IHS Literature and the Arts
Searching the World Wide Web
***adopted from OWL Purdue**

Searching the World Wide Web can be both beneficial and frustrating. You may find vast amounts of information, or you may not find the kinds of information you're looking for. Searching online will provide you with a wealth of information, but not all of it will be useful or of the highest quality.

The World Wide Web is a superb resource, but it doesn't contain all the information that you can find at a library or through library online resources. Don't expect to limit your search to what is on the Internet, and don't expect search engines to find everything that is on the Web.

Studies of search engine usage show that search engines are increasing exponentially in their indexing of new Web sites and information. Indexing is the Web term for finding and including new Web pages and other media in search results. However, search engines still only index a fraction of what is available on the Internet and not all of it is up to date. Search engines may only "crawl" sites (or revisit them for purposes of indexing) every month or so; information that has been updated since that time will be invisible to the search engines. After you try several search engines, you will see that you get different results from different sites. Also, remember that some information appears and then disappears from Web sites. Finally, search engines don't always search the entire page; if a page is larger than 100 to 500 k, many search engines will only index the first 100 to 500k of the page. So there could be valuable information that is being overlooked by a search engine even in pages that are indexed.

Not all of the information located on the Internet can be found via search engines. Researchers Chris Sherman and Gary Price call this information the "invisible Web" (another name that is frequently used is the "deep Web"). Invisible Web information includes certain file formats, information contained in databases, and pages omitted from search engines.

So, using search engines is not the only way to find material on the Web, but these search engines are one tool you can use. Knowing a few search strategies and hints can make the search more profitable. This guide provides information on the different ways of locating material on the Web including using search engines,

searching the invisible Web, and using Web directories.

The Internet is made up of a vast amount of computers networked throughout the world via data lines or wireless routers. New computers and Web sites are added every day, and no larger organizational system exists to document and catalogue them all. The Internet is a dynamic, growing, and changing system, which makes navigating it or searching it thoroughly difficult.

This is where search engines and Web directories come in. Search engines, such as Google or Yahoo, are large databases of information that store and retrieve relevant website results based on keywords. Web directories, such as the Open Directory Project, are attempts to organize the best of the existing Web sites into categories and subcategories. No search engine or Web directory will have the same sites listed in the same order, and none will have all of the possible sites on the Internet listed. Furthermore, the ranking of a Web site within a search engine (i.e. how high up on the results list it appears) has as much to do with politics as it does with quality information. The search engine rankings are determined by a number of factors including the amount of information on the site, the amount of other sites that link to it, the number of people who select that link when searching, the length of time that the site has been listed in the search engine database, and the code of the site.

Recently, search engines such as Google and Yahoo have also been providing "sponsored links"—links that appear on the first few pages of the search results and that are paid for by advertisers. This means that you may end up clicking on something that is not relevant to your search, but instead actually advertising. The image presented here gives you an example of this on Google.

Searching with a Search Engine

A search engine is a device that sends out inquiries to sites on the Web and catalogs any Web site it encounters, without evaluating it. Methods of inquiry differ from search engine to search engine, so the results reported by each one will also differ. Search engines maintain an incredibly large number of sites in their archives, so you must limit your search terms in order to avoid becoming overwhelmed by an unmanageable number of responses.

Search engines are good for finding sources for well-defined topics. Typing in a general term such as "education" or "Shakespeare" will bring back far too many results, but by narrowing your topic, you can get the kind (and amount) of information that you need.

Example:

- Go to Google (a search engine)
- Type in a general term ("education")
- Add modifiers to further define and narrow your topic ("rural education Indiana")
- Be as specific as you can ("rural education Indiana elementary school")
- Submit your search.

Adjust your search based upon the number of responses you receive (if you get too few responses, submit a more general search; if you get too many, add more modifiers).

Learn how the search engine works

Read the instructions and FAQs located on the search engine to learn how that particular site works. Each search engine is slightly different, and a few minutes learning how to use the site properly will save you large amounts of time and prevent useless searching.

Each search engine has different advantages. Google is one of the largest search engines, followed closely by MSN and Yahoo. This means that these three search engines will search a larger portion of the Internet than other search engines. Lycos allows you to search by region, language, and date. Ask allows you to phrase your search terms in the form of a question. It is wise to search through multiple search engines to find the most available information.

Know Boolean operators

Most search engines allow you to combine terms with words (referred to as Boolean operators) such as "and," "or," or "not." Knowing how to use these terms is very important for a successful search. Most search engines will allow you to apply the Boolean operators in an "advanced search" option.

AND

AND is the most useful and most important term. It tells the search engine to find your first word AND your second word or term. AND can, however, cause problems, especially when you use it with phrases or two terms that are each broad in themselves or likely to appear together in other contexts.

For example, if you'd like information about the basketball team Chicago Bulls and type in "Chicago AND Bulls," you will get references to Chicago and to bulls. Since Chicago is the center of a large meat packing industry, many of the references will be about this since it is likely that "Chicago" and "bull" will appear in many of the references relating to the meat-packing industry.

OR

Use OR when a key term may appear in two different ways.

For example, if you want information on sudden infant death syndrome, try "sudden infant death syndrome OR SIDS."

OR is not always a helpful term because you may find too many combinations with OR. For example, if you want information on the American economy and you type in "American OR economy," you will get thousands of references to documents containing the word "American" and thousands of unrelated ones with the word "economy."

NEAR

NEAR is a term that can only be used on some search engines, and it can be very useful. It tells the search engine to find documents with both words but only when they appear near each other, usually within a few words.

For example, suppose you were looking for information on mobile homes, almost every site has a notice to "click here to return to the home page." Since "home" appears on so many sites, the search engine will report references to sites with the word "mobile" and "click here to return to the home page" since both terms appear on the page. Using NEAR would eliminate that problem.

NOT

NOT tells the search engine to find a reference that contains one term but not the other. This is useful when a term refers to multiple concepts.

For example, if you are working on an informative paper on eagles, you may encounter a host of Web sites that discuss the football team the Philadelphia Eagles, instead. To omit the football team from your search results, you could search for "eagles NOT Philadelphia."

The following is a list of some of the most powerful search and metasearch engines and most comprehensive web directories.

- All4one: One of the first metasearch engines, All4One allows simultaneous searching of 10 major search engines.
- Bing: Microsoft's search engine.
- Dogpile: A metasearch engine that will search Google, MSN, Yahoo, and Ask.
- Environment Web Directory: A web directory that focuses on environmental and health issues.
- Excite: A search engine that lets you search by language, for video, audio, and mp3, and by relevant date.
- Google: Includes a new type of search, "Google Scholar," which allows you to search for more academically-oriented searches.
- Lycos: A search engine that allows for news searches but does not have many advanced search features.
- Metacrawler: A metasearch engine and will search other search engines.
- The Open Directory Project: One of the largest and most comprehensive human-edited directories in the world. Only higher quality websites will be listed here as each site submitted must be approved by a directory editor.
- WebCrawler: Another search engine that allows searching by location, domain name, and for multimedia.

The invisible web includes many types of online resources that normally cannot be found using regular search engines. The listings below can help you access these resources:

- Alexa: A website that archives older websites that are no longer available on the Internet. For example, Alexa has about 87 million websites from the 2000 election that are for the most part no longer available on the Internet.
- Complete Planet: Provides an extensive listing of databases that cannot be searched by conventional search engine technology. It provides access to lists of databases which you can then search individually.

- The Directory of Open Access Journals: Another full-text journal searchable database.
- FindArticles: Indexes over 10 million articles from a variety of different publications.
- Find Law: A comprehensive site that provides information on legal issues organized by category.
- HighWire: Brought to you by Stanford University, HighWire press provides access to one of the largest databases of free, full-text, scholarly content.
- Infomine: A research database created by librarians for use at the university level. It includes both a browsable catalogue and searching capabilities.
- MagPortal: A search engine that will allow you to search for free online magazine articles on a wide range of topics.