

From SearchEngineWatch.com

(Go to <http://searchenginewatch.com/facts/> to use the links shown below)

EVERYTHING YOU WANTED TO KNOW (AND A LOT MORE) ABOUT ON-LINE SEARCHING

1. **How Search Engines Work**
2. **Search Engine Features For Webmasters**
3. **How Search Engines Rank Web Pages**
4. **Search Engine Math**
5. **Power Searching for Anyone**
6. **Search Assistance Features**
7. **Search Features Chart**
8. **Search Toolbars and Utilities**
9. **Boolean Search**
10. **Search Engine Glossary**

1. How Search Engines Work

The term "search engine" is often used generically to describe both crawler-based search engines and human-powered directories. These two types of search engines gather their listings in radically different ways.

Crawler-Based Search Engines

Crawler-based search engines, such as Google, create their listings automatically. They "crawl" or "spider" the web, then people search through what they have found.

If you change your web pages, crawler-based search engines eventually find these changes, and that can affect how you are listed. Page titles, body copy and other elements all play a role.

Human-Powered Directories

A human-powered directory, such as the Open Directory, depends on humans for its listings. You submit a short description to the directory for your entire site, or editors write one for sites they review. A search looks for matches only in the descriptions submitted.

Changing your web pages has no effect on your listing. Things that are useful for improving a listing with a search engine have nothing to do with improving a listing in a directory. The only exception is that a good site, with good content, might be more likely to get reviewed for free than a poor site.

"Hybrid Search Engines" Or Mixed Results

In the web's early days, it used to be that a search engine either presented crawler-based results or human-powered listings. Today, it is extremely common for both types of results to be presented. Usually, a hybrid search engine will favor one type of listings over another. For example, MSN Search is more likely

to present human-powered listings from LookSmart. However, it does also present crawler-based results (as provided by Inktomi), especially for more obscure queries.

The Parts Of A Crawler-Based Search Engine

Crawler-based search engines have three major elements. First is the spider, also called the crawler. The spider visits a web page, reads it, and then follows links to other pages within the site. This is what it means when someone refers to a site being "spidered" or "crawled." The spider returns to the site on a regular basis, such as every month or two, to look for changes.

Everything the spider finds goes into the second part of the search engine, the index. The index, sometimes called the catalog, is like a giant book containing a copy of every web page that the spider finds. If a web page changes, then this book is updated with new information.

Sometimes it can take a while for new pages or changes that the spider finds to be added to the index. Thus, a web page may have been "spidered" but not yet "indexed." Until it is indexed -- added to the index -- it is not available to those searching with the search engine.

Search engine software is the third part of a search engine. This is the program that sifts through the millions of pages recorded in the index to find matches to a search and rank them in order of what it believes is most relevant. You can learn more about how search engine software ranks web pages on the aptly-named [How Search Engines Rank Web Pages](#) page.

Major Search Engines: The Same, But Different

All crawler-based search engines have the basic parts described above, but there are differences in how these parts are tuned. That is why the same search on different search engines often produces different results. Some of the significant differences between the major crawler-based search engines are summarized on the [Search Engine Features Page](#). Information on this page has been drawn from the help pages of each search engine, along with knowledge gained from articles, reviews, books, independent research, tips from others and additional information received directly from the various search engines.



2. Search Engine Features For Webmasters

December 5, 2002

The search engine features chart below is designed primarily for webmasters who care about how crawler-based search engines index their sites. It provides a summary of important factors and features that can affect how sites are indexed and ranked. Full explanations of items can be found immediately below the comparison chart.

Human-powered search engines like the Open Directory are not listed on this chart because they do not crawl the web to create their listings. See the [How Search Engines Work](#) page for an explanation of the differences between crawler-based and human-powered services.

See the [Search Engine Features For Searchers](#) page for a summary of how search engines display their results and other information that may be of interest to searchers, rather than search engine marketers and site promoters.

The How Search Engines Work section available to Search Engine Watch members provides more detailed information about the crawlers listed below. This section is just one of the many benefits that Search Engine Watch members receive. To learn more about becoming a member, please visit the membership information page.

This chart covers the crawler of AllTheWeb, AltaVista, Google, Inktomi and Teoma. Some of these crawlers power other search engines, and the relationships are shown on the Search Engine Results page.

Crawling	Yes	No	Notes
Deep Crawl	AllTheWeb, Google, Inktomi	AltaVista, Teoma	
Frames Support	All	n/a	
robots.txt	All	n/a	
Meta Robots Tag	All	n/a	
Paid Inclusion	All but...	Google	
Full Body Text	All	n/a	Some stop words may not be indexed
Stop Words	AltaVista, Inktomi, Google	FAST	Teoma unknown
Meta Description	All provide some support, but AltaVista, AllTheWeb and Teoma make most use of the tag		
Meta Keywords	Inktomi, Teoma	AllTheWeb, AltaVista, Google	Teoma support is "unofficial"
ALT text	AltaVista, Google, Teoma	AllTheWeb, Inktomi	
Comments	Inktomi	Others	

Deep Crawl

All crawlers will find pages to add to their web page indexes, even if those pages have never been submitted to them. However, some crawlers are better than others. This section of the chart shows which search engines are likely to do a "deep crawl" and gather many pages from your web site, even if these pages were never submitted. In general, the larger a search engine's index is, the more likely it will list many pages per site. See the Search Engine Sizes page for the latest index sizes at the major search engines.

Frames Support

This shows which search engines can follow frame links. Those that can't will probably miss listing much of your site. However, even for those that do, having individual frame links indexed can pose problem. Be sure to read the Search Engines And Frames page for tips on overcoming the problems with frames and search engines.

robots.txt

The robots.txt file is a means for webmasters to keep search engines out of their sites. Search Engine Watch members have access to the Blocking Crawlers With Robots.txt page, which covers the robots.txt file in more detail. The Web Robots Pages: The Robots Exclusion Protocol page also provides official information about robots.txt.

Meta Robots Tag

This is a special meta tag that allows site owners to specify that a page shouldn't be indexed. It is explained more on the How HTML Meta Tags Work page and also on the Blocking Crawlers With The Meta Robots Tag page, available to Search Engine Watch members. The Web Robots Pages: The Robots META tag page also provides official information about robots.txt.

Paid Inclusion

Shows whether a search engine offers a program where you can pay to be guaranteed that your pages will be included in its index. This is NOT the same as paid placement, which guarantees a particular position in relation to a particular search term. The Submitting To Crawlers page provides links to various paid inclusion programs.

Full Body Text

All of the major search engines say they index the full visible body text of a page, though some will not index stop words or exclude copy deemed to be spam (explained further below). Google generally does not index past the first 101K of long HTML pages.

Stop Words

Some search engines either leave out words when they index a page or may not search for these words during a query. These stop words are excluded as a way to save storage space or to speed searches.

Meta Description

All the major crawlers support the meta description tag, to some degree. The ones actually named on the chart are very consistent. If you have a meta description tag on your pages, you'll most likely see the content used in some way.

The How HTML Meta Tags Work page explains how to use the meta description tag, and the Search Engine Display Chart for Search Engine Watch members provides a more detailed breakdown of how crawler-based search engines form descriptions.

Meta Keywords

Shows which search engines support the meta keywords tags, as explained on the How HTML Meta Tags Work page.

ALT Text / Comments

This shows which search engines index ALT text associated with images or text in comment tags.



3. How Search Engines Rank Web Pages

Search for anything using your favorite crawler-based search engine. Nearly instantly, the search engine will sort through the millions of pages it knows about and present you with ones that match your topic. The matches will even be ranked, so that the most relevant ones come first.

Of course, the search engines don't always get it right. Non-relevant pages make it through, and sometimes it may take a little more digging to find what you are looking for. But, by and large, search engines do an amazing job.

As WebCrawler founder Brian Pinkerton puts it, "Imagine walking up to a librarian and saying, 'travel.' They're going to look at you with a blank face."

OK -- a librarian's not really going to stare at you with a vacant expression. Instead, they're going to ask you questions to better understand what you are looking for.

Unfortunately, search engines don't have the ability to ask a few questions to focus your search, as a librarian can. They also can't rely on judgment and past experience to rank web pages, in the way humans can.

So, how do crawler-based search engines go about determining relevancy, when confronted with hundreds of millions of web pages to sort through? They follow a set of rules, known as an algorithm. Exactly how a particular search engine's algorithm works is a closely-kept trade secret. However, all major search engines follow the general rules below.

Location, Location, Location...and Frequency

One of the the main rules in a ranking algorithm involves the location and frequency of keywords on a web page. Call it the location/frequency method, for short.

Remember the librarian mentioned above? They need to find books to match your request of "travel," so it makes sense that they first look at books with travel in the title. Search engines operate the same way. Pages with the search terms appearing in the HTML title tag are often assumed to be more relevant than others to the topic.

Search engines will also check to see if the search keywords appear near the top of a web page, such as in the headline or in the first few paragraphs of text. They assume that any page relevant to the topic will mention those words right from the beginning.

Frequency is the other major factor in how search engines determine relevancy. A search engine will analyze how often keywords appear in relation to other words in a web page. Those with a higher frequency are often deemed more relevant than other web pages.

Spice In The Recipe

Now it's time to qualify the location/frequency method described above. All the major search engines follow it to some degree, in the same way cooks may follow a standard chili recipe. But cooks like to add their own secret ingredients. In the same way, search engines add spice to the location/frequency method. Nobody does it exactly the same, which is one reason why the same search on different search engines produces different results.

To begin with, some search engines index more web pages than others. Some search engines also index web pages more often than others. The result is that no search engine has the exact same collection of web pages to search through. That naturally produces differences, when comparing their results.

Search engines may also penalize pages or exclude them from the index, if they detect search engine "spamming." An example is when a word is repeated hundreds of times on a page, to increase the frequency and propel the page higher in the listings. Search engines watch for common spamming methods in a variety of ways, including following up on complaints from their users.

Off The Page Factors

Crawler-based search engines have plenty of experience now with webmasters who constantly rewrite their web pages in an attempt to gain better rankings. Some sophisticated webmasters may even go to great lengths to "reverse engineer" the location/frequency systems used by a particular search engine. Because of this, all major search engines now also make use of "off the page" ranking criteria.

Off the page factors are those that a webmasters cannot easily influence. Chief among these is link analysis. By analyzing how pages link to each other, a search engine can both determine what a page is about and whether that page is deemed to be "important" and thus deserving of a ranking boost. In addition, sophisticated techniques are used to screen out attempts by webmasters to build "artificial" links designed to boost their rankings.

Another off the page factor is clickthrough measurement. In short, this means that a search engine may watch what results someone selects for a particular search, then eventually drop high-ranking pages that aren't attracting clicks, while promoting lower-ranking pages that do pull in visitors. As with link analysis, systems are used to compensate for artificial links generated by eager webmasters.



4. Search Engine Math

Forget power searching. Don't worry about learning to do a "Boolean" search. All most people need to know is a little basic "search engine math" in order to improve their results. Come learn how to easily add, subtract and multiply your way into better searches at your favorite search engine. The information below works for nearly all of the major search engines.

Be Specific

Before learning math, it's a helpful reminder that the more specific your search is, the more likely you will find what you want. Don't be afraid to tell a search engine exactly what you are looking for.

For example, if you want information about Windows 98 bugs, search for "Windows 98 bugs," not "Windows." Or even better, search for exactly what the problem is: "I can't install a USB device in Windows 98," for example. You'll be surprised at how often this works.

Using The + Symbol to Add

Sometimes, you want to make sure that a search engine finds pages that have all the words you enter, not just some of them. The + symbol lets you do this.

For example, imagine you want to find pages that have references to both President Clinton and Kenneth Starr on the same page. You could search this way:

+clinton +starr

Only pages that contain both words would appear in your results. Here are some other examples:

+windows +98 +bugs

That would find pages that have all three of the words on them, helpful if you wanted to narrow down a search to Windows 98 bugs, rather than on Windows 98 in general.

+star +trek +insurrection

That would get you pages about Star Trek that also specifically mention "Insurrection," the title of a Star Trek film.

The + symbol is especially helpful when you do a search and then find yourself overwhelmed with information. Imagine that you wanted to reserve a camping space in California's Yosemite National Park. You might start out simply searching like this:

yosemite

If so, chances are, you'll probably get too many off-target results. Instead, try searching for all the words you know must appear on the type of page you're looking for:

+yosemite +camping +reservations

Using The - Symbol to Subtract

Sometimes, you want a search engine to find pages that have one word on them but not another word. The - symbol lets you do this.

For example, imagine you want information about President Clinton but *don't* want to be overwhelmed by pages relating to the Monica Lewinsky scandal. You could search this way:

clinton -lewinsky

That tells the search engine to find pages that mention "clinton" and then to remove any of them that also mention "lewinsky."

Similarly, perhaps you are looking for information specifically about Windows 95 but keep getting pages about Windows 98 or Windows 3.1. You could eliminate them with a search like this:

windows -98 -3.1

Perhaps you are a fan of the original Star Trek series but instead keep finding pages about Voyager, Deep Space Nine or Star Trek: The Next Generation. Try a search like this:

star trek -voyager -deep -space -nine -next -generation

In general, the - symbol is helpful for focusing results when you get too many that are unrelated to your topic. Simply begin subtracting terms you know are not of interest, and you should get better results.

Using Quotation Marks To Multiply

Now that you know how to add and subtract terms, we can move on to multiplication. As in normal math, multiplying terms through a "phrase search" can be a much better way to get the answers you are looking for.

For example, remember above when we wanted pages about reserving a campsite in Yosemite? We entered all the terms like this:

+yosemite +camping +reservations

That brings back pages that have all those words on them, but there's no guarantee that the words may necessarily be near each other. You could get a page that mentions Yosemite in the opening paragraph but then later talks about getting camping reservations in the Grand Canyon. All the words you added together would appear on this page, but it still might not be what you are looking for.

Doing a phrase search avoids this problem. This is where you tell a search engine to give you pages where the terms appear in exactly the order you specify. You do this by putting quotation marks around the phrase, like this:

"yosemite camping reservations"

Now, only pages that have all the words and in the exact order shown above will be listed. The answers should be much more on target than with simple addition.

Likewise, remember this addition example?

+windows +98 +bugs

As you can imagine, multiplying the terms together within a phrase search would work better, because that exact phrase probably appears on good pages dealing with Windows 98 bugs. So try this:

"windows 98 bugs"

Remember the search for information about the latest Star Trek movie? We could transform that into a phrase search like this:

"star trek insurrection"

But the movie's title actually has a colon after the word "trek," and many pages might also follow this format. Thus, a better phrase search might be:

"star trek: insurrection"

Combining Symbols

Once you've mastered adding, subtracting and multiplying, you can combine symbols to easily create targeted searches.

For example, remember the person who wanted pages only about Star Trek's original series? We searched this way:

star trek -voyager -deep -space -nine -next -generation

A better search might use subtraction and multiplication:

"star trek" -voyager -"deep space nine" -"next generation"



5. Power Searching for Anyone

Search engines have a variety of ways for you to refine and control your searches. Some of them offer menu systems for this. Others require you to use special commands as part of your query.

For most people, the basic commands covered on the Search Engine Math page will be sufficient. I encourage you to read the search engine math page first. Get comfortable using the commands that are described. If you need more power after that, then review the other options on this page.

Boolean commands are NOT shown on this page. See the separate Boolean Searching page for information about these. If you don't know what Boolean commands are, don't worry about reading the page on these commands. You probably don't need them.

Not every power searching command is shown on this page, only the main ones that are most likely to be used. Read the help files at each search engine for more detailed coverage about what they offer.

Match Any

Sometimes you want pages that contain *any* of your search terms. For example, you may want to find pages that say either Ireland or Eire. The Search Features Chart shows which search engines will do this type of search by default, without you needing to specify any commands.

At some search engines, you can do a Match Any search by using a menu next to the search box or on the advanced search page. The Search Features Chart lists where this is possible.

Keep in mind that most search engines will automatically first list pages that have all your terms, then some of your terms, when you perform a Match Any search.

Some search engine specific notes are below:

AltaVista

At AltaVista, testing shows that Match Any is most likely what will happen in response to a default search. Earlier in 2001, AltaVista had said that Match Any would only occur if you searched for five words or

more. This no longer seems to be the case. The article below explains what AltaVista previously said would happen:

[Blending Vertical Results & Other AltaVista Improvements](#)

The Search Engine Report, March 5, 2001

Match All

This is a search for pages containing *all* of your search terms, rather than *any* of them. For example, you may want to find pages with references to both Clinton and Dole on the same page.

Practically all major search engines support the + symbol as a means of doing a Match All search. These are listed on the Search Features Chart. The chart also shows which search engines will perform a Match All search by default, even if you don't use the + symbol.

See the Search Engine Math page for more specific help on using the + symbol. Some search engine specific notes are below:

AOL Search

By default, AOL Search will look for any sites in its Open Directory information that contain all the words you enter. It will check both the words in the Open Directory listing and the words on the page that the listing leads to.

AOL Search will not check for matches in its Inktomi listings UNLESS there are absolutely no Open Directory listings that match all words. However, if you use AOL Search's advanced search page (see the Search Assistance page) and choose the "On the Web Only" option, then your search will be conducted against only Inktomi's listings.

Exclude

Most major search engines allow you to exclude documents that contain certain words. This is a helpful way to narrow a search.

For example, you may want a page about the philosopher Calvin, not the cartoon character Calvin. By excluding pages that mention Hobbes, the cartoon character's sidekick, you will get better results.

The best way to do this is by using the - command, which is supported by practically all major search engines. These are listed on the Search Features Chart.

See the Search Engine Math page for more specific help on using the - symbol.

Site Search

One of the most powerful features available is the ability to control what sites are included or excluded from a search. For example, imagine you wanted to see all the pages from the Mars Exploration web site run by the NASA's Jet Propulsion Laboratory. At AltaVista, you could use this command:

host:mars.jpl.nasa.gov

In response, AltaVista would display all the pages it has indexed from the mars.jpl.nasa.gov domain. More about using the site search command to find web pages from a particular web site is described on the Checking Your URL page.

Now imagine you wanted to find all the pages from the Mars Exploration web sites that also mention Venus and Jupiter. You could do that this way:

host:mars.jpl.nasa.gov venus jupiter

That tells AltaVista to list pages with the words "venus" and "jupiter" that are within the Mars Exploration web site.

You can even combine other commands, such as those described on the Search Engine Math page. For instance, look at this example:

host:mars.jpl.nasa.gov -"mars pathfinder"

Here, we are telling AltaVista to list all pages within the Mars Exploration web site that do not contain the exact phrase "mars pathfinder."

Now, imagine you are looking for information about Mars landings but are getting overwhelmed by results from NASA. You can get rid of NASA pages by doing this:

"mars exploration " -host:nasa.gov

In that example, we are looking for the phrase "mars landings" but excluding any pages from sites that end in nasa.gov. That means we will NOT get pages from sites like these

- mars.jpl.nasa.gov
- spacekids.hq.nasa.gov
- cmex.arc.nasa.gov.

We could even decide to see all pages about Mars landing from US educational sites, which end in .edu, like this:

"mars landings" +host:edu

Finally, imagine you live outside the US and want to see results that are predominately from your country. Here's how someone in the UK might search for football (soccer) information:

"football scores" +host:uk

This finds pages that say "football scores" and which are from sites that end in the .uk domain, which is used for UK-based sites.

Search Engine Specific Issues

The examples shown above all use the command that works at AltaVista. The same examples work at Google, FAST Search and some Inktomi-powered search engines, if you use the corresponding site search command that these each offer. The site search command for each of these search engines is listed on the Search Features Chart. The Checking Your URL page provides some additional search engine specific guidance.

Often, search engines that support a site search command also make this possible to do using their advanced search pages. In addition, I'd highly recommend downloading the Google Toolbar. Once you've done this, when visiting any web site, you can use the toolbar's "Search site" button to search within just that web site.

Finally, for search engines that don't offer a site search command, you may find that there is a URL Search command that provides a similar ability.

URL Search

Several search engines offer the ability to search within the text of a URL. This is very similar to performing a site search. The Search Features Chart shows which search engines have this capability and the exact command to use. Some additional search engine specific notes are below:

Excite

Excite has a "site" command as explained in the Site Search section, but this command cannot be combined with search terms in an attempt to locate pages on a particular topic from a particular web site or to filter out pages from a particular web site. For example, this query wouldn't work:

mars exploration -site:mars.jpl.nasa.gov

However, you can use the URL command to get a similar result. For instance:

mars exploration -url:mars.jpl.nasa.gov

would work to list pages about "mars exploration" but would remove any that came from the mars.jpl.nasa.gov site. Be aware that when using the URL command in this way, only the exact site listed will be removed. For example, this query:

mars exploration -url:nasa.gov

would remove pages from nasa.gov but still allow pages from mars.jpl.nasa.gov to appear, since that is a different web site.

However, when using the + command, then any sites containing the core domain will be included. In other words, this command:

mars exploration +url:nasa.gov

would bring up pages from any site that has nasa.gov in the URL, such as

- mars.jpl.nasa.gov
- spacekids.hq.nasa.gov

- cmex.arc.nasa.gov.

Google

Google's advanced search page uses the `allinurl` command for finding URLs that contain certain words, as described more on the [Checking Your Listing](#) page. However, it is the undocumented `inurl` command that you should use, if you want to find both web pages with words in the URL and within the pages themselves.

For example, let's say you want to find PDF files about mars exploration. Entering "mars exploration" isn't enough, because that could bring back both HTML and PDF pages. To solve this, you can use the `inurl` command to specify that URLs must have the word "pdf" in them, which will increase the chances of getting PDF files. Here's both commands, combined:

`mars exploration inurl:pdf`

If you used the `allinurl` command rather than the `inurl` command, this search wouldn't work.

By the way, the `allinurl` command takes its name because when using it, you are requiring that ALL the words appear IN the URL. In contrast, the `inurl` command means that ANY of the words you specify should appear.

Google also has a command that lets you narrow your search to find documents in particular formats, such works better than forcing the URL command into this role. The command is `filetype:`, and you follow it with the extension you want to search for. For instance:

`california power crisis filetype:pdf`

brings back PDF files that contain the words "california power crisis." In contrast:

`california power crisis filetype:asp`

brings back Microsoft Active Server Pages (ASP) files, while

`california power crisis filetype:html`

brings back ordinary HTML files that end in `.html`, that contain the words. It will not bring back HTML files the end in `.htm`, however. Technically, Google considers those to be a different file type, simply because the ending is different.

Link Search

Several search engines offer the ability to search for all the pages linking to a particular page or domain. The [Search Features Chart](#) shows which search engines have this capability and the exact command to use. The [Measuring Link Popularity](#) page provides some specific examples.

Many of the major crawler-based search engines allow you to search within the HTML title of a web page. This is the text that appears within the title tag of a document. For example, this page that you are reading now has an HTML title like this:

NOTE: Image was not viewable on Website from which this materials was obtained. Below is the source code from the top of the main page:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/1999/REC-html401-19991224/loose.dtd">
<html>
<head>
<TITLE>Power Searching For Anyone</title>
<!--meta http-equiv="Content-Type" content="text/html; charset=utf-8"-->
<META NAME="authors" CONTENT=" Danny Sullivan">
<META NAME="date" CONTENT="20011026">
<META NAME="channel" CONTENT="internet technology">
<META NAME="description" CONTENT="A summary of how to do advanced searches with the major
search engines, including field searching.">
```

If someone were to do a title search for "power searching," then this page might appear, because those words appear in the HTML title. To learn more about the title tag, see the [How To Use HTML Meta Tags](#) page.

All the major crawlers offer a title search option, though the exact command and how it operates varies, as explained below.

AllTheWeb, AltaVista & Inktomi (via HotBot)

AllTheWeb, AltaVista and Inktomi (as accessed via HotBot) use the *title:* command, where you follow the command with the word you want to find in the title of documents. For example, if you wanted to find all the pages listed that had the word "mars" in their titles, you would do this:

title:mars

What if you wanted to find pages that had two different words in their titles, such as "mars" and "landings"? Then you need to prefix both words with the *title:* command, like this:

title:mars title:landings

That will bring back pages that have both words in their titles, regardless of where exactly the words appear within the title. In other words, pages with these different titles would all be found:

**Future Mars Landings
Mission to Mars: The Landings
China targeting landings on the Moon, Mars**

What if you wanted pages that had the words in particular order, such as "mars landings" in that order? Then use the phrase search command (described more on the [Search Engine Math](#) page), prefixed by the *title:* command, like this:

title:"mars landings"

Performing a search like that means that only pages with those words, in that exact order, would be found. In other words, in the example page titles shown above, only this page would be retrieved:

Future Mars Landings

Please note that searching for phrases within titles as shown above does NOT work with Inktomi, but testing shows it does work with AllTheWeb and AltaVista.

Google & Teoma

To title search at Google and Teoma, you need to use the *intitle:* command. This means to find a single word like "mars" within the title of documents, you would enter:

intitle:mars

What if you want to find multiple words in the title of documents, such as "mars" and "landings"? At Teoma, all testing indicates that you can simply add these words after the *intitle:* command, like this:

intitle:mars landings

With Google, use the *allintitle:* command, which means that Google will find documents that have ALL the words you specify in their page title. The command would be used like this:

allintitle:mars landings

As for phrase searching within the title of documents, use the quotation marks to surround an exact phrase, prefixed by the *intitle:* command, like this:

intitle:"mars landings"

Other Notes

Several of the crawlers listed above provide results to other search engines, so you may find the title command will also work at their partners. The Search Features Chart lists some of the major partnerships.

Don't forget that all of the search engines listed above, plus many of their partners, also offer advanced search pages that let you perform title searching without needing to know special commands. The Search Assistance Features page has a list of these advanced search pages.

Wildcards (*)

You can search for plurals or variations of words using a wildcard character. It is also a great way to search if you don't know the spelling of a word.

The * symbol is used as the wildcard symbol at several major search engines, as listed on the Search Features Chart. The format looks like this:

- *sing** finds singing and sings
- *theat** finds theater and theatre

Some of the search engines offering wildcard search also support what is called "stemming." That means they will find terms like "singing" even if you only enter "sing." This also means you may not need to use a wildcard symbol. See the Search Assistance page for more information on stemming.

Below are some important additional details about wildcard searching at specific search engines.

AOL Search

At AOL Search, the ? symbol serves as a wildcard and will replace any single character, such as:

s?ng matches sing, sang, song

This only works to find matches in AOL Search's Open Directory information. It does not work to bring back matches from Inktomi-powered listings, as explained further below.

Inktomi

Inktomi has two wildcard commands. The * symbol will match one or more characters, such as:

sing* matches sings, singers, singing

The ? symbol matches any single character, and you can use it more than once. For instance:

s?ng matches sing, sang, song

**??ng matches ring, rang, sing, sang and
any other four letter word ending in ng**

Both commands only work reliably at iWon, at the time of this writing. They fail to function properly at AOL Search, HotBot, MSN Search or LookSmart to bring up matches from within the Inktomi listings that they use.

They also do not appear to bring up matches in wildcard fashion from any of the other data sets these services use, with the exception of AOL Search (see AOL Search section, above).

Northern Light

Like Inktomi above, Northern Light has two wildcard commands. The * symbol will match one or more characters, while % is used to match just a single character.

Anchor Search

Some search engines allow you to search specifically within the "anchor" or "link" text that appears on a web page. For example, consider this example:

Click Here For The [Mars Exploration Web Site](#)

Notice the words "Mars Exploration Web Site" are all contained within the hyperlink? This is the anchor text or the link text.

Search engines that support anchor text searching are listed on the Search Features Chart.

Proximity

Some search engines let you indicate how close words should appear to each other. Most people do not need this type of control. Usually, phrase searching is all you need. If you still feel you need control over proximity, see the NEAR section of the Boolean Searching page.

More Power Search Commands

Several of the major search engines offer additional commands that allow you to search by media type, to search within ALT text or link text, and other types of queries. Several search engines offer dedicated help pages that explain these, as listed below. All links were verified as of March 11, 2003:

AllTheWeb: Query Language

http://www.alltheweb.com/help/faqs/query_language.html

AltaVista: Special Search Terms

http://www.altavista.com/help/adv_search/syntax

Google: Advanced Search Operators

<http://www.google.com/help/operators.html>

Teoma: Advanced Search Tips

<http://sp.teoma.com/docs/teoma/about/advsearchtips.html>



6. Search Assistance Features

The major search engines have many features meant to assist beginning searchers, and even professional researchers may find these aids helpful. This page summarizes and illustrates some of the major search assistance features that are available.

Related Searches

A related searches feature is designed to help users narrow in on what they are looking for. For example, let's say you searched for "mars." When the results appeared, you might also be shown some related searches links, such as "mission to mars" or "life on mars." If you selected one of these links, a new search would be conducted, using the words you clicked on. This can help you be more specific in your query, which often leads to better results.

AltaVista

Displays related searches near the top of the results page, next to the words "Others searched for."

AllTheWeb.com

Displays related searches near the top of the results page, next to the words "Narrow your search."

MSN Search

Displays related searches in the "Popular Topics" area below the search box, on the results page.

Yahoo

At Yahoo, related searches appear at the bottom of its results page.

Clustering

Have you ever done a search and found the top results all seem to come from one site? Clustering prevents this. Clustering generally allows only one or two pages per site to be represented in the top results. This means that you get more variety and a better chance of quickly finding something of interest. The section below highlights how this feature works at the major services that offer it.

AltaVista

AltaVista clusters listings so that no more than two pages per site appear in its results. If a second page from a particular web site is listed, it will be indented under the first page. To see more results from a site, select the "Additional relevant pages from this site" link, if it appears for a particular listing.

AllTheWeb.com

Clustering is on by default and will prevent more than two pages from the same web site from being displayed. It can be overridden by changing the Site Collapsing option on the Search Customization page (see the Customizing Results section below). You can also view more pages from any particular site listed by selecting the "more hits from" link that follows the listing.

Google

Google clusters so that no more than two pages per site appear in its results. If a second page is listed, it will be "indented" under the first page. To see more results from a site, select the "More results from" link that will appear below the second page listed.

HotBot

At HotBot, clustering is on by default. However, it only works within the listings provided by Inktomi. To turn off clustering, go to the advanced search page, then in the "Best Page Only" section, check the "Disable Best Page Only Filter" box. You can also view more pages from any particular site listed by selecting the "See results from this site only" link that follows the listing.

MSN Search

Clustering at MSN Search has to be enabled from its advanced search page. Look for the "Show one result per domain" option and select it to start clustering.

Find Similar

Did you find a web page in the search results that seemed perfect -- it was exactly what you were looking for? A "Find Similar" feature tells the search engine to seek out other pages that seem similar to those you like. The section below highlights how this feature works at the major services that offer it.

AltaVista

Click on the "Related pages" link that appears at the bottom of each listing.

AOL Search

Click on the "Show me more like this" option that appears at the bottom of each page listed. This takes you to where that page is categorized within the version of the Open Directory that AOL users. That can help you find similar web sites.

AltaVista

Click on the "Similar pages" link that appears at the end of each listing.

Stemming

Stemming is the ability for a search engine to search for variations of a word based on its stem. For example, entering "swim" might also find "swims" and maybe "swimming," depending on the search engine.

The Search Features Chart shows which search engines will do stemming by default and those that allow it to be switched on as an option. Some search engine specific notes are also below.

Inktomi

Inktomi-powered HotBot & MSN Search) provide stemming as an option. To enable it, go to the advanced search pages of each search engine, then

- At HotBot, check the "Enable Word Stemming" box.
- At MSN Search, see below.

MSN Search

This appears to be on permanently, at least for some queries. For example, a search for "run," "runs" and "running" in Oct. 2001 found the same results. Oddly, using the "Enable Stemming" box on MSN Search's advanced search page actually causes no results to appear.

Search Within

Ever do a search and still feel like you have too many results? Instead of trying a new search, you might have more luck narrowing down the set of matches you've already generated. Some search engines

make this easy through a "Search Within" feature. The section below highlights how this feature works at the major services that offer it.

AltaVista

After performing a search, check the "Search within these results" box under the search box, on the results page.

Google

After performing a search, click on the "Search within results" link that appears at the bottom of the results page, next to the search box, on the results page.

HotBot

After performing a search, check the "Search within these results" box that appears n, next to the search box, on the results page.

LookSmart

You cannot search within results generated from a keyword search on the LookSmart home page. However, if you navigate to any particular category, you can then search for matching sites that appear only within that category and its subcategories. To do this, when in a category, change the drop down box at the top of the category page from "the Web" to the second option, which will be the name of the category you are in.

Lycos

At Lycos, choose the "Search these results" option which appears next to the search box, at the top of the results page.

Yahoo

At Yahoo, you can't run a search and then search within it. But you can go to any category and then choose to search just within that section. Just look for the appropriate options near the search boxes that appear within the categories.

Spidered Version

It can be helpful to see the exact version of a web page that was presented to a search engine's spider. This is good for those times when a page no longer exists, allowing you to still find the information. It's also essential if you want to determine if a search engine spider was shown something different than what a human user sees. In fact, some webmasters may "pagejack" someone else's web page, feeding it to a search engine in hopes of attaining a good ranking. For more information about this, see the pagejacking section of the Search Engines and Legal Issues page.

Only Google allows you to see the actual page it spidered, through its "Cached" feature. When you search, a "Cached" link may appear below some pages that are listed. Click on this, and you'll be shown the page that was indexed, and any of your search terms will be highlighted.

You can also bring up the spidered version using Google's cache command. Simple enter the URL of a page after cache: and omitting the http:// prefix. For instance, to see the cached version of this page, you would enter this into Google:

cache:searchenginewatch.com/facts/assistance.html

Search By Language

Sometimes you may want to find pages written in a particular language. For example, you might want travel advice about Paris written in French. If you search for "paris," you'll probably get many pages written in English, since the city is spelled the same way in English and French. However, with a search by language option, you can specify that only pages written in French should be returned.

Searching by language isn't perfect. Search engines generally use dictionaries of terms specific to different languages to identify a page's language when spidering it. That means pages with content written in several different languages may not be categorized properly. Additionally, because this is an automated process, it can suffer from the mistakes that any automated system may have.

Below is how to search by language, at search engines that offer this feature.

AltaVista & AllTheWeb.com

Use the drop-down box that appears next to the search box on the home page and results page, to search in a particular language that's offered.

Google, Lycos

Use the advanced search page to search by language at these services.

MSN Search

Use the "Language" drop-down box on the advanced search page to search by language through Inktomi's crawler-based results.

Page Translation

Some search engines allow you to translate web pages they list into different languages. That's helpful if you see a page you are interested in but it is written in a language you don't understand. Below is how to do translation at search engines that offer this.

AltaVista & Lycos

Click on the "Translate" link that appears at the bottom of each listing.

Google

Click on the "Translate this page" link that appears next to the title of pages that are not in English, when using the main Google.com web site.

Porn Filter

Some search engines allow you to filter out pages that may lead to pornographic web sites or sites with content that might be considered offensive to some people. They generally do this by scanning pages for pornographic terms at the time they are indexed. "Block" lists and human review is also conducted.

Porn filters are not perfect, but they can be especially helpful if you are working with children and want to minimize the risk of them seeing sexually explicit or offensive terms in the results that appear.

For tips on enabling porn filters at several major search engines, please see the Kids Search Engines page within Search Engine Watch.

Customize Results

Wouldn't it be nice to see more than the 10 results at a time that are usually displayed at most search engines? Perhaps you might want to see just the titles of matching web pages. Some search engines allow you to customize your results in this way, usually via advanced search pages or from menu options. The Search Features Chart shows how many results are shown at each search engine by default and up to how many you can choose to see.

Below are links to special customization pages offered by some search engines:

AltaVista Customize Settings

<http://www.altavista.com/cgi-bin/query?pref=res&stype=stext&Translate=on&sc=on>

AllTheWeb

<http://www.alltheweb.com/customize>

Google

<http://www.google.com/help/customize.html>

MSN Search

<http://search.msn.com/Preference.asp>

Sort By Date

Sort by date sounds like a great idea, but there are big problems with dates on the web. Some web servers report incorrect dates or no dates at all.

For instance, Go's engineers estimated in 1998 (back when the search engine still existed) that only 70 percent of web servers returned the correct date, while 20 percent reported the current date, regardless of when the page was created or changed. The remaining 10 percent of the time, the web servers reported no date at all. Northern Light also found similar problems, as outlined in this article: Northern Light Adds Search Functions, Freshens Index.

Still, date sorting is a nice feature to have, and one that many professionals want. When you choose the option, they list pages with newer dates first. At MSN Search, you'll find this option on the advanced search page. Use the "Sort equally relevant results by" box.

Keep in mind that often when people want to sort by date, they are often trying to get the latest information on a news topic. In these case, it is better to use a news search engine. See the News Search Engines page for a list of these helpful services.

Also see:

It's Tough to Get a Good Date with a Search Engine

SearchDay, June 5, 2002

<http://searchenginewatch.com/searchday/article.php/2160061>

Search engines have problems with calendar information. Bottom line: you may end up searching for dates in all the wrong places.

Date Range

Some search engines let you restrict a search so that only pages within a particular date range are displayed. This feature can suffer from the fact that web page dates can be unreliable, as described above. However, it can also be useful, especially as a means of determining how fresh a search engine's listings are.

For example, if you restrict a search to find pages less than a month old and don't get any matches, you have a pretty good idea that the search engine's listings are out of date.

See the Search Features Chart for which search engines offer this option on their advanced search pages.

Date Display

Along with the page description, some search engines show the date when a web page was created or modified. As noted above, these dates may not always be reliable. However, they do provide a useful clue as to how fresh or stale a search engine's listings are. Thus, search engines that show a date deserve praise for doing so.

See the Search Features Chart for which search engines display a date. When no date is reported, these search engines above will instead display the date the page was spidered.

Northern Light is an exception. In these cases, it won't report a date at all.

Directories don't spider pages, but they can display when a listing was manually added or updated, if desired.

Advanced Search Page

Most of the search services have advanced search pages designed to let you have more control over your search or to guide you into creating more complex queries. However, these are sometimes hard to find. To help, here are links that will take you directly to each service's advanced search pages. Links were verified as of March 11, 2003:

AllTheWeb: Advanced Search

<http://www.alltheweb.com/advanced>

AltaVista: Advanced Web Search

<http://www.altavista.com/web/adv>

Google: Advanced Search

http://www.google.com/advanced_search

HotBot: Advanced Search

<http://www.hotbot.com/adv.asp>

NOTE: By default, this will be the advanced search page to get Inktomi results from HotBot. Change your search engine choice by using the options under the search box, and the advanced search page will change to suit the choice you've selected.

Lycos: Advanced Search Filters

<http://search.lycos.com/adv.asp>

MSN Search: Advanced Search Options

<http://search.msn.com/advanced.aspx>

Teoma: Advanced Search

<http://s.teoma.com/AdvancedSearch>

Yahoo: Advanced Search

<http://search.yahoo.com/search/options>

Help Pages

Many services provide more information on advanced searching techniques and features within their help pages. The links below will take you directly to them. Links were verified as of March 11, 2003:

AllTheWeb: Help & FAQ

<http://www.alltheweb.com/help/>

AltaVista: Search Help

<http://www.altavista.com/help/search/default>

AOL Search: Search Help

<http://search.aol.com/aolcom/help.jsp>

Ask Jeeves: Help

<http://sp.ask.com/docs/help/>

Google: Help Central

<http://www.google.com/help/>

HotBot: Help

http://help.lycos.com/LycosHelp/help/hotbot/htdocs/hotbot_1_help.htm

Lycos Help

http://help.lycos.com/LycosHelp/help/search/htdocs/search_1_help.htm

Teoma: Basic Search Tips

<http://sp.teoma.com/docs/teoma/about/searchtips.html>

Yahoo: Search Help

<http://help.yahoo.com/help/us/ysearch/>



7. Search Features Chart

Search Engine Math Commands

Updated: March 11, 2003

(See Search Engine Math and Power Searching For Anyone for more details)

Covers: AllTheWeb, AltaVista, AOL Search, Ask Jeeves, Google, HotBot, Lycos, MSN Search, Teoma and Yahoo. HotBot references are only for its Inktomi-powered results.

Command	How	Supported By
Must Include Term	+	All
Must Exclude Term	-	All
Must Include Phrase	" "	All
Match All Terms	Automatic at	All
Match Any Terms	Via Advanced Search	AllTheWeb, AltaVista, Google, Lycos, MSN Search, Teoma, Yahoo <i>(HotBot offers but failed to work when tested)</i>
	OR	AltaVista, AOL Search, Ask Jeeves, Google, HotBot, MSN Search, Teoma, Yahoo

		(<i>must be done in ALL CAPS</i>) AllTheWeb, Lycos (<i>only works for two words</i>)
--	--	---

NOTE: By default, all the major search engines named above will match ALL of the terms you enter into a search box. This means that it is not necessary to use the + symbol in front of a particular word, though it won't hurt if you do so.

Power Searching Commands
(See Power Searching For Anyone for more details)

This section is being updated to cover the major crawler-based search engines of AllTheWeb, AltaVista, Google, Inktomi and Teoma. These crawlers also provide results to other search engines, so you may find commands on them work with their partners. Major partnerships are as follows: AllTheWeb (Lycos), Google (AOL Search, Yahoo), Inktomi (HotBot), MSN Search (Inktomi), Teoma (Ask Jeeves)

Command	How	Supported By
Title Search (Updated March 11, 2003)	title:	AltaVista, AllTheWeb, Inktomi
	intitle:	Google Teoma
	allintitle:	Google
Site Search	host:	AltaVista
	site:	Excite, Google (Netscape, Yahoo)
	url.host:	AllTheWeb, Lycos (for AllTheWeb results only)
	domain:	Inktomi (HotBot, iWon, LookSmart)
	none	AOL, Direct Hit, HotBot, LookSmart, Lycos, MSN, Netscape, Northern Light, Open Directory, Yahoo
URL Search	url:	AltaVista, Excite, Northern Light

	url.all:	AllTheWeb, Lycos (for AllTheWeb results only)
	allinurl: inurl:	Google
	originurl:	Inktomi (AOL, GoTo, HotBot)
	u:	Yahoo
	none	AOL, Direct Hit, HotBot, LookSmart, MSN Not yet updated, but may be still correct: Open Directory
Link Search	link:	AltaVista, Google, Northern Light
	linkdomain:	Inktomi (AOL, HotBot, iWon, MSN) (NOTE: measures links to entire domains)
	link.all:	AllTheWeb, Lycos (for AllTheWeb results only)
	none	AOL, Direct Hit, Excite, HotBot, LookSmart, Northern Light Not yet updated, but may be still correct: Netscape, Yahoo (n/a)
Wildcard	*	AltaVista, Inktomi (iWon), Northern Light Not yet updated, but may be still correct: Yahoo
	?	AOL Search, Inktomi (iWon)
	%	Northern Light

	none	AllTheWeb, Direct Hit, Excite, Google, HotBot, LookSmart, Lycos, MSN (MSN's help says it offers wildcard, but it failed to during testing)
Anchor Search	anchor:	AltaVista
	None	AllTheWeb, AOL Search, Direct Hit, Excite, Google, Inktomi, HotBot, Lycos

NOTE: The commands above are primarily useful when dealing with crawler-based search engines. "None" indicates any crawler-based or human-powered search engine that creates its own listings but which does not provide a particular command for searching within those listings. It may also indicate a portal that that outsources for its listings and which lacks a single command to work across the multiple datasets it uses.

Search Assistance Features
(See the Search Assistance Features page for more details)

Feature	Offered By
Related Searches	AltaVista, AllTheWeb, Excite, HotBot, Lycos, MSN, Yahoo Not yet updated, but may be still correct: iWon
Clustering	AltaVista, AllTheWeb, Excite, Google, HotBot, MSN, Northern Light
Find Similar	AltaVista, AOL Search, Google
Stemming	AOL Search, Direct Hit, HotBot, Inktomi (HotBot, MSN)
Search Within	AltaVista, Google, HotBot, Lycos
Spidered Version	Google
Search By Language	AltaVista, AllTheWeb, Excite, Google, HotBot, Lycos, MSN, Northern Light
Page Translation	AltaVista, Google, Lycos
Porn Filter	AltaVista, AllTheWeb, Google
Porn Warning	HotBot, MSN, Northern Light

Customization & Display Features
(See Search Assistance Features page for more details)

Feature	Supported By
Number Of Listings Shown (10 unless noted)	AltaVista, AllTheWeb, AOL Search (5), Direct Hit, Excite, Google, HotBot, LookSmart (15), Lycos, MSN (15), Northern Light Not yet updated, but may be still correct: iWon, Netscape, Yahoo (20)
Ability To Increase Number Of Listings?	AltaVista, AllTheWeb, Excite, Google, HotBot, MSN Not yet updated, but may be still correct: Yahoo
See 20 Results	AltaVista, AllTheWeb, Excite, Google, HotBot, MSN Not yet updated, but may be still correct: Yahoo
See 50 Results	AltaVista, AllTheWeb, Excite, Google, HotBot, MSN Not yet updated, but may be still correct: Yahoo
See 100 Results	AllTheWeb, Google, HotBot, Not yet updated, but may be still correct: Yahoo
Sort By Date	MSN Search, Northern Light
Date Range	AltaVista, Google, HotBot, MSN, Northern Light Not yet updated, but may be still correct: iWon, Yahoo
Date Displayed?	AltaVista, HotBot (for Inktomi results), Northern Light
Display Titles Only?	AltaVista, Excite, HotBot (URLs only option), MSN
Other Major Customize Options	AltaVista, AllTheWeb, Google

Boolean Commands
(See Boolean Searching page for more details)

Command	How	Supported By
Or	OR	AltaVista, AOL Search, Excite, Google, Inktomi (HotBot, MSN), Lycos, Northern Light
	None	AllTheWeb, Direct Hit, LookSmart, Not yet updated, but may be still correct: Yahoo
And	AND	AltaVista, AOL Search, Excite, Inktomi (HotBot, MSN) Lycos, Northern Light
	None	AllTheWeb, Direct Hit, Google, LookSmart Not yet updated, but may be still correct: Yahoo
Not	NOT	AOL Search, Excite, Inktomi (HotBot), Lycos, Northern Light

	AND NOT	AltaVista, Inktomi (MSN) Not yet updated, but may be still correct: Netscape
	None	AllTheWeb, Direct Hit, Google, LookSmart, Not yet updated, but may be still correct: Yahoo
Nesting	()	AltaVista, AOL Search, Excite, Inktomi (MSN), Northern Light
	None	AllTheWeb, Direct Hit, Google, Inktomi (HotBot), LookSmart, Lycos Not yet updated, but may be still correct: Yahoo
Near	NEAR	AltaVista (10 words), AOL Search (specify number), Lycos (25 words)
	None	AllTheWeb, Direct Hit, Google, Inktomi (HotBot, MSN), LookSmart
Notes		
<p>At AltaVista, Boolean only works on advanced search page. At Excite, Google & MSN, Boolean commands must be in UPPERCASE At Inktomi-powered services, set menu to "Boolean"</p>		



8. Search Toolbars and Utilities

There are a variety of ways you can search and navigate the web without first going directly to a search engine. Search toolbars provide direct access to search engines from within your browser, while meta search utilities make it easy to search more than one search engine at the same time. Search companions often help you "discover" information as you browse pages on the web.

Search Toolbars From Major Search Engines

AltaVista Toolbar

<http://www.altavista.com/toolbar/default>

Provides access to AltaVista web, news and multimedia search, page translation, term highlighting and pop-up blocking. (Review: [AltaVista Introduces Search Toolbar](#), Aug. 18, 2003)

Ask Jeeves Toolbar

<http://sp.ask.com/docs/toolbar/>

In addition to searching Ask.com, the Jeeves toolbar lets you limit your search to news, dictionary, stock market, weather, events, maps, and the Ask Jeeves Kids web sites.

Dogpile Search Toolbar

<http://www.dogpile.com/info.dogpl/tbar/>

Consolidates a variety of useful web and specialized searches into a single, easy to use interface. (Review: [Dogpile's Toolbar Fetches More than Search Engine Results](#), June 25, 2003)

Google Deskbar

<http://toolbar.google.com/deskbar/>

Provides the ability to search with Google from the taskbar within Windows. In other words, you can search without having to be in your browser.

Google Toolbar

<http://toolbar.google.com>

Special toolbar for Internet Explorer users that puts a Google search box right into your browser. In addition, you can use it to see the "PageRank" popularity score of any page you are viewing, search within the particular site you are viewing, see a previous of "cached" copy of any dead pages, find pages similar to the one you are viewing, block pop-ups and more. It loads within seconds and is well worth adding to your browser. Using the toolbar with PageRank enabled sends some information back to Google. If that concerns you, choose the "Install Without Advanced Features" option. Highly-recommended. (Review: [Google Releases Updated Search Toolbar](#), June 30, 2003)

HotBot Quick Search Deskbar

<http://www.hotbot.com/tools/>

Provides instant access to HotBot and is jam-packed with other useful tools and goodies as well. (Review: [HotBot's Swiss Army Knife Deskbar](#), June 24, 2003)

MSN Toolbar

<http://toolbar.msn.com/>

Provides the ability to search the web using MSN Search, perform site specific searches, term highlighting and pop-up blocking.

Teoma Search Bar

<http://sp.ask.com/docs/teoma/toolbar/>

Provides direct access to Teoma's search results, offers search term highlighting on the pages you visit, allows access to an online dictionary and provides the ability to email any web page you view. (Review: [Teoma Offers Free Search Toolbar](#), June 25, 2002)

Yahoo Companion

<http://companion.yahoo.com/>

Search Yahoo, access Yahoo Mail, check on stocks and more via a toolbar within your browser.

Other Search Toolbars

Advanced Searchbar

<http://www.advancedsearchbar.com>

Provides access to 60 search engines, plus push button access to disk utilities, webpage translation, a calculator and other features. Be forewarned: it's a 1MB download.

Alexa

<http://download.alexa.com/>

Alexa provides access to Google search results plus does a wonderful job of suggesting interesting sites based on the ones you are visiting -- and much more.

CNET SearchBar

<http://www.search.com/guides/sb/main.html>

This customizable toolbar is made exclusively for searching on CNET sites including download.com, news.com, search.com and shopper.com for technology news, and as a shopping search tool for products from anywhere on the Web.

Copernic Meta Toolbar

<http://www.copernic.com/en/products/meta/>

Lets you meta search the web via a toolbar integrated into your browser.

Dave's Quick Search Taskbar Toolbar Deskbar

<http://www.dqsd.net/>

Like the Google Toolbar, this gives you access to Google without having to first visit the Google web site. Unlike the Google Toolbar, it has a number of other options built into it, such as the ability to query other search engines such as AllTheWeb and Teoma, as well as a range of specialty search services. This can be done using a menu-system or by special prefix codes in front of your query. The toolbar also has a built-in calculator, currency converter, translation tool and other features. This installs into your Windows taskbar, rather than into your browser. It also requires a 400K download.

GGSearch

<http://www.frysianfools.com/ggsearch/>

GGSearch is designed to provide toolbar access to the many specialized searches offered by Google, including Google groups posts, images, links, Uncle Sam, linux, news, BSD, Microsoft, Google answers, stocks, froogle, some Google labs tools, and others.

Groowe Toolbar

<http://www.groowe.com>

Here's a toolbar that gives you easy access to searching Google and many other search engines, as well. With a click, you can query Yahoo, Teoma, AllTheWeb, AltaVista, MSN Search and others. In addition, it makes it easy to perform specialized searches with many of the search engines it supports. Get images back from AllTheWeb, or search only against Yahoo's human directory, or get news results back from AltaVista. It's a quick download and highly-recommended.

Googlebar

<http://googlebar.mozdev.org/>

Got Netscape but you want the Google Toolbar, which is only for Internet Explorer? Then check out the volunteer-created Googlebar for Netscape.

Gophoria

<http://www.gophoria.com/>

This simple utility allows you to highlight any word or words on a web page, then click to do a Google search, or to get dictionary or thesaurus definitions. Fast 100K download.

Macintosh SearchGoogle.service

<http://gu.st/proj/SearchGoogle.service/>

A simple background service to allow you to select text in (nearly) any application and press "Shift-Apple->" to launch a Google search for that text.

metaEureka A-Toolbar

<http://www.metaeureka.com/download.shtml>

Lets you get results from metaEureka, a meta search engine that hits several major web-wide search engines. Alternatively, you can also choose to search against specific search engines, as well. Beyond searching, you can use the toolbar to prescreen your email for spam, translate words into different languages, do dictionary lookups, check the time in various countries, convert currencies and much more. You can also get information about a particular URL, count links to that URL from various search engines, do a basic position check and get a keyword density report for a particular page (single words only, not phrases). Among network tools is the ability to telnet, ping, traceroute, do DNS and WHOIS lookups.

Trellian Toolbar

<http://www.trellian.com/toolbar/>

The Trellian toolbar allows you to search multiple engines, returning up to nine result pages for each engine. It also offers some web site development and search engine optimization tools, such as a PPC search tab that provides easy search of the major pay per click engines.

UltraBar

<http://www.ultrabar.com/>

Like Groowe, UltraBar gives you easy access to multiple search engines plus allows you to add your own. It also provides many of the features you'll find in the Google Toolbar, such as term highlighting and jumping to keyword buttons. Highly-recommended. (Review: [Beyond the Google Toolbar](#), July 15, 2002)

Vivisimo Toolbar / MiniBar

<http://vivisimo.com/toolbar/toolbar-download.html>

<http://vivisimo.com/toolbar/minibar-download.html>

Lets you tap into Vivisimo meta search results from your browser. The second URL lets you load a "mini" version of the toolbar, helpful for those who already have many other search toolbars installed.

Meta Search Utilities

Similar to [metacrawlers](#), the meta search utilities below send your query to more than one search engine at the same time. But unlike online meta search engines, many of the software packages can also retrieve actual web pages, allowing you to sort results, perform more analysis, eliminate dead links, save searches and more.

Copernic Agent Basic

<http://www.copernic.com/en/products/agent/basic.html>

Copernic Agent is a meta search engine, invisible web explorer, online research assistant and extensive tool box, all combined into an elegant, easy to use toolbar. (Review: [Copernic Agent: Jack of All Searches](#), July 23, 2003)

InfoGrid Internet Explorer Bar

<http://www.infogrid.com/homepage.htm>

Allows Internet Explorer users to change the search button from its default behavior of using MSN Search to instead meta searching at several major search engines via the InfoGrid site. Also provides access to news search services and other resources.

Apple Sherlock

<http://www.apple.com/macosx/jaguar/sherlock.html>

If you have a Mac with a current OS, then you've got meta search and more built in. This page at Apple explains more.

SearchWolf

<http://www.trellian.net/search/>

Allows you to search multiple search engines from your desktop.

WebFerret

<http://www.ferretsoft.com/>

WebFerret lets you search across the web. The free version has banner ads, while the paid version does not.

Subject Search Spider

<http://www.kryltech.com/spider.htm>

Metasearch utility that queries the selected set of search engines in over 35 languages. Has text highlighting and other features.

Search Companions & Discovery Tools

The search companions and navigation tools below are designed to help you find or discover information, often in non-traditional ways.

AdSubtract

<http://www.intermute.com/adsubtract/>

This software removes paid listings from search engine result pages, along with banner ads and pop-ups. Paid inclusion listings are not removed.

Annotate Net Radar

<http://www.annotate.net/html/download/download.py>

Similar to Alexa [above](#), this browsing companion suggests content from its partners that is related to what you are viewing.

Grokker

http://www.groxis.com/service/grok/g_products.html

Grokker lets you "fly" through search results grouped into different topics. It can be a helpful way to easily maneuver yourself into the right results, when there's more than one set of answers to your question.

GuruNet

<http://www.gurunet.com/>

First known as GuruNet, then Atomica, then revived in early 2003 as GuruNet once again! This tool lets you highlight words or Alt-Click on them in a document and then send the word to an online dictionary or thesaurus. (Review: [GuruNet: A Handy Information Magnet](#), June 11, 2003)

HydraLinks

<http://www.hydralinks.com/hl1001.htm>

HydraLinks is a simple utility that lets you manipulate and save search results to a customizable list, speeding up your searching and letting you easily share results with others. (Review: [Speed Up Your Searching with HydraLinks](#), June 19, 2003)

TouchGraph GoogleBrowser

<http://www.touchgraph.com/TGGoogleBrowser.html>

Enter a URL, then see a pretty picture showing you interlinks between that URL and other sites based on Google's "related" command (also called [Similar Pages](#)). However, don't misinterpret this as an illustration of actual site interlinking. Pages related to a URL are not necessarily pages that link to that URL. Instead, they are simply pages that Google deems to be similar in terms of content.

UCmore

<http://www.ucmore.com/>

This tool installs into your browser and then shows information related to the page you are viewing. It's billed as "contextual search," but it's more a discovery tool along the lines of Alexa, above. In other words, you don't search using it, but you may discover new sites related to those you like. It will suggest directory categories and actual web sites from the Open Directory that are deemed related to the page you are viewing. So, if you are listed in the Open Directory, you may show up in front of those using this tool. "Elite" listings get you a guaranteed placement -- these come from being the top bidder with FindWhat. The tool also claims not to be spyware, but content owners may not be happy to have it suggesting other sites to their visitors. Similar criticisms have been levied against Alexa, in the past.

Articles About Search Toolbars**Toolbars: Trash or Treasures?**

Online, Jan/Feb 2004

<http://www.infoday.com/online/jan04/OnTheNet.shtml>

A guide to the many search toolbars now littering our desktops, looking at what's offered, options and possible problems.

Designs on desktop search

News.com, Jan. 12, 2004

<http://news.com.com/2100-1032-5138715.html>

Now that Google's jumped out of the browser and into the taskbar, Microsoft and Yahoo are thinking of doing the same.

Beyond Google: Narrow the Search

AP, Jan. 4, 2003

<http://www.wired.com/news/technology/0,1282,61783,00.html>

A look at tools that automatically categorize and sometimes even visually present search results. There's nothing new about the concept. Northern Light did auto-categorization back in 1997, and AltaVista had a visualization tool in the same year, to name only some examples. They never caught on, but perhaps the new crop may have more luck -- though even these "new" tools mentioned are all more than a year old.

Going Deeper than Google

Fortune, Dec. 16, 2003

<http://www.fortune.com/fortune/fastforward/0,15704,563090,00.html>

Review of Grokker, which lets you "fly" through results found by Google and other search engines.

Monetizing Graphical Search

InternetNews.com, Dec. 15, 2003

<http://www.internetnews.com/IAR/article.php/3289001>

Trying to shake people out of the 10 textual search results format, a new version of Grokker provides the ability to "fly" through results items of interest, while Vivisimo has released a new toolbar to let you access "clustered" search topics from its acclaimed meta search engine.

AdSubtract to snip paid search results

IDG, Dec. 8, 2003

http://www.infoworld.com/article/03/12/08/HNadsubtract_1.html

AdSubtract removes paid listings from search results as well as blocks other types of ads.

On the Google Deskbar

The Register, Dec. 4, 2003

<http://www.theregister.co.uk/content/6/34338.html>

The Register finds that, much as it says it's depressed to admit, the Google Deskbar is a good thing. Review of features and how changing habits from browser-based searching to taskbar searching worked for the writer.

The Google Deskbar

Pandia, Nov. 7, 2003

<http://www.pandia.com/sw-2003/45-deskbar.html>

Pandia provides an overview of features in the new Google Deskbar, especially the ability to use keyboard shortcuts to search and reasons why you may want search outside your browser.

Google Plops Its Search on the Desktop

InternetNews.com, Nov. 6, 2003

<http://www.internetnews.com/IAR/article.php/3105641>

Has comments from me about Google's move from the browser and onto the desktop.

Google's Popular Toolbar

New York Times, Oct. 30, 2003

<http://www.nytimes.com/2003/10/30/technology/circuits/30goog.html>

Google's Toolbar has a great pop-up blocking feature, but that also means pop-ups you want can go unnoticed. If you're using it, remember that overriding the pop-up (easily done) may be the solution to a web access problem you may be having. And designers, as this article covers, take note that you may need to consider dumping the pop-ups.

Google And The Big Brother Nomination

The Search Engine Update, April 2, 2003

<http://www.searchenginewatch.com/subscribers/articles/03/article.php/2174961#spyware>

With "advanced features" on, the Google Toolbar sends information about the page you are viewing in order to update its PageRank meter. When you installed the Google Toolbar, this fact was highlighted and an alternative for those concerned with privacy issues was offered. Nevertheless, some people still may not have realized that the toolbar monitors page viewing in this way. This article examines accusations that the toolbar is "spyware."

Google Toolbar Alternatives

Pointeronline, Jan. 24, 2003

<http://www.poynter.org/column.asp?id=32&aid=18194>

Use Netscape? Here are tips on accessing search engines directly from the URL box.

Clearing Your Search History From Google And Other Search Engines

The Search Engine Report, Jan. 7, 2003

<http://www.searchenginewatch.com/sereport/article.php/2165181>

Ever gone to Google or another search engine using the Internet Explorer browser and noticed that some past searches you've performed were displayed below the search box? Don't blame (or thank) the search engine. It's your browser that's keeping the record, and it's one you may find yourself wishing to delete, from time to time. This article explores the procedures involved.

Search Tool & Utility Reviews

Within Search Engine Watch, this page has older reviews of software-based search tools that you may find interesting.

Removing Adware/Spyware

Readers sometimes ask how their "default" search engine got changed in Internet Explorer or why they are getting "pop-up" search results from a search engine they've never heard of.

The answer is that there may be some program installed on your computer, perhaps without you knowing it, that made these changes.

To correct this behavior, try running Ad-Aware on your system (it's free) to see if there's any "adware" present. Counterexploitation also has a good page listing some programs that might be causing your problems



9. Boolean Searching

Boolean search commands have been used by professionals for searching through traditional databases for years. Despite this, they are overkill for the average web user. The commands described on the Search Engine Math page provide the same basic functionality as Boolean commands and are also supported by all the major search services. If you are new to searching, start off learning how to search better by first reading the Search Engine Math page, rather than trying to learn Boolean commands. I'm certain you'll find it easier.

In fact, many professionals might benefit by abandoning Boolean commands when using web search engines. But since there is a comfort level in using what is already familiar, this page covers how Boolean commands are implemented at the major search services. It assumes you are already familiar with Boolean searching, although some resources that provide further help appear at the end of the page.

OR

The Boolean OR command is used in order to allow any of the specified search terms to be present on the web pages listed in results. It can also be described as a Match Any search. You use the command like this:

ireland OR eire

Search engines that support OR are shown on the Search Features Chart. For those that don't, see their advance search pages, where an option to search for any of your terms is often available.

Also be aware that some search engines perform an OR search by default, as shown in the Match Any section of the Power Searching For Anyone page. Search engine specific notes are below:

AOL Search

OR failed to work correctly at the time this page was written. For instance, a search for "ireland OR eire" failed to yield a much larger set of results that should have appeared when compared to "ireland AND eire".

Google

OR will not work to find different phrases, such as "bill clinton" OR "hillary clinton"

AND

The Boolean AND command is used in order to require that all search terms be present on the web pages listed in results. It can also be described as a Match All search. You use the command like this:

clinton AND dole

Search engines that support AND are shown on the Search Features Chart. For those that don't, using the + symbol is generally a good alternative.

Also be aware that some search engines perform an AND search by default, as shown in the Match All section of the Power Searching For Anyone page. Search engine specific notes are below:

AOL Search

When using AND, you may find a slightly different number of documents will be retrieved when compared to using the + symbol. This appears to be because AOL Search will check both its own listings and Inktomi listings when using AND but only Inktomi listings when using the + symbol.

NOT

The Boolean NOT command is used in order to require that a particular search term NOT be present on web pages listed in results. It can also be described as an Exclude search. You use the command like this:

clinton NOT dole

Search engines that support NOT are shown on the Search Features Chart. For those that don't, using the - symbol is generally a good alternative. Search engine specific notes are below:

AOL Search

When using NOT, you may find a slightly larger number of documents will be retrieved when compared to using the + symbol or no commands at all. This shouldn't happen, but it did at the time this page was written.

NEAR

The NEAR command is used in order to specify how close terms should appear to each other. You use the command like this:

moon NEAR river

Please consider whether you really need to control proximity within your searches. Most search engines will try to find the terms you indicate next to each other, or within close proximity to each other, by default. Also, all of the search engines support phrase searching through use of quotation marks. See Search Engine Math page for more information about phrase searching.

Search engines that support NEAR are shown on the Search Features Chart. Search engine specific notes are below.

AltaVista

NEAR means that terms will appear within 10 words of each other.

AOL Search

You can control the exact number of words apart by using NEAR/#. For instance, NEAR/5 would mean the terms should be five words apart. If you don't specify a number, then the terms must appear right next to each other.

Lycos

NEAR means that terms will appear within 25 words of each other. Lycos also supports an extensive range of other adjacency commands. See the site's help pages for Boolean searches for further details.

Nesting ()

Nesting allows you to build complex queries. You nest queries using parentheses, like this:

impeachment AND (clinton OR johnson)

Search engines that expressly say that they support nesting are shown on the Search Features Chart. I have not tried to verify this information. Be aware that the major search engines may process nested queries differently than each other.

Other Notes

AltaVista

Boolean searching can only be done from the advanced search page, as listed on the Search Assistance Features page.

Excite, Google & MSN

Boolean commands must be in uppercase. That's why I show them that way on this page. If you always use uppercase, you won't have problems when going between services.

Inktomi- services (HotBot, MSN Search)

You must set the menu option on the home page or advanced search page to "Boolean phrase" when using Boolean commands.

Lycos

Lycos says it supports many Boolean commands, and I haven't verified these, because of the difficulty of determining exactly which datasets might be processed. In addition, AllTheWeb -- which powers many of the search results at Lycos -- doesn't support Boolean. This makes it unclear how Lycos itself might then do this.

More Resources

Full Boolean at AlltheWeb

Search Engine Showdown, Jan. 21, 2003

<http://www.searchengineshowdown.com/newsarchive/000629.shtml>

Recaps new boolean support at AllTheWeb.com, along with other changes.

Boolean Searching on the Internet

<http://library.albany.edu/internet/boolean.html>

An easy-to-read and comprehensive guide to Boolean searching on web-wide search engines.

The Internet Search-Off

Searcher, Feb. 1998

<http://www.infotoday.com/searcher/feb98/story1.htm>



10. Search Engine Glossary

Boolean search: A search allowing the inclusion or exclusion of documents containing certain words through the use of operators such as AND, NOT and OR.

Concept search: A search for documents related conceptually to a word, rather than specifically containing the word itself.

Full-text index: An index containing every word of every document cataloged, including stop words (defined below).

Fuzzy search: A search that will find matches even when words are only partially spelled or misspelled.

Index: The searchable catalog of documents created by search engine software. Also called "catalog." Index is often used as a synonym for search engine. Index is commonly pluralized as "indices." However, Search Engine Watch instead uses the alternative plural form "indexes."

Keyword search: A search for documents containing one or more words that are specified by a user.

Phrase search: A search for documents containing a exact sentence or phrase specified by a user.

Precision: The degree in which a search engine lists documents matching a query. The more matching documents that are listed, the higher the precision. For example, if a search engine lists 80 documents found to match a query but only 20 of them contain the search words, then the precision would be 25%.

Proximity search: A search where users to specify that documents returned should have the words near each other.

Query-By-Example: A search where a user instructs an engine to find more documents that are similar to a particular document. Also called "find similar."

Recall: Related to precision, this is the degree in which a search engine returns all the matching documents in a collection. There may be 100 matching documents, but a search engine may only find 80 of them. It would then list these 80 and have a recall of 80%.

Relevancy: How well a document provides the information a user is looking for, as measured by the user.

Search Engine: The software that searches an index and returns matches. Search engine is often used synonymously with spider and index, although these are separate components that work with the engine.

Spider: The software that scans documents and adds them to an index by following links. Spider is often used as a synonym for search engine.

Stemming: The ability for a search to include the "stem" of words. For example, stemming allows a user to enter "swimming" and get back results also for the stem word "swim."

Stop words: Conjunctions, prepositions and articles and other words such as AND, TO and A that appear often in documents yet alone may contain little meaning.

Thesaurus: A list of synonyms a search engine can use to find matches for particular words if the words themselves don't appear in documents.

Downloaded November 29, 2004