

## FINDING INFORMATION ON THE NET: SEARCH ENGINES AND SUBJECT GUIDES

Finding useful and accurate information on the Net requires a certain amount of skill, as well as access to a variety of research and return engines. The first decision a searcher must make is to determine the nature of the inquiry. If, for example you are looking for particular pieces of information that likely reside on only one particular machine, such as the entrance requirements for a specific graduate school, or the location of articles published in the *Journal of Distance Education*, then you would begin your search at a multipurpose *search engine* (such as Google or Alta Vista). If, however, you were interested in more general information, such as what schools offer graduate education programs or a broad source of information about distance education, then your most effective search tool would be a hierarchical *subject guide* (such as Yahoo!).

In hierarchical directories, or subject guides (such as Yahoo!), you would scroll down through hierarchical menus until you arrive at the site you are looking for. In simple operational terms, subject guides are organized around directories, which are created and populated by humans who have reviewed the listed sites and determined the particular subject headings under which they are most usefully classified. The benefit of subject guides over search engines is that the searcher does not have to wade through hundreds of documents to find relevant information. Directories provide the searcher with a straightforward, hierarchical means of information retrieval on the Web. However, this is precisely the reason that subject guides are not used with the frequency that search engines are. While subject guides can save the searcher much time by selecting only certain Web sites for each category, this also means that many sites are excluded. It is not possible for any group of people to accurately and continuously categorize and re-categorize all information available on the Net and accurately assign it to all possible relevant categories. As such, subject guides are not bias-free, because someone has made a decision about which resources will be included in the guide and under what categories.

On the other hand, search engines (such as Google, Alta Vista, Infoseek, HotBot, Excite, Lycos) use each word in the document and all meta tags contained in the page's introduction to create very large databases. These databases are developed and maintained automatically by software robots commonly known as *search bots* or just *bots*. The search algorithmically links your request with the titles and phrases found on the Web sites and stored in the databases. In simple operational terms, search engines work with robots that index the contents of nearly every document on the publicly available Web. Then the contents are entered into databases on very large and fast search engines. When you enter a query at a search engine Web site, your input is checked against the search engine's keyword indices. The best matches are then returned to you as hits. Most search engines use search-term frequency as a primary way of determining the order in which the hits are organized, by listing first the document with the most uses of the keyword. If, for example, the word or one of the words in your query appears numerous times in a Web document, it is reasonable to assume that the document will likely turn up near the beginning of the search engine's list. Some search engines are designed to search for both the frequency and the positioning of keywords to determine relevancy, reasoning that if the keywords appear early in the document or in the headers, the document is more likely to be on target.

Finally, some of the better search engines use a technique (such as that developed by Google.com) that tracks which sites are actually visited from the hundreds that are returned as hits and places these "consumer choice" sites higher on the list with each subsequent search. Thus, these systems become more accurate and useful with use and are especially useful for finding often-searched sites and terms. The search bots constantly update the indexes, usually visiting popular sites every few weeks. It should also be noted that some search engines index every word on every page, while others index only part of the document, such as the title, headings, subheadings, hyperlinks to other sites, and the first twenty lines of text. Further, search engines that have full-text indexing systems claim to pick up every word in the text except commonly occurring stop words such as *a*, *an*, *the*, *is*, *and*, and *or*, and some search engines discriminate uppercase from lowercase, whereas others store all words without reference to capitalization.

At this point, it should be noted that the distinction between a search engine and a subject guide is becoming blurred, given that most search sites now offer both search options-as they try to provide everything to everyone as one-stop search portals. In spite of this new trend, the basic operational difference between search engines and subject guides remains. Specifically, search engines use robots to search for, and record, as many Web sites as possible. Whereas subject guides tend to wait for Web pages to be submitted by the author, which are then assessed and placed in the appropriate hierarchical subject category.

A search engine, rather than a subject guide, is usually used to find particular information on the Net. However, because search engines index almost every document on the Web, retrieving what you are looking for often results in an onerous activity of selecting from hundreds, thousands, or even millions of Web pages that are returned from search queries. Wading through these Web pages can consume a tremendous amount of time. To get relevant and useful results, it is important to make use of the search engine's advanced search features including quotation marks and Boolean operators. For example, typing the phrase *online research* into Alta Vista produces about 2,660,000 potential sites with either the word *research* or the word *online* mentioned in the documents. We can reduce this number substantially by refining the search to documents in which the words *online* and *research* are both found, by joining the two words with the Boolean operator *and*. For example, typing *online and research* now produces 138,000 successful hits-though, still not few enough to be useful. Fortunately, we can further refine the search by placing the words in quotation marks, thus instructing the search engine to select only pages where the word *online* precedes the word *research*. Now the search for *online research* produces only 62,000 hits but still too many Web sites to be useful. At this point we can continue adding *and* words to further reduce our search and hone in on pages that are particularly interesting. For example adding the words *and surveys* reduces the number of hits to 643; while the words *online research and focus group and education* produce a manageable 128 hits.

Relevancy ranking is critical for useful retrieval of information, and becomes more so as the number of Web documents grows. Most of us do not have the time to sort through the possible millions of hits to determine which Web documents are relevant and useful. Obviously, the more relevant and useful the results are, the more we are likely to value the search engine and, in turn, regard the Web as a useful information resource. In particular, when you are not getting the desired results from a search engine, you may have better luck using a subject guide, as subject guides provide the option to refine your search based on a particular topic. For example, searching for e-research within the field of higher education, a subject guide returns only pages about e-research in higher education, not e-research in elementary, middle school, business, or any other field. Searching within a hierarchical category of interest allows you to quickly narrow in on only the most relevant pages.

Overall, the best strategy for retrieving useful and relevant information is to use the same tactics that you would use in a library search. First, begin by analyzing your needs. What are you looking for? If the topic is very specific, search engines such as Google will likely result in hits that are relevant and useful. Alternatively, if you are looking for a broad topic, a subject guide such as Yahoo! would likely return the most successful results. Next, conceptualize your search question and then isolate the keywords in the question by eliminating words that are irrelevant.

You may find it useful to use a special purpose search tool that can most accurately meet your needs. Many search tools are designed with specific aims in mind and, as such, there is no best search tool. Rather, which search tool is best will depend on the kind of search you are doing. The following are a few examples of the different kinds of search tools available with specific aims.

- MedNet (<http://www.mednets.com/>) is devoted exclusively to medical information
- Kids Search Tools (<http://www.rcls.org/ksearch.htm>) has been developed exclusively for children's searches, to be used by children
- Search Engine Colossus (<http://www.searchenginecolossus.com/>) provides search engines focused on particular regions and countries.

- If you are not sure which search engine to use, or even what search engines are available, NoodleQuest (<http://www.noodletools.com/noodlequest/>) is a Web site with an automated Web form that will generate a list of appropriate search engines based on both your Internet skills and your search needs
- Searchability (<http://www.searchability.com/>) and NeuvaNet (<http://www.noodletools.com/noodlequest/>) are Web sites that provide listings of specialty search engines with advice on how to choose the search engine most appropriate for your needs. For example, if you are looking for a few good hits fast, NeuvaNet recommends the use of Google (<http://www.google.com/>), Vivisimo (<http://vivisimo.com/formi form=Advanced>), and Ixquick (<http://ixquick.com/>).
- If you are looking for a general and broad academic subject and need to focus it, Neuva Net recommends search guides such as:
  - Encarta Online (<http://www.encarta.msn.com/reference/>)
  - Encyclopaedia Britannica (<http://www.britannica.com/>)
  - Northern Light (<http://www.northernlight.com/search.html>)
  - Librarians' Index to the Internet (<http://lii.org/>), or Infomine (<http://infomine.ucr.edu/>)
- If you are looking for biographical information, try using Lives (<http://amillionlives.com/>), Biography.com (<http://www.biography.com/search/>), or Biographical Dictionary (<http://www.s9.com/biography/>).
- There is rapid progress being made in the cataloging and retrieval (through meta tag descriptors) of graphic images that you can use to enhance a research presentation. Two of the largest collections of graphic images and pictures are available at <http://ditto.com/> and <http://www.altavista.com/sites/search/simage>.
- Sound and music files are also difficult to find due to problems in classification. Moodlogic ([www.moodlogic.com](http://www.moodlogic.com)) creates search applications that allow you to