

MAKING THE MOST OF YOUR SEARCH

By Robert Blakeley, Director Internet Development
The Direct Marketing Association

Knowing the ins and outs of the most popular search engines on the Internet can help you tame these huge databases and cut down the time it takes to find the information you need.

Marketers use the Web to find information as well as promote products. On the Internet, with a little skill and some luck, you can use a search engine to find what you're looking for from amongst the thousands of pages of information that are available. A search engine is the Web equivalent of a library card. Search tools become very important when you think about how much time you might spend looking for information. Your time is valuable and the time you spend learning how to use a search engine will more than pay for itself.

A search engine is software of a type called an "agent" that searches the Web to find the information in which you have an interest. Generally speaking, you type in a word or phrase in an online form and the search engine returns a list of the documents on the Web containing those words or phrases. When you click on a document in that list, you are shown that page in your Web browser.

Finding information on the Web is easy. Locating the particular piece of information you want can, however, be time-consuming. In fact, most techniques for Web searching concern limiting the information returned by the search engine. For example, if I type in the word jobs in the engine called AltaVista, I get twenty-one million results. This is not useful. There are too many options to deal with. Fortunately, search engines have features that help you refine and focus your search. Unfortunately, most people don't know about them. I discuss these features in detail in this article. It is also helpful to know how the tools are actually designed. Each search engine has a different design and so varies the way it conducts searches. Sometimes one kind of search is more appropriate to one search engine than to another. Your ability to find information on the Web will be enhanced if you can understand the differences among the search engines.

There are hundreds of search engines on the Web. Most of these are specialized and of little use to the average user. I discuss the three most useful search engines here. They are AltaVista, Google, and Yahoo. To find any one, just type in its name (in lower case) in the location bar of your browser.

ALTAVISTA

AltaVista is one of the more comprehensive search engines on the Web. While the searches are inclusive, you also get a lot of extraneous documents that may have only a limited connection to your topic. AltaVista staff has made no editorial decisions regarding content. This search engine is best for finding either obscure information or something very specific, like a book title or a name. AltaVista provides both basic and advanced searches.

AltaVista uses text-based whole words to conduct its searches. For an effective basic search, you should use multiple query terms or phrases to qualify your subject. Because AltaVista tends to return large numbers of search documents, this will increase your odds of finding the exact thing for which you are looking. Also, the more precisely you choose your words the better the results. For example, if you were interested in Rembrandt the artist, you could use the words *Rembrandt*, *artist*, and *Dutch* for the query. This would eliminate "Rembrandt Plus Superior Bleaching System" and "George Rembrandt" who lives in Seattle. You could have chosen to use the word *painting* instead of *artist*. Artist is a word more often associated with information about Rembrandt the man and painting will give you somewhat more information related to his paintings. How would you know this? This is part of the art of searching, which you will pick up by experience.

Capitalization is a way to limit or broaden your searches. In AltaVista, search terms entered in lowercase letters are not sensitive to case, but the use of capitalized terms (or accented letters) makes the term case-sensitive. Using the word *CitiBank* finds only the terms spelled exactly with that capitalization, while *citibank* spelled without capitalization finds all occurrences of the term, regardless of capitalization.

Phrases allow you to look for groups of words such as a full name. To make a phrase, enclose the search words within quotation marks. "Abraham Lincoln" finds an occurrence of the name Abraham Lincoln. If these words were not enclosed in quotation marks AltaVista would look for all occurrences of the word Abraham or Lincoln. You might get Abraham the Patriarch and Lincoln the car as part of your search results. You could use the same method to find documents that contain a phrase such as *from these honored dead we take increased devotion*. Searching for this particular phrase will find you the Gettysburg Address.

To require that one of your terms be included in the document use the plus sign (+) as a prefix to the term, as in *+citibank*. There must not be a space between the + and the term. For example, using the query *+citibank stock investment* locates only documents with the word citibank, but that may also contain the word stock or investment. To prohibit the inclusion of a term in a document, use a minus symbol or a hyphen (-) as a prefix. For example, to find a reference to Pablo Picasso without reference to bulls try using *"Pablo Picasso" -bull*.

You may enter a wildcard character (*) at the end of a phrase which substitutes up to five additional lowercase letters. For example, *dig** will find you all occurrences of dig, digging, digress, digital, etc. The asterisk cannot be used at the beginning or in the middle of words in AltaVista.

Once you have determined your query terms and submitted your search, AltaVista will assign a confidence ranking to the search results. This ranking is determined by the occurrence of the query terms found among the first few words of a document (especially those that appear in the document title), the title of the document and something called Meta Keywords. These "key" words are embedded into the page by the author and help search engines catalog the page. The confidence ranking also considers the proximity of query terms to each other and the number of times a term occurs within a document relative to other selected documents. All these factors are weighted, and the document with the highest confidence rating is given a score. This is the document that appears at the top of the list. All others are given lower scores and presented in descending order from the top document. This does not mean that the document listed at the top is the one you want, but if you

chose your query terms wisely, the document you are looking for should be in the first ten or twenty.

AltaVista permits control of the presentation of the returned document lists in a standard or a compact format. The difference is in the amount of descriptive information provided about the documents the search engine has selected.

Fortunately, search engines have, features that help YOU refine and focus your search.

In addition to the basic search techniques described, AltaVista has an advanced search mode. Advanced searches are available by selecting the **Advanced Search** option from the main AltaVista page. These searches include all the features of the basic search, and also allow you the use of Boolean searching, proximity operators, logical groupings with parentheses, and the use of key words. Using the Boolean method allow you to do really targeted searching and eliminate much of the extraneous material often returned in basic searches.

Boolean operators are a series of special words that tell the search engine how to evaluate your query terms. AltaVista uses the Boolean operators, AND, OR, NEAR, and NOT to include or exclude terms. If any of these Boolean operators in a query refers to a phrase, the phrase must be enclosed within quotation marks. For example, you could construct the query *baseball AND "Abbot and Costello" AND team*. AltaVista would search for all documents that contained all three of the query terms. You would find results related to the famous "Who's on first?" dialog. You could be even more explicit and use the query *baseball AND routine AND "Abbot and Costello" AND team*. The NEAR operator (or Boolean proximity operator) makes sure that both query words appear within ten words of each other, for example, *baseball AND (routine NEAR "Abbot and Costello") AND team*. In this case, using NEAR returns fewer results than just using the AND operator.

You will notice the use of parentheses in the last example. This is called a logical grouping. This tells the AltaVista search engine that you want what is between the parentheses to be treated as a unit. In this example you are asking for documents that contain both the words *baseball* and *team*, and that have the word *routine* within ten words of the phrase *Abbot and Costello*. In another example, the query *(NOT mangos) AND fruit*, returns documents containing the word *fruit* but not the word *mangos*. The query *NOT (mangos AND fruit)*, however, eliminates documents that contain both *mangos* and *fruit*. You must also use parentheses with the operator NOT to exclude a term, as in *Washington (NOT George)*.

With advanced searches you may also specify keywords you want AltaVista to use in order to assign a confidence ranking to your results. This feature lets you control which items are ranked at the top of the returned document list. Before submitting a search, type the keyword into the box entitled **Sorted By**. AltaVista will place documents that contain these words toward the top of the returned search list. This does not affect your search results by including or excluding items. It only influences the order in which the list is presented.

GOOGLE

Google is arguably the most useful search engine available at the moment. Although its searches are not as broad as AltaVista's, Google uses a combination of search methods to produce very good results.

Google, like AltaVista, uses a text-based word search. The more descriptive your search terms, the more targeted your search will be. By default, if you include two words in your search term, Google treats them like an AND search. That is, it will require both words in the document. AltaVista treats them like an OR search. However, Google also supports the use of the Boolean operator OR. Thus, if you enter *CRM Loyalty* it will look for documents that contain both those terms. If you use *CRM OR Loyalty*, it will return documents with either word, or both words. Google is not case sensitive. Again, this is different than AltaVista. The searches for the terms *Direct Marketing Association*, *direct marketing association* and *DIRECT MARKETING AsSoCiAtIon* all return exactly the same results.

You can indicate required terms and prohibited terms. To require a query term to be included in a document, simply put a plus (+) symbol in front of the term. To exclude a term, use the dash or minus sign (-). These symbols function exactly as they do in AltaVista and Yahoo. So, *+pets +dog -cats* would return documents containing the word dog and pets but exclude all those that contain the word cats.

Google also supports grouping terms within parentheses to create complex logic. For example, you can use the query *(quick AND brown AND fox) OR (lazy AND poodle)*. This query will return documents that contain all the words quick, brown, and fox and documents that contain all the words lazy and poodle. Google also supports the use of phrases within quotation marks such as *"The quick brown fox"*.

The more precisely you choose your words the better the results.

If you have found the right Web site but are not sure where the information is located within the site, you can try restricting the search to that site. Simply enter your search terms followed by colon followed by the domain name of the site. For example, *e-mail promotions:www.the-dma.org* will find information about E-mail on The DMA Web site. This is useful to find documents on very large Web sites.

The Google **Advanced Search** can do even more. You can restrict your results to documents created in the past three, six, or twelve months. You can specify if your search terms should occur on the page, in the title or in the URL (the Web site address). For example, I could restrict my search to documents with the words *e-mail AND privacy* in the title and that have been created in the last three months. This would tend to give me the most current documents specifically addressing e-mail privacy issues.

The returned document lists are ranked in an unusual way. Google interprets a link from another site as an indication of value. Therefore, the more sites that link to a page, the higher it will be on

the result list. If the site that links to the page is itself a highly linked site, then they weight that even more, ranking the page still higher.

If you find what you are looking for and want to see more sites like that, you can select the **Similar pages** link next to the listing. This will return a list of sites that Google feels has content like the one you selected. This can also be useful if you have found something close to what you are looking for but not exactly what you want.

In addition to its ability to search the Web via text, Google has an indexed Directory section set up similar to a topic outline. You can simply browse the subject indexes until you find a document you want to review. For example, to find information about *Irish Traditional Dance*, you would select the **Art** category, then **Dance**, then **Folk Dancing** to find documents referencing Irish dance.

You can also perform a Google query from within a Directory section. These particular functions are similar to those of the Yahoo search engine. If you searched for *Levy* in the **Astronomy** section, you would get a lot about the Shoemaker-Levy comet. If you searched the whole directory you would also find information about the Levy Economics Institute.

YAHOO

Yahoo is one of the Internet's earliest search engines and still the best known. Yahoo organizes its information by subject into an outline format. It has an editorial staff that reviews sites for placement in the appropriate category. Because of the way it organizes data, Yahoo is the best way to find information by topic, like *New York City* or the *United States budget*, or *jazz music*. You can browse the indexes and sub-indexes to narrow and focus your search, or query within any index or sub-index. Yahoo, like AltaVista, has both basic searches and advanced searches.

In a basic search, if you are looking for information on astronomy, but aren't exactly sure what you need, you can select science from Yahoo's main index and then astronomy from a sub-index. This produces a list of many documents that pertain to astronomy. Most listings contain brief descriptions of the site, provided by the sites' owners.

You can query Yahoo's entire index or query just the sub-index you are viewing by typing in a search term. For example, from the **Astronomy** index, you could use the query term *Jupiter*. This would tend to return a list of documents about Jupiter the planet. If you were to query the entire Yahoo site for *Jupiter*, you would also find information about Jupiter the Roman god as well as Jupiter the planet.

Yahoo will look for a partial word by default. That is, if you use the word *men*, Yahoo will return documents that contain the word men, such as mentor, menu, augment, development.

In using multiple query terms, Yahoo uses an OR default. That means it will search for documents containing any one of your query terms. As with both AltaVista and Google, you can use the required and prohibitive qualifiers (+ and -) as prefixes to include or exclude terms from your search. So again, *+pets +dog -cats* would return documents containing the words dogs and pets but

exclude all those that contain the word cats. Using the + or - is the same as the Boolean AND or the Boolean NOT.

You can also look for phrases in Yahoo by using quotation marks. "The quick brown fox" will return a list of documents that contain that phrase. A wild card search uses the * operator. This finds documents that contain terms that are part of a word. For example, doc* will return doctor, doctors, document, documentation, etc.

Yahoo has two additional operators. Yahoo can search for matching document titles only. Just enter a query term and use t: as a prefix to that term. Likewise, if you use u: as a prefix to a query term, Yahoo will search only a document's URL (Uniform Resource Locator), which is its address that appears in the browser location bar on your screen. The search terms bush, t: bush, and u: bush return very different results. The first retrieves document with the term bush in the body of the document. In the second, the term must be in the document title. In the third, the term bush must be in the web address of the document (I.E. www.bushgardens.com)

Yahoo returns a results list divided into three types of information. The first section contains a list of sponsor links. The owners of these sites have paid to be at the top of the page. Next is an alphabetical list of matching Yahoo categories. This is followed by a list of documents that match the search terms.

Users can also conduct advanced searches by selecting **Advanced Search** from Yahoo's regular search page. From there, you can select four query options. These correspond to the Boolean AND, OR, NOT and a search for an exact phrase. Another search option allows you to select a search area, either Yahoo categories or other Web sites.

CONCLUSION

This column is only a brief look at three search engines. AltaVista is an engine that lends itself to broad searches for specific or obscure pieces of information. Yahoo is best when you are not sure what information you are looking for, but know the general topic, while Google is a general, all-purpose tool that is useful when your needs fall between AltaVista or Yahoo. There are many other search engines. Many provide documentation on their site that will tell you how they work and how best to use them.

Remember that search engines are tools. The more experience you have with these tools, the quicker you will learn to quickly find the information you need. So, too, the better your research skills, the more effective will be your use of the tools.

Robert Blakeley is Director of Internet Development for the Direct Marketing Association in New York City. Mr. Blakeley has worked in the Internet industry for 9 years and has worked with many companies and government agencies. These include Atlantic City and the Agency for Economic Development and its U.N. member nations. He can be reached at rblakeley@the-dma.org.