

Smarter Searching

How to find what you're looking for on the internet by searching smarter

They say that clearly defining any problem is half of the solution. It's the same with searching on the internet. The more specific you are, the faster you can get to the information you need.

Are you looking for information about a person? A company? A software product? A health-related problem? Do you want to find a job? Get a date? Plan a vacation? Do you need to research a term paper? Document a news story? Size up your company's competition? You may need to use different search tools and search engines to find what you're looking for. And to do that, you're going to need to think the way a search engine does.

1. Why do I need a search engine?

A search engine plays the same role as a card catalogue in a library. There are lots of great and useful information in a library, but it's physically impossible to examine all the books personally. Not even the most avid web-surfer could hyperlink to all the documents in the aptly named World Wide Web. There are millions of pages and billions of words on the Web. And more are being posted every minute of every day. The search engines and directories help you sift through all those billions of 1's and 0's to find the specific information you need.

2. If it's impossible to examine all the documents on the Web, how do the search engines do it?

They use software programs known as robots, spiders or crawlers. A robot is a piece of software that automatically follows hyperlinks from one document to the next around the Web. When a robot discovers a new site, it sends information back to its main site to be indexed. Because Web documents are changing constantly, spiders also update previously catalogued sites. How good they are in carrying out these tasks varies from one search engine to the next.

3. What's the difference between a Web directory like Yahoo and a Web search engine like Google?

There is less difference now than there used to be, because many search engines have built large subject catalogues to help you search. But think of a Web directory as a subject catalogue--something like the subject catalogue in the library. Yahoo attempts to organize Web by dividing it into topics and subtopics. Some examples include: Arts, Science, Health, Business, News, and Entertainment. If you're looking for information on the Web that fits neatly into an obvious subject or category, go first to Yahoo or one of the other Web directories.

Think of a Web search engine as an index like at the back of a text book that enables you to seek out specific words and phrases. With the search engine's help, you can locate individual appearances of such words in documents all over the Web. This can be both a blessing and a curse--usually the latter! You might get way too many hits. Or you might discover that your keyword has meanings you didn't anticipate. Or you might get no hits at all.

So now that you have decided whether to search through either a directory, using categories, or through a search engine, using index words, it's time to organize your thoughts into priority order and start surfing.

2. Fine-tune your keywords

Be specific. Try to meet the search engine halfway by refining your search before you begin.

Example: If you want to buy a car, don't enter the keyword "car" if you can enter the keyword " Toyota ." Better still, enter the phrase "Toyota Dealerships" AND the name of the city where you live.

Remember that search engines use software robots to survey the Web and build their databases. Web documents are retrieved and indexed. When you enter your keyword at a search engine website, your input is checked against the search engine's keyword indices. The best matches are then returned to you as hits.

There are two primary methods of text searching--keyword and concept.

Keyword Searching

This is the most common form of text search on the Web. Most search engines do their text query and retrieval using keywords.

When you're entering search terms, the more specific your words or phrases are, the better. Many times, search engines will ignore words that are too common because they slow down the search by returning too many hits. Focus your search by adding terms. You want to arrange your terms from left to right, with the most important one first because some search engines prioritize words in that order. Make sure you spell the terms correctly as search engines are very literal – if you're using Google, it will offer alternate spelling suggestions.

Also, think of the form of the word you're entering. Searching on the word "color" won't necessarily retrieve pages with the words colors or coloring. A number of sites support what is called wildcard characters, so you can enter color* search all the words beginning with the word color. This is also called "stemming" – Yahoo! Is one of the few sites that stems automatically. Frequently Asked Questions are a wealth of information provided by many sites. Adding FAQ to the end of any string of keywords and it can help you in a search for general information.

The Problem with Keyword Searching

Keyword searches have a tough time distinguishing between words that are spelled the same way, but mean something different (i.e. hard cider, a hard stone, a hard exam, and the hard drive on your computer). This often results in hits that are completely irrelevant to your query. Some search engines also have trouble with stemming and can't work through the complicated English language where singular and plural versions of a word can differ dramatically or in the case of infants' names to adult names. (gosling to goose to geese)

Search engines also cannot return hits on keywords that mean the same, but are not actually entered in your query. A query on heart disease would not return a document that used the word "cardiac" instead of "heart."

So what about Concept-based searching?

Unlike keyword search systems, concept-based searches try to determine what you mean, not just what you say. In the best circumstances, a concept-based search returns hits on documents that are "about" the subject/theme you're exploring, even if the words in the document don't precisely match the words you enter into the query.

This is also known as clustering -- which essentially means that words are examined in relation to other words found

nearby.

How does it work? Well, search engines that use this search method have created software that sticks to a numerical approach. It determines meaning by calculating the frequency with which certain important words appear. When several words or phrases that are tagged to signal a particular concept appear close to each other in a text, the search engine concludes, by statistical analysis that the piece is "about" a certain subject.

For example, the word heart, when used in the medical/health context, would be likely to appear with such words as coronary, artery, lung, stroke, cholesterol, pump, blood, and attack.

If the word heart appears in a document with others words such as flowers, candy, love, passion, and valentine, a very different context is established, and the search engine returns hits on the subject of romance.

Warning: This often works better in theory than in practice. Concept-based indexing is a good idea, but it's far from perfect. The results are best when you enter a lot of words, all of which roughly refer to the concept you're seeking information about.

Refining Your Search

Most sites offer two different types of searches--"basic" and "refined." In a "basic" search, you just enter a keyword without sifting through any pull down menus of additional options.

Search refining options differ from one search engine to another, but some of the possibilities include the ability to search on more than one word, to give more weight to one search term than you give to another, and to exclude words that might muddy the results. You might also be able to search on proper names, on phrases, and on words that are found within a certain proximity to other search terms.

Read the help files found on each search engine site and take advantage of the available refining options that can be found there.

Certain operators and symbols tell the search engine how to treat your search terms. AND, OR and NOT are Boolean logic terms – named after mathematician George Boole – you can use these terms to link or exclude certain words from your search. In some search engines, the plus sign and the minus sign take the places of AND and NOT.

The Boolean AND means that all the terms you specify must appear in the documents, i.e., "heart" AND "attack." You might use this if you wanted to exclude common hits that would be irrelevant to your query.

The Boolean OR means that at least one of the terms you specify must appear in the documents, i.e., you might use this if you didn't want to rule out too much.

Also learn to EXCLUDE with the Boolean NOT. Excluding is particularly important as the Web grows and more documents are posted. Don't be discouraged if you need to run your initial query over again several times, each time adding further refinements to narrow down your list of relevant hits – this is to be expected.

Example: If you want to find out how medical details about a family member's diagnosis of Alzheimer's disease, try entering "Alzheimer's" AND "symptoms" AND "prognosis." If you want to find out about Alzheimer's care and community resources, query on "Alzheimer's" AND "support groups" AND "resources" NOT "symptoms."

Here's another example. You might start searching for a new laptop by entering "laptop OR notebook review." OR tells the search engine to retrieve pages with either word. The results are plenty of links to follow that will help you to compare laptop models. Say you decide to go with Dell. You can then use the Boolean terms to further refine your search by entering laptops OR notebooks AND Dell AND price. You now have all the information you need to select the laptop that fits your needs and your budget. Phrases are another entry you can use to search on the web: The ability to query on phrases is very important in a search engine. Those that allow it usually require that you enclose the phrase in quotation marks, i.e., "space the final frontier." This is particularly useful when you're searching for a name. If you

type Geoff Smith , you will get millions of links that have separate references to Geoff and to Smith. "Geoff Smith" works better. Here's an example: When Bill and Melinda Gates had their baby girl, lots of people wanted to find out her name. But if you enter "Bill and Melinda Gates" baby, you would get pages about the Bill and Melinda Gates Foundation and babies in general, so you need to add more definitive terms. If you refined your search to "Bill and Melinda Gates" daughter, the first five results were announcements about the new Gates baby.

Capitalization: is also essential for searching on proper names of people, companies or products. Unfortunately, many words in English are used both as proper and common nouns—the name Bill, or the dollar bill, the name Lotus, or the lotus flower, Digital, digital—the list is endless.

All the search engines have different methods of refining queries. The best way to learn them is to read the help files on the search engine sites and practice!

Relevancy Rankings

Most of the search engines return results with confidence or relevancy rankings. In other words, they list the hits according to how closely they think the results match the query. However, these lists often leave users shaking their heads on confusion, since, to the user, the results often seem completely irrelevant.

Why does this happen? Basically it's because search engine technology has not yet reached the point where humans and computers understand each other well enough to communicate clearly.

Most search engines use search term frequency as a primary way of determining whether a document is relevant. If you're researching diabetes and the word "diabetes" appears multiple times in a Web document, it's reasonable to assume that the document will contain useful information. Therefore, a document that repeats the word "diabetes" over and over is likely to turn up near the top of your list.

If your keyword is a common one, or if it has multiple other meanings, you could end up with a lot of irrelevant hits. And if your keyword is a subject about which you desire information, you don't need to see it repeated over and over—it's the information about that word that you're interested in, not the word itself.

Some search engines consider 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, this increases the likelihood that the document is on target.

As far as the user is concerned, relevancy ranking is critical, and becomes more so as the sheer volume of information on the Web grows. Most of us don't have the time to sift through scores of hits to determine which hyperlinks we should actually explore. The more clearly relevant the results are, the more we're likely to value the search engine.

The five main tips to searching smarter are: Use Directories for general information and Search Engines for specific terms. Use precise terms such as computer tech support, Campbell River, instead of Computers, Vancouver Use wildcard characters to find all forms of a word (*) Use Boolean operators (AND, OR, NOT, + and -) to weed out irrelevant results Put exact phrases and proper names in quotation marks.

If you understand how search engines organize information and run queries, you can maximize your chances of getting hits on web sites that matter.

Your best bet is to be familiar with a variety of engines and directories if you need to do a thorough search because at this point, no one service can index the web completely, and remember, there is no wrong way to surf the web as long as you get the results you're looking for.