SEARCH STRATEGIES FOR EFFECTIVE INTERNET BROWSING

M. Chandrashekara* and Y. L. Somashekara**

1. Introduction

The growth of the Internet has been phenomenal, particularly since 1993 with the introduction of the World Wide Web as a user-friendly, graphical interface to Internet resources. The Internet now links more than 1,50,000 networks in over 130 different countries. The web has established itself as a highly important, if not an essential, source of information in online searching.

Today's information professional has to be ready to conduct an online literature search on any given subject requested by the library user, and must be able to produce good quality results with speed. This might necessitate the searching of any computerized database server, or on the World Wide Web.

2. Steps in Internet Searching

The steps involved in a typical internet search are :

- Analysis of the information requirement,
- Selecting the search engine,
- Translation of the concepts into syntactical search statements of the selected search engine,
- Performing the search,
- Refining the search based on the results,
- Visiting the actual site and saving the information (Use File- Save option of the browser).

* Lecturer, Dept. of Studies in Library & Information Science, University of Mysore, Manasagangotri, Mysore-570 006
** Assistant Librarian, Mysore University Library, Manasagangotri, Mysore-570 006

Before Starting Internet Search, the browser need to consider the following points.

- Performing the search,
- Consider what's Not Included in Internet
- Content of Adobe PDF and formatted files
- The content in sites requiring a log in CGI output such as search results from a database
- Information within Intranets/Firewalls
- Institutional resources requiring membership
- Materials protected by copyright
- Commercial resources with domain limitations
- Sites that use a robots.txt file to keep files and/or directories off limits
- Non-Web resources (E.g.: Text files)

3. Teachiques of Searching

Below are a number of general tips to help the searcher to make web searching more effective.

a) Be Natural : Type what you want to know, rather than a list of synonyms. Search engines are being taught to understand the same.

If you would've asked a fellow human *"Is alphabet soup nutritious?"* Then ask the search engine "alphabet soup" AND "nutritious" rather than alphabet soup nutrition food health.

b) Capitalize : In general, always use lowercase. If you enter 'star', you will receive hits for Star, STAR, star and so on. However, if you are seeking information on 'India', the country, start it with a capital 'T'. This will exclude a lot of sites which focus on Indian tableware.

c) Use Rare Words : The more unusual or uncommon the keywords you use are, the more specific the results will be. Taking a moment to think of a valid yet uncommon word is a valuable technique.

Example: 'alcohol' returns 912,620 hits (AltaVista) 'vodka' fetched 120,740 and it narrows down to 2754 hits when you enter 'Stolichnaya'.

Note: For a few engines the word order is important, so always enter the rare word first.

d) Require Words : Adding a "+" before a search word will require it to be in every resulting hit.

Say you were after information on an actor from the X-Files, Forbes Angus. Entering those words "Forbes Angus" x-files cast fox brings out 2.7 million sites, with nothing about occasional cast member Forbes Angus on the first page. But if you add "+" to the least common word +"Forbes Angus" x-files cast fox then the top results of the 80 sites found will give you the info you require.

e) Exclude Words : By using a "-". Say you sought the homepage of Bruce Willis, a plumber in Arkansas. To avoid all the millions (actually 134,928) of pages dedicated to the film star, use this: "Bruce Willis" plumber Arkansas -"Die Hard" -movie -superstar -Demi

f) Spell it Right : Google.com allows you to see how much time's keywords or phrases have been searched for. Here is an example of a hard to spell place name Mysore, Karnataka

Mysore 1,13.000
Mysore 4402 (correct spelling)
Mysore 212
Mysore 191

Be aware of the differences between English and American spellings, such as colour & color. In such cases use (colour OR color).

Some search engines, like Ask Jeeves, will check your spelling for you .

g) Recognise "Stop Words" : Search engines ignore the most common words in an effort to speed things up. Several hundreds of these are deemed to be "stop words". They vary from engine to engine, but always contain words like' the', 'of', 'web', 'a', 'to', 'in', & 'is'. It doesn't matter whether they are embedded in a phrase or if they have a + before them, they will not be included in the search. Usually this doesn't matter, but it is smart to be aware of the process.

"Searching the web" contains two stop words: the & web. Consequently the search engine will only look for "searching". If you are aware of this, you can add a more relevant keyword to narrow your search, like: "people search".

Note: Google will let you know of any words it has excluded

FAST does not use stop words

h) Use Wild Cards : The asterisk *covers all possible extensions.

> funk* will find funk, funky, funkiest and funkadelic
>
> str*ed will find stretched, straightened and strapped
>
> "God * toga" will find "God of toga", "God on toga" and "God sold toga"

Note: Lycos uses a "$". Not supported by Excite, Google or Infoseek

i) Reverse Questions : Search engines look for pieces of text that match your query. Web pages are more likely to contain answers than questions - so search for the answer. Phrase your query how you would expect the answer to read - the difference appears slight, but it makes a huge difference .

> "IRS stands for" rather than "What does IRS stand for?"
>
> "man first landed on the moon in" rather than "When did man first land on the moon?"
>
> "sky is blue because" instead of "Why is the sky blue?"

Note: Ask Jeeves is the exception, and comes up with excellent answers to common, natural language questions .

j) Solve Dead Links : Try shortening the URL to the next subheading. Keep doing so until you get to the point that works. Then browse from there to see if you can track down the file that you want.

If http://www.spock.com/jim/life/not_we_know_it.html returns an error, try http://www.spock.com/jim/life/ar. You still get an error, try http://spock.com/jim/ and so on down to the root domain http://www.spock.com.

Note: Google has most of the web cached. If a link is dead, clicking on the Cached link will bring up how it looked when it was indexed.

k) Deal With Huge Pages : Sometimes the reason a page appears in the results is because it is one very long page of text, briefly mentioning hundreds of subjects. Sometimes these are useful, such as in genealogical searches. Often they are not...

In general, the most useful pages will be between 10k and 80k .To find that which you seek within a huge page, use the "Find in Page" option of your browser for Explorer & Netscape it is in the Edit menu.

Google results list a maximum size of 101k. Many of these will be much, much larger and take forever to download.

l) **Use Boolean Phrases :** Named after George Boole, Boolean phrases are a system of logical combinations, using words like AND, OR & NOT. *It is best to always capitalise them.*

AND or "+"

Example: Larry AND Curly AND Moe

        Larry +Curly +Moe

        AND requires the word to be present

OR

Example: Chico OR Zeppo

        OR allows either word to be present

NOT or "-"

Example : Marx NOT Brothers

        Marx -brothers

        NOT excludes words. In this example results should display sites about communism and not comedy.

NEAR

Example: "Salman Rushdie" NEAR teatowel

        Finds keywords within 25 (Lycos) or 10 (Alta Vista) words of each other. Not supported by the other engines.

NEST THEM!

Example: Marx NOT (Brothers OR Moscow)

        ("Jesus Christ" NOT Humor) AND (Mary OR Magdalene)

BUT DON'T GO TOO FAR !!

Example: ((alphabet AND Soup) NOT (twinkies OR "KFC")) AND nutritious

... is too confusing. Use "alphabet soup" AND nutritious

... and if you get a lot of KFC hits, refine the results to exclude them (most top engines support this).

4. Problems of Internet Searching

- Huge and varied information available
- Lack of knowledge of search engines and their capabilities

- Lack of knowledge of search techniques
- Lack of understanding of various components of web pages
- Search engine/catalog distinction; meta-search tools
- Coverage and overlap
- Inclusions, exclusion, ranking, placement
- Caching and replication
- Date of materials
- Transience

5. Choosing the Right Engine.

Don't just use any old search engine - they are all very different . Use the advanced engines These are never on the front page but they should be – It is certain that 95% of users would have no difficulty in understanding them. Ditch the standard engines and bookmark the advanced ones.

Play the field. You will naturally make one engine your favourite, but when you have time you should play with the others. All the engines on this site have a unique advantage, and if you can learn what they are, your searches will become easier.

Don't flog a dead horse. If you are having difficulty in finding the site you want, try the same keywords on another engine (before resorting to Boolean or meditating on more appropriate keywords). Or use a metasearch engine.

According to Nature magazine (1999) the web contained roughly 800 million individual pages and 180 million images. (Amazingly, only 1.5% of the web is pornographic). The best engine listed below only covered 16% of the web.

Search Engine	% of Web Covered	Popularity Ranking	% of Dead Links
Northern Light	16% Best	7	9.8%
AltaVista	15.5%	5	6.7%
Inktomi (Snap)	15.5%	-	2.8%
Inktomi (HotBot)	11.3%	6	2.2% Best
Inktomi (MSN)	8.5%	-	2.6%
Infoseek	8.0%	3	5.5%
Google	7.8%	-	7.0%

Inktomi (Yahoo)	7.4%	**1 Most Popular**	2.9%
Excite	5.6%	2	2.7%
Lycos	2.5%	4	**14.0% Worst**
Euroseek	**2.2% Worst**	-	2.6%
<u>All the above combined</u>	<u>42% of the Web</u>		

In ranking web pages, search engines follow a certain set of rules. These may vary from one engine to another. Their goal, of course, is to retrieve the most relevant pages at the top of their lists. To do this, they look for the location and frequency of words and phrases in the web page document and, sometimes, in the TML META tags. They check out the title field and scan the headers and text near the top of the document. Some of them assess popularity by the number of links that are pointing to sites. The more links, the greater the popularity, i.e., value of the page.

6. Rating Internet Search Engines

The table given provides description, usability and effectiveness of few search engines.

Product	Description	Usability	Effectiveness	Remarks
<u>InfoSeek</u>	Pay-as-you-search sites ($9.95 a month with 100 free queries) with natural-language search strings.	****1/2 Interface, instructions, and customizability excellent throughout.	***** Plenty of hits with unusually high accuracy.	****1/2 The best of the bunch, but it isn't free.
<u>Lycos</u>	Popular site, with good customized searching on over 3 million pages.	**** Provides excellent descriptions of results.	**** High number of hits, and precision of results is usually excellent.	*One of the best free tools for searching the Web.*
<u>WebCrawler</u>	Large database with simple search form, acquired by America Online.	**** Excellent interfaces throughout and easy to use.	***1/2 Very high numbers of hits but too many slightly off topics.	*Strong, and getting better all the time.*
<u>WWW Worm</u>	Over 3 million URLs constantly searched.	***1/2 Easy-to-understand front page with forms-based but limited customization.	***** Excellent quantity of hits and usually very high precision.	*One of the top Web-based search tools.* ***1/

All-in-One Search Page	Colorful maetapage with well-designed interface.	**** Interface reflects unique elements of individual engines.	*** Convenient to access several engines, but slow at retrieving information.	2 Excellent starting point for Internet searching, but not multithreaded.
Internet Exploration Page	Combination multithreaded and individual search page.	*** Good for searching multiple sites.	**** Time limit determines quantity and quality, with strong results.	***1/2 A good start and often all you need.
Open Text	Full-text search engine of over 1 million Web pages.	***1/2 Easy-to-use interface and solid search capabilities.	*** Results always plentiful but far too frequently off target.	***1/2 Continually improving; good for precise search strings.
Yahoo	Superb search site and Web subject index.	*** Easy to use, but limited in search types.	**** Quantity and quality of results consistently good.	***1/2 Always a useful place to start.
AliWeb	Archie-like search engine that holds copies of FTP directories and files for fast searches.	*** If you liké Archie, you'll like AliWeb..	** Always a large number of hits, but far too few directly on topic.	*** Good place for broad searching, but not for precision.
CUSI	Single page with forms for searching Web sites, Gopher, Internet users, and shareware.	*** Easy to use, but little cutomizability and few advanced features.	*** results depend entirely on destination search tool; often slow processing	*** Good site to use to access individual engines.
EINet Galaxy	Well-established search site and subject index.	*** Strong customization, but not intuitive.	*** Consistently high number of results, but unpredictable precision.	*** Better for its subject catalog than its search abilities.
Jump Station II	Innovative search form lets you access large database of agent-acquired Web sites.	***1/2 Strong search form and customizability make you want to return.	** Relatively low number of hits, but reasonable accuracy.	*** Mainly useful to confirm results of other searches.
SavvySearch	Friendly site that searches multiple engines simultaneously.	*** Simple interface, but results are hard to read.	*** Disappointingly low number of hits, but very good precision.	*** Excellent promise for this multithreaded page.
CUI W3 Catalog	One of the oldest search sites; uses Perl strings.	** A pretty good place to start, but who wants to learn Perl?	** Hits are useful but limited in number.	** Rarely a candidate for in-depth search tasks.

7. Feature Comparision Chart of Search Engines.

The table gives the comparative features such as, content size, full text, default word, Book collectors, phrase search truncation words included word elimination, duplicate detection etc.

Database	AltaVista	Excite	Google	HotBot	Lycos	Northern Light
Scope						
Content Size	250M pages & media objects	250M pages	1.25 billion sites	110M sites	50M pages	200M sites
Full-text	Yes	Yes	Yes	Yes	No	Yes
Logic						
Default word	OR	OR	AND	AND	and	and
Boolean connectors	AND, NEAR AND NOT,	AND, AND NOT,	limit including & excluding words	OR, NOT	or, not, adj., near, before, far	or, not
Phrase search	Quotation marks	Quotations marks	Quotation marks	Quotation marks	Quotation marks	Quotation marks
Truncation	No, use *	No	Automatic	No, ? for one right-hand letter; * for left-hand	No	Automatic
Case sensitive	Yes	No	No	Yes	No	No
Words included	Use +	Use +	Use +	Use +	Use +	Use +
Word elimination	Use -	Use -	Use -	Use -	Use -	Use -
Duplicate detection	Grouped under one title	Yes	Grouped Under Categories	Grouped under one title	Yes	Yes
Special features	Limit by date, language, or format field followed by a colon	Concept searching suggests terms	Search any language	Limit by date, language location, page depth	Search f for image and sound files	Custom folders

8. Evaluating Web Pages

a) Checking the Source

You can expect to find everything on the web: silly sites, hoaxes, frivolous and serious personal pages, commercials, reviews, articles, full-text documents, academic courses, scholarly papers, reference sources, and scientific reports. To sort it all out? And to check the authenticity of a page, you need to know how to read a URL expand Let's look at the URL for this paper :

http://www.sc.edu/beaufort/library/bones.html

Here's what it all means:

"http" is the protocol

"www" is the host computer name

"sc" (University of South Carolina) is the second-level domain name

"edu" is the top-level domain name

"beaufort" is the directory name

"library" is the sub-directory name

"bones" is the file name

"html" stands for hypertext mark-up language (that's what the computer reads)

Only a few top-level domains are currently recognized, but this is changing. Here is a list of the domains generally accepted by all, particularly in united states.

edu — educational site

com — commercial business site

gov — U.S. governmental/non-military site

mil — U.S. military sites and agencies

net — networks, internet service providers, organizations

org — U.S. non-profit organizations and others

Because this Internet was created in united states, the country code "US" was not originally assigned to U.S. domain names; however, it is used to designate state and local government hosts, including many public schools. Other countries have their own two letter codes as the final part of their domain names, e.g., UK for United Kingdom; CA for Canada; FR for France, etc.

For a list of ISO Internet Country Codes, go to: HotBot's list of Domains and Country Codes

To assess the authenticity and to present accurate information one you can generally rely on the GOV, MIL and EDU hostnames. The NET, ORG and COM are more uncertain and might require additional verification.

b) Checking the Vital Information

A reputable Web page will usually provide you with the following   information:

* Last date page updated
* Mail-to link for questions, comments
* Name, address, telephone number, and email address of page owner

If the page owner is not readily recognizable, it does not provide you with credentials or authority.

c) Checking the Content

On the web, each individual can be his/her own publisher, and many are. Don't accept everything you read just because it's printed on a web page. Unlike scholarly books and journal articles, web sites are seldom reviewed or refereed. It's up to you to check for bias and to determine objectivity. Who sponsors the page? The Flat Earth Society? Hmmm.. .... Who is linking to the page, and what links to other pages does the page itself maintain?

Look to see if the page owner tells you when the page was last updated. Is the information current? Can it be verified at other, similar sites?

Try to distinguish between promotion, advertising, and serious content. This is getting to be more difficult, as an increasing number of pages must look to commercial support for their continuance.

Watch out for deliberate frauds and hoaxes. Some folks really enjoy playing games on the Web.

d) Assessing Web Page Stability

There is no way to freeze a web page in time. Unlike the print world with its publication dates, editions, ISBN numbers, etc., web pages are fluid. There's no bibliographic control on the web. The page you cite today may be altered or revised tomorrow, or it might disappear completely. The page owner might or might not acknowledge the changes and, if he relocates

the page, might or might not leave a Forwarding address. Try to assess the stability of the pages you reference. Again, one of the best ways to do this is to look closely at the page sponsor, last dated updated, and the authority of the author(s). When you are writing a paper and using web pages as source material, keep a backup of what you find on the web, (either as a printout or saved to disk) so that you can verify your sources later on if need be.

9. Conclusion :

The browsers of Internet are unaware of the various 'role indicators' comfortable to the specific search engines. In the absence of the knowledge and skills to use the role indicators and the use of Boolean algebra it is difficult to get information which is precise, accurate, relevant and comprehensive. Therefore, it can be concluded that the user community be given periodical orientation/training courses for effective and economical use of Internet services.

References :

1. Forrester, William H and Rowlands, Jane L. The Online Searching Companion. London: Library Association,2000.

2. Quercia, Valerie. Internet in a Nutshell. Cambridge: O'Relly,1998.

3. URL : http://searchinginez.com. Revised : June 27,2001.

4. URL : http://www.sc.edu/beaufort/bones.html

5. Sullivan,Danny.Introduction to Search Engine Optimization. At: URL: http://www.searchenginewatch.com .Updated : June 26,2001.

6. URL : http://www.Internet.com

7. Randall, Neil. Search Engine that Could. At: URL : http://www.Ziff.com

☆ ☆ ☆ ☆ ☆