Box 7.3 Databases versus Search Engines

Databases and search engines have some distinct differences, although they may not be clear to student researchers. In fact, the distinctions may be blurring as the information keeps changing.

**Databases**

**Collects material:** Utilizes submission processes based on peer review, editorial boards, and other review processes

**Indexing:** Information and data are stored in a uniform way utilizing “fields.” Fields allow for the retrieval of specific pieces of information.

**Retrieval of Information:** Finds results by matching information listed in fields. All fields are searchable

**Scope:** Databases are organized with a special purpose or subject in mind or are intended for a specific audience or membership. All entries in the database relate to its subject, purpose, or some kind of specialization.

**Authority:** A submission process is in place to ensure authority

**Updating:** Every day new information can be added to a database easily.

---

**General-Purpose Search Engines**

**Collects material:** Uses a Web crawler or by submission

**Indexing:** Retrieval is facilitated across several formats

**Retrieval of Information:** Find results by searching for keywords found in text. (Relevancy determined by algorithm or similar such criteria.)

**Scope:** General-purpose search engines provide across-the-board coverage.

**Authority:** Search engine does not evaluate information or Web sites.

**Updating:** New Websites and Web content created every day. Search engines are always crawling for new content and revisit Web sites on a regular basis, although frequency varies from one search engine to another.

---

**Cost:** Free or fee-based access may depend upon membership in a professional association

**Examples:** America: History and Life (fee-based); Historical Abstracts (fee-based); Africabib (free)


---

**Cost:** Free access.

**Examples:** Google; Yahoo

attempt to provide access to Pol Pot's security police archives, which document a portion of the 1.7 million Cambodians who were killed in the genocide between the years 1975 and 1979. Google will list the URLs for the databases themselves, but it will not be able to search within their contents, for instance, to find an individual's name. Another example of a hidden database to which common search engines might not go is Northwestern University's collection of photographs on the Siege and Commune of Paris, 1870-1871. Doing a search for "siege Paris" would find the web page. However, if you were looking for a photograph of the Chateau de Meudon, contained in the database, searching for "chateau meudon" might not find the photographs contained in this site. To search within a database, you must fill out a search query form with your desired search terms. Searching through fields unique to that database, the internal search engine dynamically generates a list of results; a webpage specifically tailored to your query. These results exist only for the one time you use them. Once you close your browser window, those results no longer exist as a single site. The url is considered dynamic, not static or stable, as normal websites or even the journal articles in JSTOR. A dynamically generated or temporary url will have symbols like "?" or cgi in it. For instance searching in the website/collection of primary resources that make up *American Memory* for civil war hospitals yields 1,293 results with a url of http://memory.loc.gov/cgi-bin/query. This url cannot be entered into a browser to produce the same results at a later date. (You may find some of the referenced sources in this book have dynamically generated urls. To find the source, you may have to do an Internet search for its title instead of typing in the url.) Many times dynamically generated urls are very long. As technology changes, the ability to search within these kinds of databases may also change.9

Open Directory. http://dmoz.org (accessed September 6, 2011). Billed as the “most comprehensive human-edited directory of the Web,” Open Directory has the usual categories. It also has a number of links to directories in other languages.

INFOMINE. http://infomine.ucr.edu (accessed September 6, 2011). Directory created by librarians from a number of university libraries for students and faculty using the Internet for research. The search interface is a bit different from most, looking much like an online catalog with options to search in author, title, subject, description, and full text.


Yahoo. http://www.yahoo.com (accessed September 6, 2011). The original directory. It is vast, though not approaching the size of a search engine. When it comes to scholarship, there are good finds as well as lots of junk in Yahoo. This directory is not as selective as the previously listed ones, but its organization is good. Use your evaluation skills carefully.

General History Search Directories, Weblogographies, and Print Bibliographies

WWW-VL History Central Catalog. http://vlib.iuc.edu/history/index.html (accessed September 6, 2011). Part of the World Wide Web Virtual Library, a massive directory of sites focused on history and covering all eras and countries imaginable. There is no search feature included, so you must browse through the categories. Some marginal sites are included in this site’s attempt to be comprehensive. Still, a gem.


Box 7.2 Google

Habitually, many of us go straight to the search engine Google to perform our searches. In fact “google” has become a part of speech, for example, to google, googling, googled, and so on. According to Siva Vaidhyanathan, author of The Googlization of Everything: (And Why We Should Worry), “Google is on the verge of being indistinguishable from the web itself” (3). The Google search engine has revolutionized not only the way we find information, but how we expect to acquire that information—fast, free, and highly relevant. The technologies that Google has invented and uses, as well as the company’s practices can shape public conversation, as well as your personal views. Because Google collects and organizes information/websites, it has no editorial responsibility to seek and highly rank authoritative sources. News outlets, such as newspapers, wire services, and television news are expected to vet their stories and judge the credibility of the information they have gathered (33–35).

Google’s ranking is based upon criteria that includes popularity of a link, your past search strategies in google and the sites you have looked at, where you are located (Google knows where you are searching from and makes assessments as to what US residents want compared to their European counterparts), and how many other pages link to a particular page (140). Google often seems to defer to Wikipedia, whose articles often end up on the first page if not in the first five results (63). Google’s ranking has the potential to give false importance to information because of a website’s ranking within the results. To be fair, Google’s desire is to make information past and present, printed and digital, easily and accurately accessible to a wide audience. Google Books and Google Scholar, despite their failings (see chapters 3 and 4 of this book) are a step toward creating a digital library that allows us to do research in different ways. Many other search engines do have similar practices to Google. Because of this, you must use more than one search engine and creatively evaluate not only the accuracy of the information presented but also what information may be left out.

site also contains a number of primary source collections as well. In the same way, some of the chatter on the H-Net (http://www.h-net.org/) series of listserv groups cites cutting-edge research. What the Internet will not provide, generally, is access to published, refereed journal articles or books. While there are exceptions, usually access to such materials requires a paid subscription by your library. An exception is Google Scholar (http://scholar.google.com/), a separate section of the search engine Google, which does provide links to some full-text academic journals (see chapter 4 for a discussion). There are also some academic journals that do publish freely online and will not be indexed in a periodical index. As discussed above, the search engines will not search in a fee-restricted database or in an online journal. For a list of online history journals, some free, some subscription, see http://aac.sub.uni-goettingen.de/en/journals/history/ and http://www.doaj.org. Many universities are creating scholarly commons or institutional repositories that collect and make freely available via the web, the published and unpublished works of their scholars. This is done with the permission of both the scholars and the publications in which the works appeared (for an example, see University of Nebraska’s DigitalCommons, http://digitalcommons.unl.edu/disciplines.html). An Internet search may turn up these sources.

- Communication. Many scholars belong to a listserv, and some read blogs or create their own blogs. With a listserv, you subscribe to a list on which you can ask, view, and reply to questions about topics around a central theme, for instance, early modern history. Blogs are public, personal diaries in which anyone can participate. Finally, RSS (Really Simple Syndication) allows you to receive notifications, much like a personal newsletter, of new postings in selected blogs. More about blogs and listservs below.

- Syllabi. Sometimes online course syllabi can be useful in research. Professors create specialized guides for their students to use in a course. Be very careful to evaluate your source. Make sure that the person writing the document has some authority and is not a student completing an assignment. Be careful using websites by other students. Some will be of excellent quality, while others may be poorly constructed and researched. It is difficult to judge quality if you do not yet know much about the subject you are studying. Be cautious about using syllabi and the documents linked to them as sources. See the online American History survey History Matters

Box 7.1 How to Read a URL

Your browser “reads” a URL (uniform resource locator) or the address of a website, which is the location of the computer where the web page you want resides. Knowing how to read a URL is essential to understanding where you are on the Internet. See figure 7.1 for an example of a URL.

http://www.dohistory.org

http://www.dohistory.org/on_your_own/toolkit/writing.html

Figure 7.1. The parts of a URL (uniform resource locator). By stripping each of the file names, separated by a forward slash (/), you can discover the origin or hosting computer.

Using the diary of Martha Ballard, an eighteenth-century midwife, this website illustrates how the past can be pieced together using fragments of evidence. The first URL takes you to the main website, an organization called “dohistory” that provides materials to study Martha Ballard’s diary. The second URL above leads you to a guide on deciphering eighteenth-century handwriting. By reading the URL itself, you will see the name, or host, of the website you are visiting, the kind of site (domain) it is (.com, .edu, .org, etc.), and which particular part of the file you are looking for. You can actually use your knowledge about a URL’s structure to evaluate a website. Even though you know nothing about the content of the website, by peeling off the file names back to www.dohistory.org, you know that an organization, in this case the Center for History and the New Media Film Study Center at Harvard University cited at the bottom of the page, has made this website available on the Internet. That should prompt you to ask questions about the purpose and intent of the organization and the influence that might have on the content of the site.
The Information-Literate Historian

A Guide to Research for History Students

Second Edition

Jenny L. Presnell

Miami University (Ohio)

New York Oxford

OXFORD UNIVERSITY PRESS