WHAT IS THE WEB?  WHAT DO SEARCH ENGINES SEARCH?

The Internet is a system of protocols (a protocol is a specified means of communication) which allow networked computers to communicate transparently across multiple, linked packet networks. These protocols are known as TCP/IP (Transmission Control Protocol and Internet Protocol) and are comprised of telnet, email, ftp, http, USENET, gopher, and others. These protocols allow thousands of different types of computers linked to the Internet to speak to each other. Packet switching forces the information request into packets (or small chunks of information), each packet being sent along the least busy segments of the network and reassembled into a full document when it reaches its destination.

The World Wide Web (WWW) or “Web” is really a subset of the protocols that make up the Internet. The Web protocol is called http, it uses html coded “pages”, and is a distributed hypermedia system that allows distribution not only of text, but of graphics, sound and animation. It is a network of information within the network of the Internet. The Web uses a hypertext system to make navigation of that information possible. When a document is displayed, words on the screen may be linked to other locations in that document, or to other documents or sources. To see information on the Web, you must know an address. This address is referred to as a URL (Uniform Resource Locator) and is in the form of http://www.asbury.edu/ In order to view these pages on the web, you need a browser (such as Netscape, Firefox, or Internet Explorer) which will format the coding of the html into a layout that is readable by a person, viewable on a computer screen.

When you “search the web” you are actually searching this smaller segment of the total internet comprised of the html coded web pages. While the amount of information contained on these pages is very large, it is not the total amount of information that is available. Everything is NOT on the internet or the web. What is available on the Internet is often what can most easily be translated into electronic form, which is not the same as the best information. Numbers, statistics, facts, etc can easily be loaded onto a computer. However, carefully researched, thoughtful scholarship takes longer to produce and is not as likely to make it to the web.

RECOMMENDED APPROACH TO SEARCHING

Step #1 - Analyze the topic to decide where to begin

Does your topic have:

- distinctive words or phrases?
- have NO distinctive words or phrases you can think of? You have only common or general terms that get the "wrong" pages.
- seek an overview of a broad topic?
- specify a narrow aspect of a broad or common topic?
- have synonymous, equivalent terms, or variant spellings or endings that need to be included?
- Make you feel confused? Don't really know much about the topic yet? Need guidance?
Step #2 - Pick an appropriate starting point

<table>
<thead>
<tr>
<th>Topic Analysis:</th>
<th>Search Engines</th>
<th>Subject Directories</th>
<th>Specialized Databases &quot;Invisible Web&quot;</th>
<th>Find an Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinctive or word or phrase?</td>
<td>Enclose phrases in &quot; &quot; Test run the word or phrase in Google.</td>
<td>Search the broader concept, where does the term fit in a larger idea?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO distinctive words or phrases?</td>
<td>Use more than one term or phrase in &quot; &quot; to get fewer results.</td>
<td>Try to find distinctive terms in Subject Directories</td>
<td>Want data? Facts? Statistics? All of something? One of many like things? Schedules? Maps?</td>
<td>Look for a specialized subject directory on the topic. E-mail the author of any appropriate web pages. Ask a discussion group or expert. Never hurts to seek help.</td>
</tr>
<tr>
<td>Overview or broad topic?</td>
<td>NOT RECOMMENDED</td>
<td>Look for a specialized Subject Directory focused on your topic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narrow aspect of topic?</td>
<td>Boolean searching as in Yahoo! Search.</td>
<td>Look for a Directory focused on the broad subject.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synonyms, equivalent terms, variants</td>
<td>Choose search engines with Boolean OR, or Truncation, or Field limiting.</td>
<td>NOT RECOMMENDED</td>
<td>Look for a specialized database on the Invisible Web. See a Librarian for assistance.</td>
<td></td>
</tr>
</tbody>
</table>

Step #3 - Learn as you go and vary your approach with what you learn

Don't assume you know what you want to find. Look at search results and see what you might use in addition to what you've thought of.

Step #4 - Don't bog down in any strategy that does not work

Switch from search engines to directories and back. Find specialized directories on your topic. Think about possible databases and look for them. Ask for help from knowledgeable librarians.

CHOOSING A SEARCH ENGINE

Search engines are used to perform keyword searches in large databases of information culled from the Internet. To effectively use any type of search engines, it is important to understand what they are, how they work, and how they differ. When searching for sources to use for research purposes, note that the online databases supplied by the Kinlaw Library through noah.asbury.edu/library and the materials in the Library Online Catalog have already been carefully evaluated by librarians. Librarians have also reviewed the web sites listed on the subject pages on noah at: http://noah.asbury.edu/library/bysubject.htm Those resources are recommended to be used in conjunction with general searches done on the Internet.
Search engines can be grouped into four distinct categories (active, passive, meta, and specialized). Each search engine will perform differently based on its underlying techniques. It is important to know how a search engine works as that will effect the types of results that are produced. Below are listed the four types of search engines with examples for each type.

Active (Robot Generated) Search Engines [quantity over quality]

Active search engines rely on computerized retrieval mechanisms. Referred to as "spiders", "crawlers", or "robots", these mechanisms visit Web sites and retrieve relevant keywords to index and store in a searchable database.

Advantages
- Revisit Web sites on a regular basis to look for changes or updates.
- Provide a ranking of search results based on relevancy.
- Useful for accessing hard-to-find information.

Disadvantages
- Rankings are sometimes questionable because they are computer-generated rather than human produced.
- Lack of human involvement can lead to poor quality control.

Examples
- Google - http://www.google.com
- Yahoo - http://www.yahoo.com
- Ask - http://www.ask.com
- All The Web - http://www.altheweb.com
- Alta Vista - http://www.altavista.com
- GigaBlast - http://www.gigablast.com

Passive (Human Selected) Search Engines [quality over quantity]

Passive search engines, often referred to as directories, are human controlled and do not roam the Web directly. They rely on individual submissions that are reviewed and indexed by subject category.

Advantages
- Provide users with a subject directory and a short description of each site.
- Human involvement usually produces a higher relevancy of results.

Disadvantages
- Contain far fewer records than active search engines.

Examples
- Librarians Index - http://lii.org
- InfoMine - http://infomine.ucr.edu
- LookSmart - http://search.looksmart.com
- About.com - http://www.about.com
- Yahoo (classic) - http://dir.yahoo.com
Meta Search Engines

Meta search engines do not create their own databases. They rely on existing active search engine indexes to retrieve search results. In a meta-search, keywords are submitted to the individual search engines and they are searched simultaneously. Since most meta-search tools only retrieve the top 10-50 hits from each search engine, the total number of hits retrieved may be considerably less than found by doing a direct search on one of the search engines. Some meta-searches offer the ability to customize which search engines are queried. The result list will be sorted according to ranking features or in some kind of lumping format. Some "smarter" meta-searching technology includes clustering and linguistic analysis that attempts to show themes within results, with some fancy textual analysis and display that can help you dig deeply into a set of results. Use these with caution and be sure to check the features of the meta-search engine before using.

Advantages
- Remove most duplication from search results.
- Save users time by searching several search engines at once.

Disadvantages
- Redundancies can occur in search results.
- Displayed results are often confusing.
- It is not always clear how the search system works.
- Possibility that they are not very thorough or exclude major search engines.

Examples
- Clusty - http://clusty.com
- DogPile - http://www.dogpile.com
- Kartoo - http://www.kartoo.com
- Mamma - http://www.mamma.com
- MetaCrawler - http://www.metacrawler.com
- Fazzle - http://www.fazzle.com (advanced search is small - at very top)
- IxQuick - http://www.ixquick.com
- ZapMeta - http://www.zapmeta.com

Specialized Search Engines

Specialized search engines offer additional features that active, passive and meta search engines do not provide.

Kartoo - http://www.kartoo.com
- Visual presentation of results.
- Groups results by keywords and concepts.
- Shows results based on relevancy.

Ask - http://www.ask.com
- Allows users to pose search requests in the form of questions rather than keywords.
- Attempts to direct users to the exact page or pages that answers the question.
- Has a back-up metacrawler component that will search other indexes if a match is not found within its own database.

One Page Mega Search - http://www.bjorgul.com/
- Single page with many search boxes.
- Offers many choices not listed elsewhere.

THE INVISIBLE WEB
There's lots of helpful information locked away in databases that can never be indexed by search engines because this information is buried deep under many layers of web pages. LARGE amounts of other information is located in databases that have subscription fees. This information is contained in the "invisible web" or "deep web." NO search engine can reach it or search it. A good searcher should be aware of this type of information. See http://brightplanet.com/infocenter/largest_deepweb_sites.asp for a listing of these types of sites. See: http://www.completeplanet.com/ to search for the appropriate deep web sites by subject of each search site. See Widener's page on what search engines will not find: http://www.widener.edu/Tools_Resources/Libraries/Wolfram_Memorial_Library/Need_Help_/How_to_Do_Research_/What_Web_Search_Engines_Won_t_Find_/498/

Examples:
Library subscribed databases like EbscoHost at Kinlaw Library
U.S. Government data contained deep on their web sites (like census.gov)
Professional subscription databases (company or stock info or those at association web pages)
Individual articles at subscription magazine or newspaper sites (like New York Times)

For more information
Kansas City Library's Introduction to Search Engines - http://www.kclibrary.org/guides/searchengines/

FINDING THE BEST SEARCH ENGINE

Choose the best search for your purpose -
http://www.noodletools.com/debbie/literacies/information/5locate/adviceengine.html

Search Engine Features
http://www.searchengineshowdown.com/features/

Listing of many search engines
http://www.allsearchengines.com/

Anatomy of a URL
http://libweb.uoregon.edu/it/webpub/url-anatomy.html

EVALUATE THE RESULTS

All results should be evaluated with a critical eye. The web is a confusing and crazy place. Since anyone can post anything at any time, the information is ever changing and can be misleading, false, bogus, designed to provoke, or only minimally useful. Be critical. Just as you would consider the credentials of an author or publisher of a written source before accepting information, you must evaluate the source of information on the Internet. Who created the site? Is it updated? Is it simply a hoax? Anyone with technical skills and Internet access is able to put information on the Internet.

Below are some methods for deciding if the information discovered is really going to be useful, valuable, or worthy.

Challenge
Challenge information and demand accountability. Stand right up to the information and ask questions. Who says so? Why do they say so? Why was this information created? Why should I believe it? Why should I trust this source? How is it known to be true? Is it the whole truth? Is the argument reasonable? Who supports it?

Credibility
Is it a trustworthy source? What are the author’s credentials? Is there evidence of quality control? It is produced by a known or respected authority? What is the organizational support? Goal: an authoritative source, a source that supplies some good evidence that allows you to trust it.

Accuracy
Is the information up to date, factual, detailed, exact, and comprehensive? Who is the audience? Does the purpose reflect intentions of completeness and accuracy. Goal: a source that is correct today (not yesterday), a source that gives the whole truth.

Reasonableness
Is the information fair, balanced, objective, reasoned, with no conflict of interest? Is there an absence of fallacies or slanted tone? Goal: a source that engages the subject thoughtfully and reasonably, concerned with the truth.

Support
Does the site contain listed sources, contact information, and available corroboration? Are the claims supported and further documentation supplied? Goal: a source that provides convincing evidence for the claims made, a source you can triangulate (find at least two other sources that support it).

Adapt
Adapt your skepticism and requirements for quality to fit the importance of the information and what is being claimed. Require more credibility and evidence for stronger claims. You are right to be a little skeptical of dramatic information or information that conflicts with commonly accepted ideas. The new information may be true, but you should require a robust amount of evidence from highly credible sources.

USEFUL LINKS FOR EVALUATING WEB SITES

Excellent site for K-12 - LOTS of links -
http://school.discovery.com/schrockguide/eval.html

Cornell Study Guide on Critically Analyzing Information Sources
http://www.library.cornell.edu/t/help/res_strategy/evaluating/analyze.html

Checklist for Evaluation - from Univ. of Alberta

Another checklist -
http://www.lib.purdue.edu/itd/techman/eval.html

Criteria from UCLA and Esther Grassian -
http://www.library.ucla.edu/libraries/college/help/critical/index.htm

Two From Berkeley -
http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html
http://www.lib.berkeley.edu/TeachingLib/Guides/Evaluation.html

One from Johns Hopkins -
http://www.library.jhu.edu/researchhelp/general/evaluating/

Good Bad and the Ugly - Criteria and Examples -
http://lib.nmsu.edu/instruction/evalcrit.html

Module on Web Evaluation and new Tutorial-

Listing of Types of Fallacies -
http://www.nizkor.org/features/fallacies/

Research on Web Credibility from Stanford -
http://www.webcredibility.org/

Oft quoted - article on net from Hope Tilman -
http://www.hopetillman.com/findqual.html

Virtual Salt - Evaluation criteria
http://www.virtualsalt.com/evalu8it.htm

This guide is adapted from -
Berkeley Library Web Tutorial found at: http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/About.html
and http://www.library.cornell.edu/t/help/res_strategy/evaluating/evaluate.html

Created: 6/97
Revised: 9/07
JW

Kinlaw Library, Asbury College, Wilmore, KY 40390