Finding Information on the Internet

by

Tom Des Lauriers

West Valley College

x4027

Learning Objectives

Upon completion, the individual will be able to:

- **Use different search engines to find information on the Internet.**

- **Perform a Boolean search using different search engines to find information on the Internet.**

- **Explain the difference between a search engine and a meta search engine.**

- **Use the built-in help for different search engines to perform a more refined search.**

Finding Information on the Internet

Remember that you have to evaluate the value of the information you find on the Internet. The Internet is a great publishing environment, anyone can put a page up on the Internet. There is a lot of useless, incorrect, and false information on the Internet. You need to have a well-defined search plan if you hope to find the information you are looking for on the Internet. Once you find some information you have to evaluate the value of the information.

Search Engine

Definition: A search engine, in the context of the World Wide Web, is a program that seeks out, visits, and indexes URLs. The resulting index is searched for keywords or phrases entered by a user. The engine turns hyperlinks to sources whose description, title, or content match the words or phrase that people search for.

With an ordinary search engine or search tools (such as Infoseek, AltaVista, Yahoo!, Hotbot, or Excite), you submit keywords to a single database of web-pages owned by the search tool, and you get back a different display of documents from each search engine's unique database of web-pages. Results from submitting very comparable searches can differ widely, but also contain some of the same sites.

Ways to Limit or increase your search:

1. Use "   " around any word(s) or name(s) that you want searched as a phase (i.e. words in the order you type them.)

2. If there are terms or phrases within the "   " that you want present in every document retrieved, precede both by a + with no space.

3. You can also can combine words with the Boolean, AND, OR, NOT, AND NOT.

4. Use a – minus sign in front of a word to exclude words from your search (Be careful with this one).

Boolean Logic and a few tips that work with most search engines

Narrow your key words to narrow your search. If you have a phrase to search, put the phrase in quotes. Otherwise the engine will search for each word separately, e.g., "Thomas Jefferson". Without the quotation marks you will likely get thousands of sites that contain the word Thomas and/or Jefferson and may have nothing about Thomas Jefferson. Some engines use parentheses instead of quotation marks. Check the engine's Help section.

Use Boolean logic: AND, OR, NOT, NEAR, ADJACENT, FOLLOWED BY, BEFORE, AFTER. Type these Boolean operators in capital letters and use spaces between the words. "Thomas Jefferson" AND "Monticello" is an example. With many engines you can use + for AND and - for NOT. For some engines + means look for this word and - means exclude this word, so you would need to put the + or - before each word. For example, to search for Thomas Jefferson the person, you could ask the engine to look for +"Thomas Jefferson"+Monticello-"Louisiana Territory"-"Louisiana"-state. Notice there are no spaces with this type of query.

Some engines are case sensitive for capital letters and some are not. NEAR means, as it says, the key words have to be within a certain number of words of each other. Some engines allow you to specify the range; some do not and use a default number, for example, within 10 words. Use NEAR to look for words on the same page. For example, "Thomas Jefferson" NEAR "Monticello". BEFORE and AFTER work the same way as NEAR. NOT may also take the form BUT_NOT, as in "Thomas Jefferson" BUT_NOT Monticello.

ADJACENT and FOLLOWED BY both mean your key words will have to be next to each other on a site for the engine to report that site to you. Use an asterisk (*) as a wild card. Place the asterisk to the right of the word you are searching and the engine will search for that word plus variations of the word. For example if you type in row* you will find sites that have row, rower, rows, rowing, and rowed. Typing two asterisks will give you even more variations on the word. Typing **think should bring returns not only for think, thinks, thinker, thinking, but also thought, thoughtful, etc. Some search engines use the dollar sign ($) instead of the asterisk. Again, check the "Help" section of the search engine.

Some also allow the use of a question mark (?) in place of a letter in a word. You may be looking for information on the dog breed, collie, but are unsure if it ends in an "e" or a "y". You could try searching for colli?. You can also use parentheses to build a query. "Thomas Jefferson" AND (Louisiana OR state) will get you sites about Thomas Jefferson and the state of Louisiana. Be sure to check your spelling. The engine will look for words exactly as you spelled them. If you are searching for "science" but spell it "science" you will be out of luck.

Type your key words in lower case letters (unless as with Thomas Jefferson in the above examples you know what you are looking for will always be capitalized). Engines will then look for the word in both upper and lower case. If you type the key word capitalized, the engine will only look for sites where the word is capitalized. Use synonyms to broaden a search. For example, if you are looking for information on bias, you could also type in stereotype and prejudice (or prejudic*).

Be sure and read the help section at each site to see how that site operates. Most use at least some of the techniques that I have listed above. Some engines will search only by key words others will do conceptual searches. In a conceptual search you will get sites that the engine thinks are similar in topic to what you are looking for. One site even ranks the order of the search results according to whether the person registering the site with the search engine/index registered for free or paid a fee. Those paying a fee get ranked higher when search results are returned.

Some search sites have people review sites for inclusion in the index. Others use computers to search sites on the Internet and people don't review them at all, and each search site's computer (usually known as a "spider") categorizes the sites in a different way. That is why it is useful to use a meta-search site, like Dogpile, which simultaneously searches many engines and indexes. It really is amazing to see how different the results are from different search sites even when you use exactly the same key words to search.

Some SEARCH TIPS:

1. Use phrase within the "   " whenever you can to help you focus on your target.

2. Keep your searches simple (usually no more than two terms, or two phrases within the "   ", or one term and one phrase within the "   ").

3. If there are terms or phrases within the "   " that you want present in every document retrieved, precede both by a + no space or combine them with the Boolean AND.

4. Use the * at the end of words to get all forms of the word.

5. Use a – minus sign in front of a word to exclude words from your search (Be careful with this one).

6. Make sure your search terms are spelled correctly.

7. The more similar the words are that you use in a search, the more results you will get back. (Example: restaurant cafe bistro)

http://www.infoseek.com/
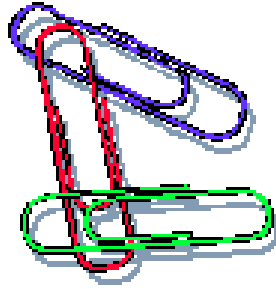
a) Use multiple words when performing your search. Infoseek's search technology will return more refined results from several words than from a single word.

For **example**, typing best pizza in San Francisco will yield more relevant results than typing only pizza or restaurant. (Keep in mind, relevant results are returned even if they don't contain all query terms.)

b) Capitalize proper nouns. Lowercase words will match any case. For **example**, typing next will return documents containing the words next, Next, and NeXT. However, typing NeXT will only search for pages about NeXT. Capitalized names that are adjacent to each other will be treated as a single name.

For example, typing Rock Hudson will return pages with the actor's name in them. However, rock hudson will return pages about the actor, as well as pages about rocks, the Hudson River, and the Hudson Brothers.

c) Separate proper nouns with a comma to indicate different names, and ensure the best results.

For example if no comma is included, the phase will be treated as White House Bill Clinton, rather than White House and/or Bill Clinton.

a) The AltaVista search service helps you find documents on the World Wide Web. Here's how it works. You tell the search service what you're looking for by typing in keywords, phrases, or questions in a search box. The search service responds by giving you a list of all the Web pages in their index relating to those topics. The most relevant content will appear at the top of your results.

b) When searching, think of a word as a combination of letters and numbers. The research service needs to know how to separate words and numbers to find exactly what you want on the Internet. You can separate words using white space and tabs.

c) You can link words and numbers together into phrases if you want specific words or numbers to appear together in your result pages. If you want to find an exact phrase, use "double quotation marks" around the phrase when you enter words in the search box. You can also create phrases using punctuation or special characters such as dashes, underscore lines, commas, slashes, or dots.

   Example: Try searching for 1-800-999-9999 instead of 1800 999 9999. The dashes link the numbers together as a phrase.

http://www.yahoo.com/
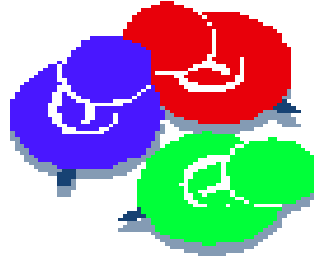
a) Yahoo!, like the Web itself, is too large to be explored entirely link by link. However, with over half a million sites divided into more than 25,000 categories, Yahoo! is both browseable and searchable. Use these two features and you will almost always find something to match your interests.

b) Yahoo, search looks for three main things: Yahoo categories, web sites listed in Yahoo!, and web pages indexed by Inktomi.
For the first two of these, Yahoo! searches for matches in its database and then ranks the results in order of most relevant too least relevant. Some of the factors that affect relevancy are:

1) The number of search words matched. The more words matched, the higher the rank.
2) Exact word matches. These are ranked higher than approximate matches.
3) Where in an entry the search words were found. A match in the title of a site is ranked higher than a match in the comments or the URL.

c) The third set of results, a full-text search of the entire Web, is provided by Inktomi, a Yahoo! partner. Inktomi uses what's called a search engine, and it specializes in indexing every single web page it can find. This gives it a lot of raw data. Search engines provide good results with very specific requests, but often-poor results with general requests. Chances are, however, that if your search terms didn't match anything in Yahoo!, you are doing a very specific search, and Inktomi's results may be right on the money.

http://www.hotbot.com/

a) Getting started is pretty straightforward: Type in search terms or keywords relating to your query, in the search box and hit the Search button. HotBot then searches the full text of the more than 110 million documents in its database, finds all documents that fit your criteria, sorts them according to their relevance, and returns a list of documents in the form of abstracts and hyperlinks.

b) HotBot's pull down menus and buttons let you modify or refine your search. Use these tools to narrow your search by placing more conditions on the query. The more you tell HotBot about what you're looking for, the more accurate your results will be.

c) Be specific. Try to pick words that are unique to the topic you're investigating. If you're looking for information about the Virginia state motto, enter all three words in your search. If you enter Virginia motto, HotBot may give you pages that discuss mottos, but not the state motto.

http://www.excite.com/

a) Search for ideas and concepts, in addition to keywords, and enter more than one keyword in your search query. Excite uses Intelligent Concept Extraction (ICE) to find relationships that exist between words and ideas, so the results will contain words related to the concepts for which you're searching.

b) If you find that one of the Web results better describes what you are searching for, click **More Like** This located next to the site title. Excite Search will use that document as the basis for a new search to find more sites similar to the result you selected.

c) Excite's list of search results may present several pages from the same site. When you click List by Web Site, your list will be compressed to display Web site titles and relevant documents contained within the sites.

List of some Search Engine's

Infoseek -- http://www.infoseek.com/

Altavista -- http://www.altavista.com/

Yahoo -- http://www.yahoo.com/

Hotbot -- http://www.hotbot.com/

Excite -- http://www.excite.com/

Netscape Search -- http://search.netscape.com/

Lycos -- http://www.lycos.com/

Snap -- http://www.snap.com/

Look Smart -- http://www.looksmart.com/

GoTo -- http://www.goto.com/

Web Crawler -- http://webcrawler.com/

AOL Net Find -- http://www.aol.com/netfind/

List of Specialized Search Engines -- http://www.searchengines.net/specialized.htm

List of Foreign Language Search Engines -- http://www.searchengines.net/language.htm

Reference Search Engine -- http://encyclopedia.com/

Search Engine Index -- http://www.merrydew.demon.co.uk/search.htm
An Index of many different types of search engine's

Meta-Search Engine

In a meta-search engine, you submit keywords in its search box, and it transmits your search simultaneously to most of the popular search engines and their databases of web pages. Within a short time, you get back a compilation of results containing matching sites from all the search engines queried. This can save you a lot of time and provide an overview of the kinds of documents out there matching and term, phrase in quotes, or set of terms and phases. **No matter how good meta-search engine's sound they do not eliminate the need to learn how to search skillfully using at least a few web search engines.**

Meta search sites let you use lots of search engines at once.

Dogpile -- http://www.dogpile.com

 Search 25 popular engines as well as Usenet newsgroups; using a Dogpile form, you can set the order that the Dogpile Meta-Search Engine will use to send your query to 25 search engines that the Dogpile Meta-Search Engine supports. Once you choose and save the order of the search engines, you will never have to do it again

Manna -- http://www.mamma.com

The mother of all search engines let's you put the top engines to work, even searching for MP3, images, or sound files. Mamma is what is called a smart meta search engine. Mamma will analyze your query and properly format the words for each source it probes. Mamma will then organize the results into a uniform format, and presents them by relevance and source. Mamma makes it easy for you to find what you need. Also note that Mamma is not case-sensitive.

Metacrawler -- http://www.metacrawler.com

Search nine of the most popular engines or Usenet newsgroups. One of MetaCrawler's advantages over other search engines is the standardization of query syntax. Although some people find the use of punctuation a little off-putting, it is really quite straightforward to use, and you can dramatically improve the quality of your results by using these features. When you specify a search, list groups of words that must occur together by surrounding them in quotes. Also, specify words or phrases that must appear in documents by prefixing them with a plus sign, and specify words or phrases that must not appear by prefixing them with a minus sign. There should be no space between a + or - sign and the word to which it applies. There should be a space between individual query terms.

ProFusion -- http://www.profusion.com

 Lets you choose the three best engines based on your query, the three fastest
engines, or all the engines. ProFusion Intelligent Search employs intelligent
agent technology to continuously monitor and adjust to the performance of
each of the search engines it uses. ProFusion analyzes incoming queries,
categorizes them, and automatically picks the best search engines for the
query based on adaptive confidence factors which encapsulate knowledge of
the past performance of each search engine for the chosen category.
ProFusion merges the search results returned by the selected search engines
based on the values reported by the search engines themselves and their
respective confidence factors. ProFusion continuously updates its knowledge
of a search engine's quality by keeping track of which search engine
provided the link that the user actually follows on the result page. Thus, the
confidence factors are continuously adjusted based on implicit collaborative
feedback from ProFusion users.

Savvy Search -- http://www.savvysearch.com

Create your own custom metasearch site. SavvySearch is best employed as a time saving tool to assess the state of available information from a single, central, convenient access point. The power of metasearch is the time it saves you.   SavvySearch collections integrate results from multiple remote sites on the fly, intelligently applying the different (and sometimes confusing) query languages on your behalf. Casual searchers may find everything they need with a single SavvySearch query. Expert searchers may appreciate a SavvySearch to quickly home in on remote engines that warrant an in-depth search. SavvySearch provides direct links to each reporting search service, allowing you to travel directly to these sources and take advantage of other advanced features they may offer. SavvySearch Structure SavvySearch sends your query to a collection of remote search engines. You can decide where you would like to direct your query by choosing a collection. If choose not to decide, your query is sent by default to a sensible selection of Internet guides or search engines that are suitable for most general queries.

List of some Meta-Search Engine's

DogPile -- http://www.dogpile.com/

Mamma -- http://www.mamma.com/

MetaCrawler -- http://www.go2net.com/search.html

ProFusion -- http://www.profusion.com/

Savvy Search -- http://www.savvysearch.com/

The Internet Sleuth -- http://www.isleuth.com/

Super Search -- http://www.robtex.com/search/query.htm

All 4 One -- http://all4one.com/

Starting Point -- http://www.stpt.com/

Multi Meta -- http://www.stpt.com/

Cyber 411 -- http://www.cyber411.com/

Highway 61 -- http://www.highway61.com/

Web search -- http://www.web-search.com/

List of some other Sources of Information

Citing online sources
http://clever.net/quinion/words/articles/citation.htm

Style Manuals in Brief: APA and MLA
http://www.utexas.edu/depts/uwc/.html/citation.html

A Style Sheet for Citing Internet Resources
http://english.ttu.edu/kairos/2.2/news/youcanuse/craig/page10.html

Copyright Clearance Center
http://www.copyright.com

Copyright for Schools
http://www.li.suu.edu/library/Plagiarism/Copyright.htm

Tutorial on Using Major Search Engines
http://www.seekhelp.com/tutorial.htm