



Searching the Web

The tools people use to search the web have changed even in the last few years. In the late 1990's, there were multiple search engines to choose from, each with a slightly different method of collecting information. Today, while there are still many search engines, the most robust engine is Google. Google uses a context algorithm to help determine the best matches for your search term. It also remembers which pages were the most popular among others that used the same search term. Generally, a search engine collects web page information by using a "spider." A spider is a small piece of software that roams the web and records the information that appears on each page. In this section, we'll see how you can improve the accuracy of your searches and introduce you to some valuable resources on the web that can make your searches even better.

Searching Tips

You are probably familiar with the problem of getting too many results when searching for a specific term. Oftentimes the search engine will display web pages that have nothing to do with your actual topic. Here are some guidelines for tightening your search.

TIP:

Remember that many search engines, including Google, offer a service that allows individuals or companies the option to pay the search engine to ensure that their page shows up first. These will often be under a heading like "Sponsored Links." Be aware that these pages are generally not the best match for your search term.

Use multiple terms. Generally, when you enter several terms into the search blank, the search engine will try to find pages that have all the terms you listed. This can help you avoid pages that are in the general subject area you want, but that don't have quite the information you are looking for. Use uncommon terms. If you are looking for something specific, be sure to include any uncommon terms that are not likely to appear on pages unrelated to the specific subject you need. Use the + sign. With most search engines, placing a plus sign (+) at the beginning of a word will force the engine to return only pages that do contain that word. Use the - sign. The minus sign (-) is just the opposite of the plus sign—when you enter a minus sign before a word, the search engine should ignore pages which contain that word.

Search for phrases, rather than individual words. If you put quotation marks (") around a group of words, the search engine will try to find pages which contain that exact phrase—in other words, those words together in that order.



There are also a few things you can do to increase the number of pages found, but in ways that will not stray too far from your search topic. Use wildcards. When you put an asterisk (*) at the end of a word, the search engine will allow any letters after that point in the word. For example, searching for fisher* will actually yield any pages which contain words like fisherman, fishermen, fisherwoman, fishery, etc.

Use parentheses. When you want to search for any word from a group of words, you can put the entire group inside parentheses, separated by commas. To find a group of words similar to those in the previous example, you might search for (fisherman, fishermen, fisherwoman, fisherwomen). This would only find pages with one or more of those four words on them.

Using Boolean Searches

Computers carry out operations based on logical statements. This has been incorporated into almost every search engine in the form of Boolean Logic. Boolean Logic, named after the Irish mathematician George Boole, uses three primary "operators" or words. These words are OR, AND, and NOT. Used with search terms, these operators can aid in narrowing or broadening your search. Refer to the handout from the University at Albany for more information.

Deep Web

There is much information that is not recorded by search engines available on the web. This is known as the "deep web"; publically available information that is not accessible through normal search engines. This can take the form of images, PDF files, and online databases. Some search engines such as Google allow you to try to search for deep web items but even so it is often difficult to find these resources. We'll take a look at a deep web search tool below. It is also recommended that you read the Deep Web Tutorial from the University of New York.

Helpful Links

Here's a collection of links to various sites that may be useful to you. Go through a few of these and consider the ways that you could change how you search online for the better.

[The Spider's Apprentice](http://www.monash.com/spidap.html) (<http://www.monash.com/spidap.html>)

This page offers a FAQ on how search engines work, as well as information on how to plan the best search strategy.

[Finding Information on the Internet](http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html)

(<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>)



This site by UC Berkeley has information on maximizing your Google search as well as a style sheet that explains how to cite Internet sources.

[Bare Bones 101](http://www.sc.edu/beaufort/library/pages/bones/bones.shtml) (<http://www.sc.edu/beaufort/library/pages/bones/bones.shtml>)

The University of South Carolina has created a fine, comprehensive set of lessons on how to maximize web searching. The site is designed as a class with separate lessons for each topic.

[Internet Tutorials](http://library.albany.edu/internet/) (<http://library.albany.edu/internet/>)

These pages were created by the University of New York. They utilize many graphics and charts to describe how web searching works and tips and tricks for maximizing searching. This site is highly recommended.

Search Engines

Google isn't the only way to conduct general web searches. Here's a short description of not only Google but four other search engines that use a slightly different approach.

[Google](http://www.google.com/) (<http://www.google.com/>)

As mentioned above, Google is the most popular search engine today. They have a fairly revolutionary way of sorting the results from searches. Google rates web sites according to what they call “popularity”—in other words, the amount of times that other sites link to the site in question. So if a site gets linked to frequently, it will show up higher on the ratings. This is a very effective system. If you are looking for information which is not very obscure, you will probably find it easiest here.

[Ask Jeeves](http://www.ask.com/) (<http://www.ask.com/>)

Ask Jeeves is an interesting site, because it allows you to form your search as a question. It also has a unique way of displaying results. When it has performed the search, it will bring up a screen with several different types of questions on it. Within each type, there are lists of possible questions. You can simply find the question that is nearest to the question you need answered, and then click on it. There is one significant problem with the Ask Jeeves site, however: it opens other webpages within a frame, so that it seems like you are still in the Ask Jeeves site. This also prevents you from bookmarking the resulting site.

[Yahoo!](http://www.yahoo.com/) (<http://www.yahoo.com/>)

Yahoo! is slightly different from other search engines. While you can search for specific keywords, Yahoo! also allows you to browse for web pages in certain categories and subcategories. This is useful when you know a general topic area but may be unsure of the best keywords to use.



[Infomine](http://infomine.ucr.edu/) (<http://infomine.ucr.edu/>)

Infomine is similar in concept to Yahoo! in that it breaks up web pages into various subjects. However, as it is geared towards scholarly research, the web pages in this directory are of generally higher quality than you would find on Yahoo!.

[Clusty](http://clusty.com/) (<http://clusty.com/>)

This new search engine categorizes results by placing them into dynamically created clusters. For instance, searching for "Jefferson" will create clusters such as Thomas, Jefferson County, Jefferson Airplane, and Hotels. Within each cluster you will find web pages that pertain to the cluster title.

TIP:

Remember that pages with dynamic content (job postings, online message boards, blogs, etc.) are not generally recorded by search engines. You will have to go to those specific sites and then search within the site for your search term.

Searching the WSU library

Our libraries here at WSU give you access to mountains of potential research material. There is an online tutorial located at Eastern Washington University. If you need more information, feel free to ask a reference librarian. You may also find it useful to take GenEd 300 (Accessing Information for Research), a one credit course designed to thoroughly introduce you to advanced research methods. Don't be afraid of the library! You will find information of a much higher quality using their online journals and printed materials than you will ever find on a general search engine.

Assignment: Search the Web

Your assignment this week is to try some advanced web searches and to report your results. Choose a topic of interest to you. Then, search for your topic online using at least three different search engines, with one of those being a non-conventional engine (such as accessing the "deep web" or using WSU's online journals). Report your results in one or two paragraphs in a Word document, noting differences and similarities. Now, using Google, do another search for your topic. This time, use at least three different concepts from today's class to see how the search results differ (for example, Boolean Logic, quotes, and wild cards). Report on your results in one to two paragraphs in the same Word document.