Successful Search Strategies
(Published in the TechEdge, November 2007)
by Wesley A. Fryer

Maps from the the 16th Century are intriguing to examine not only for the land forms they outlined, but also for the parts of the globe they failed to map. Like the cartographers of ancient days, modern search engine creators are attempting to develop tools which assist human navigation of vast, unknown territories in cyberspace. The internet is composed (by the best estimates) of over 500,000 websites, containing between 35 and 60 million webpages. Growing at a rate of 20% per month, by the end of 1998 the web will contain more words than is presently contained in the entire US Library of Congress (Steinberg, "Seek and Ye Shall Find (Maybe)," Wired, May 1996). Even the largest search engines are unable to comprehensively index the internet. Searching for and finding information on the web can be a daunting task, but is none the less an essential skill in our modern age.

Readers should be familiar with basic navigation of internet websites and comfortable using the basic features of a web browser before learning the skills described in this article.

This article describes search strategies which can be employed to effectively meet the directives of the Texas Essential Knowledge and Skills for Technology. Beginning in grades K-2, the TEKS for Technology require students to "apply keyword searches to acquire information" and " select appropriate strategies to navigate and access information for research and resource sharing" (TEKS Chapter 126.2.B.4). Because the internet is not mentioned by name in the TEKS for these primary grades, searchable CD-ROM encyclopedias or other searchable computer media could be used to fulfill this "information acquisition" requirement.

"Like the cartographers of ancient days, modern search engine creators are attempting to develop tools which assist human navigation of vast, unknown territories in cyberspace."

In grades 3-5, however, students must use "keyword and Boolean search strategies" on networks "including the Internet...for research and resource sharing" (TEKS Chapter 126.3.B.4). The same search strategy requirements are

also part of the Technology TEKS for middle and high school students (TEKS Chapters 126.12.B.4 and 126.22.B.4). To effectively teach students to locate reliable and relevant information on the web, educators need to understand some basic concepts about internet search engines as well as the need for teacher supervision.

"Boolean" searches are named for George Boole (1815-65), a British mathematician who developed a system of algebraic logic that has been applied beneficially in various areas, including computer logic circuits and software applications.

SUPERVISION REQUIRED

Whenever students are accessing the web, they should be supervised by a certified teacher present in the classroom. Even if filtering software is being used to "block" potentially offensive internet websites, the potential liability risk of students accessing objectionable sites justifies constant supervision at all levels. Students (and their parents) should also complete liability release forms before accessing online resources. This is especially important before students' first names, school work, and/or photographs are published on a school website. (An example waiver form is available for download at http://www.wtvi.com/teks/tools.html.)

SHORTCOMINGS OF SEARCH ENGINES

If you have ever searched for a topic on the internet, you know that many of the websites referenced by a search engine can be disappointing. The primary reasons for this disappointment are the search engines' general inability to account for synonyms and homonyms in the English language. It is this shortcoming which causes a simple keyword search about "space" to reference webpages about "the music space on the web," the movie "Space Jam," and VRML (Virtual Reality Markup Language.) The same search might not reference pages about the MIR space station or Cape Canaveral, Florida.

Educators and their students should use a variety of search engines when attempting to locate internet information. It is important to become familiar with the search syntax of different engines so that advanced searching techniques can be effectively employed.

An important attribute of the internet is that it enables virtually anyone to publish his/her ideas for a global audience. This fact often makes it difficult to assess the validity of information contained on a webpage. Discerning truth is also difficult when a webpage is not located contextually. For example, a webpage of tax information linked to from the IRS website would be contextually located. On the other hand, a keyword search for "freedom" could yield webpages from the US Veterans Administration as well as US militia groups. Webpage "context" might not be readily apparent or verifiable. Searching the web clearly highlights the need to teach students critical thinking skills, but it also demonstrates the need to utilize advanced search techniques which pinpoint desired information as accurately as possible.

Although they are struggling to do so, no search engine has comprehensively indexed the entire internet. It is impossible to objectively declare one search engine the "best of the web," although comparative contests are frequently held between them. Educators and their students should therefore use a variety of search engines when attempting to locate internet information. It is important to become familiar with the search syntax of different engines, however, so that advanced searching techniques can be effectively employed. In this article, I will address the syntax requirements for three search engines which have differing strengths: Yahoo, AltaVista, and Excite.

GENERAL SEARCH GUIDELINES

Regardless of the search engine used to locate information, try to use unique words that specifically describe the type of information you are seeking. Avoid common, general terms. Use all lowercase characters, unless the word is a proper name which must be capitalized. Whenever possible, employ "phrase searching" (searching for more than one keyword) or a Boolean search strategy (described later), to further refine your search. If you link to a webpage buried within a website, "truncate" the web address in the location window at the top of your browser to view the main webpage. For example, if you link to http://www.wtvi.com/teks/tools.html, to truncate the web address, delete the filename and directory name so you can link to http://www.wtvi.com/. In this way, you can check the source of the webpage, which can possibly affirm the authority and reliability of the webpage's information. Lastly, remember to use several search engines since results from each will vary.

BROWSING FOR WEBSITES

If you know the general topic, it may be best to use a "directory site" to browse the internet. One of the most popular and well known directory sites is Yahoo (http://www.yahoo.com). Websites in Yahoo are classified into discrete categories which help provide context for the websites you can find in its index. This categorization process is done by Yahoo's employees, which makes the process very labor intensive. Almost all other search engines employ a computer algorithm to index websites. This is much faster, but cannot provide an equivalent, human-input, contextual organization.

Table C1 : Search Syntax for Yahoo

Main URL Address :	http://www.yahoo.com
Advanced Search Options Address :	http://search.yahoo.com/bin/search/options
Phrase Searching :	Enclose phrases in double quotes (Example : "space travel")
Special Characters :	Use + before a word that must appear, Use - before a word which must NOT appear
Search Limits :	To search within a title, use t:word (t:whales) To search within a URL, use u:word (u:kansas)
Wild cards :	Use "*" to specify a root word and retrieve various word forms (sing* will retrieve sing, sings, singing, singer, as well as singe, singular, etc.

Yahoo has over 200,000 websites classified into approximately 20,000 different categories. A keyword search in Yahoo provides related subject categories as well as individual websites. Yahoo tries to categorize the main webpage of an organization or group, and attempts to avoid search results which include more than one webpage from the same site. The size of the Yahoo database is relatively small, however. For this reason, it is useful for browsing websites, but not as useful for searches requiring advanced Boolean syntax. (Refer to table C1 for a description of the syntax required by Yahoo as well as the advanced search page URL address, which permits AND/OR phrase searching.)

ADVANCED BOOLEAN SEARCHES

Beginning in the 1998-99 academic year when implementation of the TEKS is required, third grade through high school students in Texas will be required to use Boolean search strategies to locate information on the internet. A Boolean search is similar to an algebraic statement of equivalency, but is composed of words (instead of numbers) and conjunctions (instead of operations like addition and subtraction). A Boolean search for information about Allied

aircraft in the Gulf War might search for webpages containing the words "aircraft" and "Persian Gulf," but not "Iraq."

Other Search Engines to Try:

http://www.infoseek.com

http://www.opentext.com

http://lycos.cs.cmu.edu/

http://www.webcrawler.com/

http://www.mckinley.com/

http://www.stpt.com/

http://www.search.com/

http://www.dogpile.com/

Search engines categorize websites by different information contained in webpages. The most comprehensive engines index sites' "full-text," which means they catalog websites by their URL address, title, and all the text contained in them. Three of the most popular full-text search engines are AltaVista, InfoSeek, and Open Text. Of these three search engines, AltaVista is the most comprehensive (presently containing over 60 gigabytes of webpage information) and allows "nested Boolean" searches, desirable for advanced internet queries (Nested Boolean searches use parentheses to separate components of the query). AltaVista is the only search engine (presently) which allows searches within one of twenty-five different languages. Additionally, you can submit a search to AltaVista and then choose to "refine" your search to further narrow the results.

A simple keyword search in AltaVista can frequently yield results in excess of one million different websites (My simple keyword search on "space" yielded 4,113,780 webpages.) The AltaVista simple search does not support Boolean operators, however, although it permits phrase searching (like Yahoo, by enclosing phrases in double quotation marks). Due to the enormous size of the AltaVista database, it is best to use the advanced search located at http://www.altavista.digital.com/cgi-bin/query?pg=aq. The Boolean syntax for AltaVista, search limits, and the "refine" option (described in Table C2) permit very specific searches which can efficiently reduce the number of search result documents.

Table C2: Search Syntax for AltaVista	
Main URL Address:	http://www.altavista.digital.com/
Advanced Search Options Address:	http://www.altavista.digital.com/cgi-bin/query?pg=aq
Phrase Searching:	Enclose phrases in double quotes (Example: "space travel")
Special Characters: (simple search ONLY)	Use + before a word that must appear, Use - before a word which must NOT appear
Search Limits: (advanced search)	You can limit results to one of over 20 languages, search the web or Usenet News (Internet-wide discussion groups), as well as inclusive dates webpages were last updated.
Wild cards:	Use "*" to specify a root word and retrieve various word forms
Boolean operators:	Use AND, OR, NOT, and NEAR to specify the relationship between query terms. (Words can be upper or lower case.) Nest operators with parentheses.
"Refine" search:	Choose "refine" after performing either a simple or advanced search to include or exclude specific words which are present in your search results (listed by percentage of frequency)

A thorough Boolean search in AltaVista should include wild card operators (an asterisk) to display alternative word forms, parentheses to "nest" query terms, and "operators" including **AND**, **OR**, **NOT**, or **NEAR** which specify the relationship of the query words. An example search for grant sources for educational technology in AltaVista might look like:

(grant* or fund*) and education* and (technolog* or comput*)

Performed in AltaVista, this search retrieved 246,463 documents that matched at least some of the terms. Since these results needed further refinement, I selected the "refine" button. From the resulting pop-up menu, I chose to:

- require the words "grant" and "education"
- exclude the words "college" and "university"

After specifying these refinements, I performed the search again and was presented with only 91 documents. The first five results were highly relevant websites, including the grant writing guidelines and application information I wanted to find.

AltaVista provides an excellent online tutorial for learning to use its advanced search syntax. The advanced tutorial is located athttp://www.altavista.digital.com/av/content/help_advanced.htm. Practice using the

various Boolean operators when performing searches and try to narrow (by refining your query) search results to reasonable numbers of documents.

Table C3: Search Syntax for Excite	
Main URL Address:	http://www.excite.com
"Power Search" Address:	http://www.excite.com/search/options.html
Phrase Searching: (simple search)	Enclose phrases in double quotes (Example: "space travel")
Special Characters: (Power Search)	Do NOT use special characters like quotation marks, +, -, or operators like and, or, etc. in the power search.
Search Limits:	Include or exclude specific words or phrases by selecting pull down menu choices and typing in corresponding form fields. Search the web, current news, or Excite's international databases in Europe.
Boolean Searching:	Boolean operators (and, or, etc.) CAN be used in the normal Excite search window. Nested Boolean searches are NOT permitted. An Excite "power search" allows Boolean searching without the use of advanced query syntax.

Another search engine which should be utilized when teaching students to locate information on the internet is Excite. Unlike many other search engines, the algorithm employed by Excite performs a statistical analysis on webpages to create classification categories based upon words that occur together frequently. The result is that Excite searches seem to produce more relevant results than many other engines, because it finds webpages with related concepts.

One type of advanced search in the Excite database is a "power search." Basically, a power search enables complex Boolean queries to be composed by novices or research experts without advanced query syntax. As described in Table C3, specific words or phrases can be included or excluded, and the webspace to be searched can be limited. Alternatively, the normal Excite search window can be used with Boolean operators. Nested Boolean searches (which use parentheses) cannot be performed by Excite, however.

Excite provides a limited directory of sites (called "channels") which can be browsed, similar to those of Yahoo. This human-indexed directory contains around 80,000 websites at present. One pedagogical benefit of browsing websites in a directory site is that students become familiar with hierarchical organization. To find a review of their favorite movies, for example, they must look in entertainment, then in movies, then in current releases, etc. Teachers

can help students mentally transfer the hierarchical organization methods of website directories to other academic applications, such as those used in science and social studies.

SEARCH WITHIN A DOCUMENT

Once you have located a webpage that appears to be relevant to the inquiry, it is easy to search within the webpage for one or more terms in the text. From the "Edit" menu in Netscape, select "Find." Enter the word you want to search for on the webpage, and Netscape will highlight the first occurrence of the word(s). Students can then use the skills described in the August 1997 TechEdge article,"Web Browser Skills for Research," to acquire textual, graphical, and other information from the webpage to use elsewhere.

EXPLORATION WITHOUT END

The cartographers of history may have made incomplete maps, but they had a reasonable expectation that with continued exploration, the uncharted spaces would eventually be filled with the outlines of continents and oceans. In cyberspace, however, the frontier of information appears to be perpetually expanding. It can be our joy, as well as our task as educators, to not only efficiently navigate through the internet's vast expanse of information, but also meaningfully contribute to its contents. We are explorers of new lands in an revolutionary age. Good luck in your search for knowledge!

A linked list of online references used in the preparation of this article is available at http://www.wtvi.com/teks. Additionally, links to all the search engines mentioned in the article are provided so you only have to type in one URL: the address of this column's website!

Wesley Fryer is an elementary educator and internet consultant in Lubbock, Texas. He welcomes your questions and comments about this article or about classroom technology integration in general!

--
This content is licensed by Wesley Fryer under a Creative Commons Attribution 4.0 International license.
http://creativecommons.org/licenses/by/4.0/deed.en_US
http://www.wesfryer.com
Republished in January 2014 on http://publications.wesfryer.com