

## Web of Information

All about the Internet, histories, glossaries, legal issues, and more. See <http://www.isoc.org/internet/>.

## Evaluating Web Sites

Thousands of sources of information exist on the Internet – some of them reputable, others not. To find the information you need from a reputable site, consider the following questions:

- Is an author identified?
- What are the credentials of the author or organization?
- Is there a bias or a commercial interest?
- Who is the intended audience?
- How current is the information?
- Has the page been well maintained?
- Are references, citations, or links to other resources included?

More information about evaluating Web resources is available from Evaluating Web Pages: Techniques to Apply & Questions to Ask (<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Evaluate.html>).

**Hanford Technical Library** - <https://library.pnl.gov/index.stm>

- ❑ Electronic Resources - <https://library.pnl.gov/subjects.asp>
- ❑ Electronic Journals - <https://library.pnl.gov/journals.asp>
- ❑ Electronic Reports - <https://library.pnl.gov/resourcebysubject.asp?subjectid=56>
- ❑ Library Catalog – <http://libcat.pnl.gov>

## Search Engines

Use search engines to narrow information in topic areas, find specific sites or opinions about issues, politics, etc., and obtain a large number of results.

### Search Engines

- Search engines are tools called “spiders” that gather information automatically on the Internet. A **spider** is a computer program that travels the Internet to locate Web documents. It indexes the documents in a database, which is then searched using a search engine (such as AltaVista). A spider can also be referred to as a **robot**. Each search engine uses a spider to build its database.
- These search engines look for the presence of a user's search terms in the source document and rank the document accordingly.
- Search engines generally use term relevancy to rank results; however, each search engine determines relevance differently.

## First-Generation Search Engines

### Examples

All the Web – <http://www.alltheweb.com>

Cuil - <http://www.cuil.com/>

Altavisa – <http://www.altavista.com>

Gigablast – <http://www.gigablast.com/>

Bing - <http://www.bing.com/>

Hakia – <http://www.hakia.com>

### **Creating Search Statements**

- +/- are implied Boolean symbols
  - ❑ Use + to search all the words entered. **Example: +Hanford +tanks**
  - ❑ Use - to specifically exclude words or concepts. **Example: +Hanford -PNNL**
- Phrase searching
  - ❑ Use quotation marks (" ") to search for terms in the exact the order specified. **Example: "radioactive waste management"**
- An asterisk (\*) is usually used for truncation. **Example: fish\*** finds fish, fishes, fishing, etc. However, search engines differ. Consult the search engine's *Help* function to determine that engine's method for truncation.

Best Search Tools Chart: <http://www.infopeople.org/search/chart.html>

### **Second-Generation Search Engines**

#### **Examples:**

Google - <http://www.google.com/>

Ask – <http://www.ask.com>

A Google “toolbar” is available at <http://toolbar.google.com>. The toolbar is an add-on feature to Internet Explorer that allows you to search Google while viewing any Web page, search Google's non-Web page catalogs, and search the site you are currently viewing.

### **Ranking Schemes**

- These search engines offer improved results by using a variety of ranking schemes, such as ranking by links from other sites, popularity, concept, and domain.
- Google ranks relevant Web sites based on the link structure of the Internet itself. A page becomes highly ranked when other highly ranked sites link to it.

Try these tools to compare the overlap of results retrieved by popular search engines.

Thumbshots – <http://ranking.thumbshots.com>

TwinGine - <http://twingine.com/>

### **Metasearchers**

Metasearchers are Web sites that access several search engines from one place with one search command.

#### **Examples:**

BestSearch – <http://www.bestsearch.com/>

MetaCrawler – <http://www.metacrawler.com/>

Dogpile – <http://www.dogpile.com>

Search.com – <http://www.search.com/search>

Jux2 – <http://www.jux2.com/>

Searchzee – <http://www.searchzee.com/>

### **Second-Generation Metasearchers**

Clusty – <http://clusty.com/>. This search engine automatically organizes Internet search results into folders and transforms long lists of search results into logical categories for the user to view.

Kartoo – <http://www.kartoo.com>. This search engine provides a visual representation of search results, which allows you to gain additional insights about your Web search.

### **For Scientific Information Only:**

Scirus - <http://www.scirus.com/> Scirus is a science-specific search tool that not only searches web pages, but also abstracts of journals published by Elsevier, Scirus' parent company. To see if the Hanford Technical Library subscribes to a specific journal title, check our online catalog, Leona at: <http://libcat.pnl.gov>

## Directories and Portals

Directories and portals are human creations – these sites are chosen by people (e.g., by librarians or information professionals who carefully select and annotate each entry) rather than by computer programs.

Use directories and portals to search a broad area of topics; search selected, evaluated, and annotated collections; find quality sites over quantity; and locate Web sites rather than individual pages within Web sites.

### Examples:

About.com - <http://www.about.com/> Librarians' Internet Index - <http://www.lii.org/>

INFOMINE - <http://infomine.ucr.edu/> Yahoo! - <http://dir.yahoo.com/>

## Deep Web

The "Deep Web" or "Invisible Web" primarily refers to content contained in databases connected to the Internet. For example, the EDGAR Database (<http://www.sec.gov/edgar.shtml>) is a database of publicly available filings submitted to the Securities and Exchange Commission from January 1995 to the present. This content cannot be gathered by search spiders and is therefore not included in normal search engine results.

Use the Deep Web for information that is likely to be stored in a database or that is dynamically changing in content.

### Examples:

Complete Planet- <http://www.completeplanet.com/> IncyWincy - <http://www.incywincy.com/>  
DeepPeep - <http://www.deeppeep.org/> Intute - <http://www.intute.ac.uk/>

For more information on the Deep Web Sites, go to <http://www.internettutorials.net/deepweb.asp>

### General Search Tips

- Plan your search strategy before searching.
- When applicable, use more than one word. **Example: +Battelle +Richland.**
- Use more than one search engine.
- Use the search engine's Help function.

To view the latest information on search engines and useful Web sites, visit

- ❑ Alt Search Engines – <http://altsearchengines.com/>
- ❑ Internet Scout Project – <http://scout.cs.wisc.edu/>
- ❑ Search Engine Watch – <http://searchenginewatch.com/>

## Who produced the site?

.com = commercial site

.edu = academic site

.gov = government site

.org = typically a non-profit site

.net = networked service provider

.mil = a military site

## Country-/Region-Specific Domain Names

A list of domain names specific to countries and regions (e.g., .uk for United Kingdom) is available from the NORID Web site (<http://www.norid.no/domenenavnbasert/domreg.html>).

For more information on searching the Internet, contact the Hanford Technical Library at (509) 372-7430 or by email at <mailto:pnl.techlib@pnl.gov>. For information about advanced internet search techniques, see [https://library.pnl.gov/tipsheet/adv\\_internet.pdf](https://library.pnl.gov/tipsheet/adv_internet.pdf).