excite. The search processes are completely automated, deploying thousands of software robots (known as "parallel scanning technology"). It has a database of searchable information on Australia, with references to more than 11.5 million documents. It also has almost 100 million web pages indexed globally. Web Wombat has been operating as an Internet search engine since 1994. It was one of the world's first Internet search engines, established before Alta Vista, Inktomi and Excite. It was the first search engine in the world to introduce non-search-related content to its web site, making it the world's first search portal. It was also the world's first regional search engine.

Web Wombat has recently coined the term "boutique search engine". This refers to small "vertical" search tools that allow users to search within a specific area of interest (ie accommodation, loans, wine, football, finance)

2.   Evaluation Criteria

According to **Smith (1999)** recall and precision are two ways of evaluating the effectiveness of search engines. Recall is the proportion of possible relevant documents that a search actually retrieves. If there are 100 documents "out there" that are relevant to an information need, and the search engine finds 20 of them, the recall is 20%. Precision is the proportion of the documents recovered in a search that are relevant to the information need: If 10 documents are found, and 6 are relevant, then the precision is 60%. Search engines generally retrieve a large number of documents, then rank them by relevancy. The total number of documents found is less relevant than which ones appear on the first screen of results. For this reason, studies of precision in relevancy ranked search engines (e.g. **Leighton 1999**) usually calculate a "first 10" or "first 20" precision: the proportion of relevant documents in the first 10 or 20 presented by the search engine. I have followed this trend by choosing the first 50 hits, to give the evaluation a wider statistical sample. As I have personally found out, it is easy to overlook real gems if one only sticks to the first 10-20 results of a search.

Relevancy is also a very subjective term. In reality the only person who can determine whether a document is relevant to an information need is the person who has the information need. Perhaps the most rigorous study of relevance in Internet search engines was done by **Gordon (1999)**, who asked faculty to submit actual information needs, and assess the results of searches done by expert Internet searchers. I have also followed this approach in my evaluation, using work colleagues to provide me with the query, and then letting them state how relevant the results were to their needs.

Recall is in fact almost impossible to determine, since it depends on knowing how many relevant documents exist on the Web in total, and this could only be determined by examining the entire Web. In practice "Relative recall" is calculated (e.g. **Clarke 1997**). The results of searches across several search engines are pooled and assumed to approximate the total number of relevant documents, and the recall of individual search engines is calculated from this. Since this evaluation had a limited time scope, recall is not one of the evaluation criteria I used.

I admit that I am neither a brilliant logician, nor a specialist at software evaluation. The criteria chosen below for evaluating the performance of the two Australian search engines, Anzwers and Web Wombat, are based on casual questioning of colleagues at work as to what are the most important (and frustrating) aspects of their online research. I made sure that their answers were not limited to one particular field. The interests ranged from finding legal resources (discussed in detail in my Research Report for this module) to low-

maintanance gardening, feminist issues, knowledge management and who else is a DJ in Brisbane. Below is a comprehensive list of criteria considered important in a search tool:

- How good is the user interface?
- What is the scope of information gathering?
- How easy is it to enter simple or advanced queries?
- Can you modify search results?
- Query formulation:
  - Boolean logic or similar function
  - phrase searching
  - wild cards and truncation
  - upper case/lower case sensitivity
  - result limitation
- How fast is the search?
- How many hits were returned?
- How many false hits did you retrieve?
- How many hits are relevant?
- Are hits listed in order of relevance?
- How are the results structured?
- How current are the hits?
- Are the hits "live" sites?
- Are the help screen and directions easy to understand?
- Does the search engine have additional features?

3. The Query Structure

The queries used were divided into three categories to examine the capabilities of the engines with a number of different subjects and styles of syntax, following the system proposed by **Nasios, Korinthios & Despotopoulos (1998).**

- Category 1:
This category included queries which were posed to the search engines in quotes ("*phrase*"). This forces the search engine to retrieve documents containing the exact phrase included in the quotation marks, provided it supports the format.

- Category 2:
This category included queries posed to the search engines in natural language, that is without quotes, operators or other additions. Since both search engines specialise in Australiana, I did not avoid focusing on any particular field,

- Category 3:
This category includes queries given to the search engines along with additional Boolean operators (i.e. AND, OR, NOT) and special signs (+ -). Finally the ( ) operator is used for more complex operations and particularly combinations of the above operators.

Each search engine had a set of three queries per category applied to it. The queries were carefully formulated after an initial reference interview, as I strongly believe in the GIGO (Garbage In Garbage Out) theory. If the user, allegedly being an intelligent creature, doesn't know what s/he wants, then one should not expect the search engine, based on a dumb, keyword harvesting robot software, to be any more intelligent.

The queries were as follows:

Query 1: The user needed information on symptoms and effects of Ross River Fever. The information had to be easily understood, as the user was not a medical practitioner.

Query 2: The user needed to know which entities in Australia engaged in providing Knowledge Management consulting to corporations and public service.

Query 3: The user wanted to see if there was a follow up on "a guy from Melbourne who in 1997 said he developed an algorithm that allowed whole videos to be transferred through phone lines". The user could not remember the "guy's" name.

ALL SEARCHES WERE PERFORMED ON THE AUSTRALIAN DOMAINS ONLY.

4. Results

Table 1: Ranking of Anzwers and Web Wombat

Criteria	ANZWERS	WEBWOMBAT
User Interface	Bland – also had to go to a different page to do "power search"	Very cluttered – too much graphics, some of which refuse to download
Information Gathering Scope	Gives choice of either Australian, New Zeland (separately)or global search. Has 110 million documents in its database	Gives choice of either Australian and New Zeland sites (together) or global search.
Ease of query entry	Simple search is easy to enter. However, Power search is convoluted, with a few boxes and drop-down menus to go through before one gets to the point.	Easy. Advanced Search is a joke – it is actually the Boolean version of Simple Search Simple search box allows limited number of characters
Modification of Query Results	One can apply the Filter on "Power Search" to eliminate common words that clutter the results	None
Query Formulation:		
Boolean logic	Supports full Boolean searches using AND (by default), OR, NOT and NEAR. Uses () for complex Boolean queries – according to Help files.	Supports AND, OR NOT and NEAR through clicking the appropriate boxes. Complex searches are not supported.
Natural language search	Supports – eliminates stopwords	Supports
Phrase searching	Supports phrase searching ""	Does not support
Wildcards and truncation	Does not support either	Does not support
Upper/lower case sensitivity	Does not support	Does not support
Result limitation	By date (created or modified)/domain/language/feature/scriptlanguage/title. Searches Usenet by Group Name	None
Speed of search	First page returns at acceptable speed. Further pages take for ever.	Good
No of hits returned	Regardless to number, only first 100 are accessible	Does not show number of results, only number of hits per each keyword entered. One has to continue clicking on Next until one runs out of steam.
Avarage % of relevant hits (first 50 hits only)	10%	3%
Avarage % of false hits (first 50 hits only)	90%	97%
Ranking	By word frequency- length of document/search word in <meta>	Unknown

	and \<title> tags/penalises spoofing and spamming	
Results structure	Returns a list of documents in the form of abstracts and hyperlinks. Should be providing page size, however, all sites seem to have 0 bytes.	Returns hyperlinks and an abstract made of the few first words.
Currency of hits	Recent (1998-2000)	Recent (1998-2000)
% of dead links (first 50 hits only)	1%	4%
Help Files evaluation	No examples given. Boolean Operators and Queries Modifiers basically do identical things. Has two links to Help files (one for search tips and the other for FAQ-mostly for website submission)	No Help Files provided
Additional features	Subject directory and news service available. Trying to jump-click results gives a 404 (page cannot be found error)	Automated classifieds advertising system, jobs search database facility and ad serving software. Subject directory available. Free e-mail Womabat owners are so full of themselves, one can read the same discription of how great they are 3 times on their website (About the company, about the software, Press releases)
OVERALL RANKING	GOOD	POOR

Table 2: Results of Queries Performed on Anzwers and Web Wombat

QUERY	ANZWERS	WEB WOMBAT
Query 1: Phrase *"Ross River Fever Symptoms"*	The query produced 1 irrelevant result (0% relevance)	Not Supported
Query 1: Natural Language – Ross River Fever Symptoms	109 recalled, only 100 could be accessed from the search page. Of first 50, 12 were duplicates, 8 of which were on the first page. Of the reminder, 21 sites were irrelevant (34% relevance)	The first remotely relevant item was ranked at 34th point (at 78%). Unknown number was recalled (see above). Of the first 50 sites, 11 were duplicates, 2 were dead links, and 29 sites were irrelevant (16% relevance)
Query 1: Boolean *Ross NEAR River NEAR Fever AND Symptoms* **then** *"Ross River Fever" AND Symptoms* *Ross River Fever Symptoms (all words/exact order then any order for Wombat Advanced Search)*	35 recalled, 3 were duplicates, 1 was a dead link, and 29 were irrelevant (6% relevance) 43 recalled, of which 9 were duplicates, and 29 were irrelevant (11%)	(exact order): Returned no results. (any order) produced identical set of results as the Natural Language Query above.
Query 2: Phrase *"providers of knowledge management consulting in Australia"* *"knowledge management consultants in Australia"*	No results No results	Not Supported

Query 2: Natural Language *Knowledge management consultants in Australia*	5350 returns. Of the first 50 hits, 3 were duplicates, and 41 were irrelevant (12% relevance)	Unknown number returned. Of the first 50 items polled, 4 were duplicates and none was relevant to the search (0% relevance)
Query 2: Boolean *"knowledge management" AND consultants AND Australia* *"knowledge management consultants" AND Australia (or consulting)* *knowledge management consultant Australia (all words/exact order for Wombat Advanced Search)*	220 results returned. Of the first 50 hits, 3 were duplicates, and 1 a dead link. 32 hits were irrelevant (28% relevance) 4 results returned; 1 relevant (25% relevance)	(exact order) no results returned (any order) of the first 50 hits, 10 were duplicates and 34 were irrelevant (12% relevance)
Query 3: Phrase *Melbourne "video compression through telephone lines"*	No results returned	Not Supported
Query 3: Natural Language *Melbourne video compression telephone*	62 results returned. The first item gave the newest story on the query. 4 were duplicates. Only the first item was relevant. (2% relevance)	Unknown number of results. Of the first 50 hits, 1 was a duplicate, 1 dead link and 1 was relevant (2% relevance)
Query 3: Boolean *Melbourne AND "video compression" AND "telephone lines"* *Melbourne video compression telephone (all words/any order for Wombat Advanced Search)*	3 results returned. None of the results was relevant (0% relevance)	produced identical set of results as the Natural Language Query above
Avarage Relevance	**10%**	**3%**

5. Mene, mene, tekel …. USA?

Australia's Web searchers suffer from an identity problem - to customers they all look the same. The main difference between all the search engines is the way they physically search the Web. Unfortunately, that difference is completely invisible to the user, and so has extremely low marketing value **(Saville, 1999)**.

There are millions of Australian Web pages indexed in our search engines. Competition is so intense the search-engine companies are spending about $2 million a month on promotions, according to AIS Media's Steve Allen. Not that the companies themselves will tell you that - in fact, no-one will reveal to you their revenues, profits, or staff numbers. What they will tell you is how popular they are. But a high proportion of the local users are forgetting to add the ".au" to their addresses, so that much of the Australian traffic for companies such as Yahoo and Excite goes to the US parents.

However, that is changing, with the percentage of Australian search engine sites visited by Australians growing from 36 per cent in April to 41 per cent in May. The latest figures from Sinewave's Top100.com.au show that Yahoo.com has the biggest share of local visitors, with an audience reach of 30.5 per cent in May. AltaVista.com is third with 7.06 per cent, trailed by excite.com, at fifth, with 5.76 per cent, yahoo.com.au (sixth) with 3.95 per cent (but rising), looksmart.com.au (eighth) with 4.06 per cent, and anzwers.com.au in 10th place with a 2.63 per cent share. At this early stage, the search engines will need to find money from advertising, and that market is soft. It's as if the number of TV channels were indefinitely multiplying, leading to an oversupply of advertising time.

AltaVista's Paul Whiteway says Australian search engines are earning about 80 per cent of their revenues from advertising and 20 per cent from e-commerce; being mainly tenancy fees from the partner Web sites and commissions on sales generated. Yahoo's Faure says that "it's very easy to look across all ad impressions and come up with average numbers [for cost]. That's fairly misleading". So the fight is on for market share in an industry with the sort of current revenues that barely pay for lunch. Total online advertising revenues in Australia was a meagre $12.5 million in 1998, according to Internet research company www.consult.com.

US and Australian market leader Yahoo is presently painting buses purple and buying space on giant billboards in order to get across the message that Yahoo is the place to be.

6. Conclusions

An article in the Sydney Morning Herald evaluated Anzwers as one of the top searching engines available, giving it 3 out of 4 points. Although one needs to trawl through a few screens to find what one is seeking, but it has plenty of filtering options to help you narrow down the results. It allows looking for exact phrases, images, audio and video content, or focus on particular domains. It returns particular indexed pages more than home pages. This closely agrees with my own findings.

In a recent paper, an Australian scientist from CSIRO tentatively put forward the suggestion that commercial search engines are more interested in lowering their overheads and impressing public with speed and recall, than in algorithms and relevance ranking (**Hawkings, 1999**)
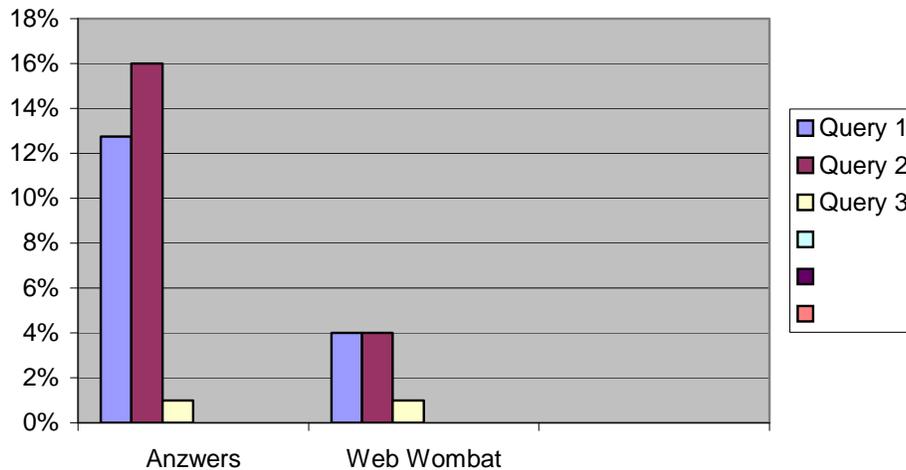

Fig. 1 Relevance Percentage for 3 Different User Queries

As can be seen from the figure above, relevance varies perceptibly between Anzwers and Web Wombat. The result for the last query is rather unusual, and has nothing to do with the way the search engine looks for information. There are thousands of articles about video compression and telephone lines, however, the particular person for whom the user was looking was not written much about, as the media are fearing to be dragged into a scam. This explains the very low relevance ratio for the third query.

Of the two search engines, Web Wombat is slightly better than useless, while Anzwers provides results on par (if not occasionally better) with Yahoo! (using Google) and AltaVista. The downside of Anzwers is that although Inktomi technology alleges that it can eliminate duplicates, Anzwers not only duplicates same content at different URLs, but identical URLs as well.

It is obvious that for a lay person who uses these search tools for a day-to-day information gathering, the sheer amount of irrelevant information one has to wade through is frustrating. It would be of benefit if search engines came up with better abstracting technology (some of the summaries just don't make sense). But then, maybe we are expecting too much of technology and as such lowering our own standards. It is an unfortunate thing that while everyone with access to a PC and telephone line can spread information (and misinformation), very few people possess the skills to make this information easily categorised and retrieved.

7. References

Chu, H. & Rosenthal, M. 1996. Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology. [Online] Available WWW: http://www.asis.org/annual-96/ElectronicProceedings/chu.html

Clarke,S. J. & Willett, P. 1997. *Estimating the recall performance of Web search engines.* **Aslib Proceedings** 49(7, July/August), pp. 184-189

Fallon, D. *Top ten search sites.* **Sydney Morning Herald***:* 24.02.2001 (ICON)

Gordon,M. & Pathak, P. 1999. *Finding information on the World Wide Web: the retrieval effectiveness of search engines.* **Information Processing and Management** 35(2, March), pp. 141-180.

Hawking,D. et al. 1999. Results and Challenges in Web Search Evaluation. [Online] Available WWW: http://www8.org/w8-papers/2c-search-discover/results/results.html

Leighton, H. V. & Srivastava, J. 1999. Precision among World Wide Web Search Services (Search Engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos. [Online] Available WWW: http://www.winona.msus.edu/library/webind2/webind2.htm

Leighton, H. V. 1996. Precision ANOVA of infoseek, lycos, and webcrawler Index Services. [Online] Available WWW: http://www.winona.msus.edu/library/webind.htm

Nasios, Y., Korinthios, G., & Despotopoulos, Y. 1998. Evaluation of Search Engines. [Online] Available WWW: http://piper.ntua.gr/reports/searcheng.pdf

Peterson, R. E. 1997. Eight Internet Search Engines Compared. [Online] Available WWW: http://www.firstmonday.dk/issues/issue2_2/peterson/index.html

Saville, M. *Australian Search Sites Look for Identity.* **Sydney Morning Herald:** 17.06.1999

Schlichting C. & Nilsen, E. 1996. Signal Detection Analysis of WWW Search Engines. [Online] Available WWW: http://www.microsoft.com/usability/webconf/schlichting/schlichting.htm

Stojanovski, J. 2000. Search Engines In-Depth. [Online] Available WWW: http://www.carnet.hr/cuc/cuc2000/radovi/prezentacije/F/F2/f2_f.pdf

Smith, A. G. 1999. Discovery Downunder: Choosing Search Engines for Australasian Topics. [Online] Available WWW: http://www2.auckland.ac.nz/lbr//conf99/smith.htm

Webster, K. & Paul K. 1996. Beyond Surfing: Tools and Techniques for Searching the Web. [Online] Available WWW: http://magi.com/~mmelick/it96jan.htm#tools

Zorn, P. et al. 1996. Advanced Searching: Tricks of the Trade. [Online]. Available WWW: http://www.onlineinc.com/onlinemag/MayOL/zorn5.html