

Search Engines

Web search engines can provide fast starts on a research problem. In fact, in a few short years World Wide Web search engines have become the most popular tools by far for finding information because they are convenient, can be used without much skill, and frequently produce useful results. Remember these 5 important points about search engines:

1. *Search engines cannot see the contents of most academic databases.* They are only designed to find html and similar “pages.” They usually can’t retrieve the contents of databases such as library catalogs or **Arts Abstracts** or any of the nearly 400 databases that the University subscribes to. (There are a small number of exceptions to this rule. See for example **Google Scholar**, below.)

2. *Most web resources are self-published and vary widely in quality,* so you have more work to do in evaluating them than you would with other publications. Trade and academic publishers put effort into assuring the quality of their books and magazines in order to assure that they are profitable. Only a very small percentage of web sites attempt those standards. When you use web sites you are taking on extra responsibility for judging quality. Here are some points to consider when evaluating web sites:

Authority

- Who is the author of the piece?
- Is the author the original creator of the information?
- What is the author’s occupation, experience, position, education, or other credentials?

Affiliation

- What institution (company, organization, government, university, etc.) supports this information?
- Does the institution appear to exercise quality control over the information?
- Does the author’s affiliation with this particular institution appear to bias the information?

Currency

- When was the information created or last updated?

Purpose

- What appears to be the purpose for this information? To inform? Explain? Persuade?

Audience

- Who is the intended audience?

Compared to what?

- What does this work/site offer compared to other works, including non-internet works?
- Is the information in this site supported by other sources?

3. *Discourse communities can be difficult to identify on the Internet.* The faculty in your department expect you to develop some understanding of their discourse community (those people who contribute to the knowledge of a particular discipline or topic.) A source like **Artbibliographies Modern** focuses on the publications where artists, critics, art dealers, and art historians publicly contribute to their field of study. While good contributions can also be found on the Internet they can be difficult to identify there because of the mix of hobbyists, entrepreneurs, and others contributing.

4. *Using search engines well requires skills* that are very similar to the ones needed for searching databases. Because search engines scan millions of items, they almost always return some results even if very simple search statements are entered. But search engines usually have powerful features that are not invoked unless you specify them. Use the most unique terms that relate to your topic and learn how to search phrases – usually surrounded by quotes “like this.” Learning to read and shorten URLs (web addresses) is an important skill. Another is choosing the right search engine for the right job. One easy way to learn some of these tricks quickly is to use the guide posted by a non-profit group called Infopeople. Their **Search Tools Chart** (<http://www.infopeople.org/search/chart.html>) selects a small number of good search engines and web guides, explains what they are searching, and describes the search features of each. A more inclusive guide to search engines, and what jobs they are suited for, is **Noodle Tools’ Choose the Best...** (<http://www.noodletools.com/debbie/literacies/information/5locate/adviceengine.html>) Try opening the **Search Tools Chart** or **Noodle Tools** in one window and experiment with different search engines and techniques in

another.) Never settle for just one search. Experimentation is very important in web searching. For example, type any 3 unrelated search terms (frog metal wheat?) as a search statement in **Google** (<http://www.google.com>) then change the order of the three terms a few times and see how the results of the search change. Or try Googlehacking! (It's a sport. Look it up.) Trial-and-error learning is especially important with search engines since search engine companies they tend to be so secretive about their workings and features change overnight.

5. **Google is not the best choice for every task.** Some new search engines have specialties which can make them much more effective than **Google** for a particular need. Here are some examples related to academic research:

A9.com's Open Search. (<http://www.a9.com/-/home.jsp>)

This convenient "aggregator" or "meta-search engine" run by Amazon.com searches several sources simultaneously and returns the results in separate columns running down the screen. Featured sources at present are Wikipedia, a general search engine (Live.com), the Internet Movie Database, the Amazon book database, a blend of free reference sources, etc. You can also select from hundreds of others. The speed is impressive.

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

This search engine clusters the results into groups based on their similarity. So the hundreds of results from a term like "architecture" are grouped into categories such as: architects, schools, networking, software, etc. Searches can be focused on images, blogs, and other forms. A similar search engine called **KartOO** (<http://www.kartoo.com>) takes this clustering approach further by creating malleable animations of the clusters.

Google Book Search (<http://books.google.com>)

This is a very important project. **Google** has been working with large research libraries (like Penn State) to digitize books in their collections. **Google** has also been working with a number of publishers regarding the texts of their books. A large number of books has already been posted with Google-style searching of their complete texts. Usually you can read a passage from the book that contains your keywords. Often you can read all of the passages in the book that contain the words. Sometimes you can see every page of the book. Even though only a fraction of the planned books are completed, search results can be very impressive for some topics. Only a bit of this content seems to appear through regular **Google** searches.

Google Scholar (<http://scholar.google.com>)

Also important. Uses the technology of the **Google** search engine but tries to concentrate on reliable sources that meet scholarly expectations for quality. It seems to do this in two ways: 1) By focusing **Google** on the official postings of research organizations and university departments. 2) By taking advantage of a new protocol for making the contents of a few databases visible to **Google**. The databases selected for inclusion include a collection of most American library catalogs and also include a few of the databases of electronic journals that we subscribe to at Penn State. By connecting to these resources, **Google Scholar** can find an article in one of Penn State's electronic journals. However, it only can see a handful of the more than 400 database the library has and it is very weak in its coverage of the arts. Also, we have discovered that **Google Scholar**, still labeled a "beta" test after 3 years, is very incomplete. For example, sometimes it finds one article in an electronic journal, but not another in the same journal – even though both should be available. It is very useful, but don't trust it.

Marketleap Link Popularity Check (<http://www.marketleap.com/publinkpop>)

If you have found a web site that is of particular value to your research, you can use specialized search engine tools to determine what other web sites have linked to it. This one returns the results from some major search engines. Type in the URL of the site you are interested in and follow the instructions. The software creates a table of results with the results for your site in the top row. That row shows the number of linking web sites found on each of the search engines. Click on the numbers to see the actual list of web sites. Something similar may be done in **Google** ("link:[url]") but this only returns results from **Google**.