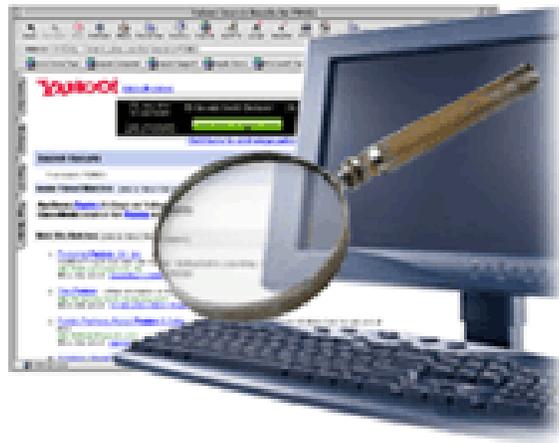# The Quest for Information: A Guide to Searching the Internet

James Day, DDS, MEd

## Abstract

Searching the Internet effectively necessitates the use of contemporary software programs commonly known as "search engines" to locate desired and meaningful information. Knowing how to effectively and efficiently utilize such search engines is critical for effective and efficient knowledge-seeking activities by the practitioner. Three categories of search engines are described in this article along with an explanation of strategies to maximize the efficiency of searching by appropriately using keywords and supplemental search operators. Examples of major search engines are included in the discussion.

**Keywords:** Boolean searching, searching, search engines

**Citation:** Day J. The Quest for Information: A Guide to Searching the Internet. J Contemp Dent Pract 2001 Nov;(2)4: 033-043.

## Introduction

In the real world, finding something as common as your favorite "French Roast" coffee hardly ranks as much of a challenge. Just head for your local supermarket or Starbucks and scan the shelves for the package or bin with the appro- priate "label." Your choices are really quite straightforward and simple. Just grab the package and you're done.

However, looking for that special item online can be an altogether different experience. Searching for the answer is akin to the search for the Holy Grail without the map. You know it's out there, you just can't find it in the over 1 billion published web pages currently online. After attempting a search on the Internet, responses can come back that do not match your intent, while others are obscured in a jumble of undesirable links that aren't what you really want. Truly, this is the proverbial "needle in a haystack" experience.

To overcome the problems of "searching" the Internet, learn about the variety of software tools that can help you. The correct tool with the appropriate strategy can help you "surf" the web to full advantage. The goal of this article is pro- vide an understanding of the basic categories of search software on the web and suggest approaches that use these resources in a produc- tive, efficient, and meaningful way.

## Inner Workings

The web is an example of client/server computing with both computers sharing the work in perform- ing a task. The server transmits the document. The client software displays it. You use a client program (e.g., Internet Explorer or Netscape Navigator) to retrieve information from this server computer. Your Internet Service Provider typically represents the server portion of the equation.

Given the context of scanning the Internet for useful information, a search engine is a program, or series of programs that scan and index web pages on the Internet at the server level. In the context of the World Wide Web, this means the search engine program searches the "HTML lan- guage" of the Internet contained in documents gathered by another software program known as a "robot" and then the client software (browser) displays it.

Based on the latest statistics from Zona Research in Redwood, Ca. (2000), the search engine was used 77% of the time by people looking for information. That makes the search site the #1 tool for infor- mation retrieval. And with over a billion pages on the web, information appearing on the top ten in one engine does not necessarily endorse a consistent score or ranking. For example, for a selected topic, sites ranked at the top in Altavista.com are not necessarily the same as those found at the top in Google.com.[1]

Traditional search software retrieves pages of software and then creates indices using tools like a thesaurus. When one enters the word "coffee," the results include other words like "latte." Often search engines require the user to try keyword after keyword to find what they are look- ing for. Some allow you to add modifiers to this unique word. It is preferable to make the search engines more sophisticated rather than making the searcher more sophisticated.

Some search engine and directory sites sell placement ranks within the results themselves (i.e., one can buy a higher page rank). Consider this as a factor when you look for an easy, hon- est, and objective way to find high-quality web- sites with information relevant to your search.

## Search Engine Types

For the most part, Internet search tools fall into two main groups.[2] The online directory is typified by "Yahoo" and "Lycos." A second category, the search engine, is represented by other web sites such as "Hotbot" and "Alta Vista." The difference between these two is the strategy they use to compile their site listings. However, a third group of the search utilities are hybrids of these two

approaches and combine the two into a comprehensive package. The idea behind the hybrid search engine is to offer "the best of both worlds."

**Directory-Based Search Services**
For directory-based search services, the primary frame of reference is the subject matter. The site listings are compiled and reviewed manually. For example, Yahoo, the best known Internet Directory, dedicates staff to review and categorize site suggestions and then adds them to a specific directory on Yahoo. The directory structure is hierarchical and starts with a general subject heading such as dentistry. Successive sub headings are more specific. For example under Dentistry topical areas such as dental implants, fluoridation, amalgam, organizations, orthodontics, and tooth whitening are included. These databases are comparatively small and the frequency of the updating is relatively low.

Examples of directory-based search services include:

Yahoo!
http://www.yahoo.com/
You can find almost anything at Yahoo. This is an excellent hierarchical index that makes navigation simple.

InfoSeek
http://www.infoseek.com/
This is another large and popular subject index.

**The Search Engine**
In contrast, the unmanned Search Engine completely automates the process of indexing the sites and totally removes the human component. A software robot called a spider or crawler gathers sites from across the web as it scans pages and connects to associated links. One particular advantage is that the spider will automatically return to the same site periodically to check for new content or new pages.

The results from this "spidering" are then saved in the engines index and serve as the basis to orient each query. Given the automation process and the size of the Internet, these indices grow to upwards of 250 to 500 million pages. These efficiencies enable the search engine to cover a wide variety and number of sites that are maintained current through regular visits by the robot. The

information may not be as exacting and the quantities may be voluminous compared to a directory-based site.

Use this type of resource to obtain very specific information that is current. Robot sites also offer current linkage to other sites as referenced in the returned information.

Examples of unmanned search engines include:

Google
http://google.com
The mother of all search engines that puts the world of the Internet at your fingertips. Results are uncanny in their accuracy and placement in rank is not altered through purchase of priority rankings.[3,4]

Alta Vista
http://altavista.digital.com/
Alta Vista is a huge, fast web index provided as a service to the Internet community by Digital Equipment Corporation. It indexes more than 30 million web pages and 14,000 Usenet news groups.

**The Hybrid: Directory With Search Engine**
The other search engines are really modifications or hybrid models of the directory and the search engine models. The directory with search engine uses both the subject and keyword search for its process. The tool initially follows a directory path for the keyword search. It progressively narrows the search field with repeated visits to a selected site. This approach works well if the searcher is uncertain of either the subject or the appropriate keywords to use.

An example of a directory with a search engine site is:

Excite
http://www.excite.com/
Excite combines keyword searching and subject browsing in a unique and powerful way.

**Multi-Engine Searching**
Another approach to searching the vast Internet is the use of the multi-engine search, or meta-search engine that harnesses a number of search engines simultaneously. A common or natural language request is electronically and transpar-

ently distributed to multiple search engines, each directed to find information the searcher requested. The relevant hits are integrated into a single list that returns the abundance harvested from the many. This methodology allows the user to cover a great deal of material in a very efficient way. The protocol is also tolerant of imprecise search questions or keywords.

An example of a multi-engine site is:

MetaCrawler
http://www.metacrawler.com/
MetaCrawler refers your search to seven of the most popular search engines (including Alta Vista and Lycos), then compiles and ranks the results for you.

### Searching the Internet

As search engines improve, are we headed toward an age of intelligent machines? Although computers have not yet developed the power to mimic the human thought process, is it possible for them to understand the context behind an inquiry? For example, consider the combination of the words "tiger" and "woods." This association clearly means one thing to a naturalist and quite another to a golf fan. Can computers use this type of association and associative contexts to supply meaningful replies to one's queries on the Internet?

Contextual hints to specific meaning are now often included in advanced applications of artificial intelligence to searching software. For example, if someone types in the word "survivor" in the health section of an indexed search engine, they are probably not looking for information on the television show, "Survivor." Such linkage of topics within a specific domain based on context illustrates how computers are becoming "intelligent."

Technology is able to filter out the irrelevant links and exclude sites that don't fit a certain profile. The software notes the placement and pattern of words and then sends a "crawler" software agent into the specified pages on the web that matches the specified criteria. This strategy also works when the search software considers the

searcher's last "stop" on the web. Chances are, if you're looking for health information, the previous URL in your browser was a health related site. The software of today is able to use this context in its criteria for the search parameters.[5]

Keyword searching is a good strategy when you have a specific word or phrase in mind; for example, the name of an organization or product. Browsing in subject indexes may be the best strategy if you are still figuring out what information you need.

### Search Strategies

What can you search for? Nearly everything! It would be improper to state that the Internet contains the sum total of all human knowledge, however, it is rapidly moving in that direction.

As with any cognitive processes, the most efficient and productive approach is to use strategies that best suit you and the subject matter. The simpler you can keep it, the better the results. As a guide, the following suggestions start with the most simplistic and efficient plan and progresses to more complex techniques. Start using the easiest approach on the right site.

To begin, locate one of the search engines mentioned previously. At the site of a search engine, your search begins on a web page displayed within your browser window. On the web page there is a field in which you can enter the topic of interest to you. Most of the sites have a Help section or Frequently Asked Questions (FAQ) clearly marked. If you have difficulty, either use the "back" arrow at the top left of the browser tool bar or try the help utility for the site.

You begin your search by entering the "simple" search criteria into an editable field on your screen. Then press the search button.

### Simple Search Pointers

Searching the Internet for some particular types of information can be both a frustrating experience and a rewarding one. Start looking for what you need with a particular search engine or directory. Remember there can be considerable overlap between the con-

- BE SPECIFIC
- USE NOUNS AS QUERY WORDS
- PLACE MAIN TOPIC WORDS FIRST
- USE 6-8 KEYWORDS
- TRUNCATE YOUR TERM WITH THE * TO ADD PLURALS (e.g., DENTIST* = DENTIST + DENTISTS)

tents of one engine and another, so you will find repeated references among them.

Let's say we are looking for information on a supplier of parts for a 1961 Austin Healy Mk1 3000 BT7. What keyword do you use for a search engine like Yahoo? Yahoo allows only one keyword or a phrase in their advanced search options.

Let's start by looking for "Mk1 3000 BT7." With the proliferation of the information on the web and the increased efficiency and accuracy of the search engines, you might find it. But the selection may be way too specific. If not enough information was listed under Mk1 3000 BT7, widen your search by looking for "Austin Healy." Here you may find several classic car dealers or even the parts supplier you need. You may also find someone's Homepage where they write about owning a Healy or you may find that the Rock group "Austin Healy" makes up the bulk of your return search items. In this case the more specific information returned the appropriate responses.

Possible steps to finding your desired information using Yahoo and Google search tools:

1. **Start specific**
   (e.g., Search for " Mk1 3000 BT7");
   **Results:** 56 site links using "Yahoo" while 97 site links were found using "Google" with one key word being used.
2. **Broaden your search for more references**
   (e.g., Search for "1961 Austin Healy 3000");
   **Results:** 403 site links using "Google" and 274 using "Yahoo"
3. **Broaden your search even more**
   (e.g., Search for "Austin Healy");
   **Results**: 25,200 site links using "Google" and 15,300 using "Yahoo"
4. **If you still don't find what you need, broaden your search even further**
   (e.g., Search for " Classic Cars");
   **Results:** 746,000 site links using "Google" and 940 using "Yahoo"

It should be noted that the search results may vary with each attempt so that the results stated above are approximate. The results you receive

for any search request can vary depending on the site you select and when you enter a search term within a site. In fact, the Web is so dynamic the results from the same site can vary by the minute. This vigor adds the incredible advantage of instantly distributing knowledge that is time and context dependent. The user of an online search modality should understand there is no "correct" answer to any particular query. The relevance of each response is determined by the user.

### Advanced Search Techniques

Searching for highly specific things like "Mk1 3000 BT7" is easy for most people. A problem arises, however, when you need to search for something specific and it requires more than one word and combines context.

Most search engines and directories provide for advanced searching and their methods vary from system to system. Two types of advanced searches are usually supported: Phrase searches and Boolean searches. Each of these types of searches has strengths and weaknesses. Many search engines allow you to choose application based rules within each process. Let's review a hypothetical case to determine how the two techniques differ.

### Phrase Searching

Perhaps the easiest of the advanced techniques, phrase searching, allows you to search on multiple words for one topic. For example, searching for "clinical information in placement of a dental implant" is more meaningful than looking under the single word "implant."

Most search engines will do a phrase search without needing to inform the search engine the words you are looking for need to be grouped together. The software assumes that all the words have to be present on the page and in close proximity to one another.

### Boolean Searching

Named after an English mathematician, Boolean searching refers to a form of logic applied to the search. Basically a Boolean search requires some additional words to be used, for example searching on the words "Cancer" and "treatment." This type of search allows you to exclude websites that may be about Cancer, the constellation,

or the horoscope sign by combining the two concepts in one search.

Most web site searches allow you to simply "click " on the parameters you wish the search to employ. For instance, clicking on "any" equals the Boolean "OR."  Selecting "all" is the same as choosing "AND."  The following are definitions for these search parameters:

•Selecting "AND" = Search on Term1 AND Term2
•Selecting "OR" = Search on Term1 OR Term2
•Selecting "NOT" = Search on Term1 but NOT Term2

Boolean searching, given these simple, yet powerful, capabilities allows software to quickly narrow searches so the results of a search may quickly pinpoint the information they need. As you add more terms to the specificity of the search, you will find better results returned from search engines.

## Conclusion

The good news about search engines is the user can simply interact with most web-based programs in a non-technical and normal conversation. In discovering what the user wants and likes, the search engine sites have almost advanced to a point of inherently assumed intelligence.  The complexities of calculation exist in the software itself. Conceptual knowledge and the ability to understand the output as related to context is all that is required to successfully connect to the information highway.

## References

1.   Cooper, Todd.  How to Climb the Search Engine Rankings, Infoworld 2000;  22(24):61-64.
2.   http://www.washington.edu, accessed 5.20.2001.
3.   Freeman, Robert, http://www.accessmagazine.com, accessed 4.01.2001.
4.   Zetter,K. Best of the Web 2001, P C World, August, 2001.
5.   Guernsey, Lisa.  Revving Up the Search Engines to Keep the e-Aisles Clear, New York Times, February 28, 2001.

## About the Author

James Day, DDS, MEd

Dr. Day is a lecturer in the Division of Information Technology and Research in the Department of Oral Medicine at the University of Washington.  He serves as the Vice-Chair of the American Dental Association ANSI Standards Committee for development of electronic technologies and Communication in dentistry and a member of the association's Task Force on Distance Learning. Correspondence related to this article and be sent to him at the following address:

Division of Information Technology and Research
Department of Oral Medicine
University of Washington School of Denstistry

e-mail: jimyd@u.washington.edu