



webreference suggest cool new headlines search contents

[home](#) / [web](#) / [articles](#) / [search](#) / [features](#)



## Search Engines

### III. Getting the Most Out of Your Search Engine

#### Search Engine Features

Web location services typically specialize in one of the following: their search tools (how you specify a search and how the results are presented), the size of their database, or their catalog service. Most engines deliver too many matches in a casual search, so the overriding factor in their usefulness is the quality of their search tools. Every search engine I used had a nice GUI interface that allowed one to type words into their form, such as "(burger not cheeseburger) or (pizza AND pepperoni)." They also allowed one to form Boolean searches (except Hotbot as of 7/1/96, which promises to install this feature later), i. e. they allowed the user to specify combinations of words. In Alta Vista and Lycos, one does this by adding a "+" or a "-" sign before each word, or in Alta Vista you can choose to use the very strict syntax Boolean "advanced search." This advanced search was by far the hardest to use, but also the one most completely in the user's control (except for OpenText). In most other engines, you just use the words AND, NOT, and OR to get Boolean logic.

By far the best service for carefully specifying a search was Open Text. This form has great menus, making a complex Boolean search fast and easy. Best of all, this service permits you to specify that you want to search only titles or URLs. But then there's Alta Vista's little known "keyword" search syntax, now as powerful as OpenText, but not as easy to use. You can constrain a search to phrases in anchors, pages from a specific host, image titles, links, text, document titles, or URLs using this feature with the [syntax](#) keyword:search-word. There is an additional set of keywords just for searching Usenet. (To my knowledge, Alta Vista's keywords were undocumented before 7/19/96, so tell your friends you heard it here first!)

Which Search Page Should I Use When, and How?		
Use . . .	If You . . .	Using the Feature . . .
Lycos	have no good ideas for specific search strategies	best test results for broad search terms
" "	want to find someone's e-mail	People Finder.
Magellan	have more than one broad search word, or can't pick a site from Lycos' summaries.	best available results summaries.
" "	want interactive news/ want details on today's headlines.	news with links to related sites.
OpenText	want to search only document title or perform complex searches	title search specification, best advanced search interface.
Alta Vista	are hunting for an image	image:search_word syntax.
" "	want to find all the links to your page	+link:your_site -url:your_site syntax.
Yahoo!	want the best national and international news	Reuters world headlines.
" "	want a dictionary or other reference source	Dictionaries or Reference Libraries.

What could really make engines with large data bases shine, however, would be an improvement in the way they rank and present results. All engines I tested had ranking schemes that were not well documented, based on how many times your search words were mentioned, whether or not they appeared early in the document, whether or not they appeared close together, and how many search terms were matched. I did not find the ranking schemes very useful, as relevant and irrelevant pages frequently had the same scores.

### Catalogs

I have only been disappointed by catalog services. In practice, they seem to aim for the lowest common denominator, and reflect very little thought to how and when they might be useful instead of search engines. All the ones I tested were directed toward novices and favored popular commercial sites. I would have thought they would be very good for finding software at least, but this was not the case. See the example below trying to find Web server related software.

### Advanced or Boolean Queries

Making queries very carefully in Boolean terms to narrow a search rarely produces useful results for me (but see below). In practice, other ways of specifying a search besides detailed logic are much more useful. Specification of exact vs. approximate spelling, specification that search terms must appear as section headings or URLs, using more keywords, and just specifying the language of the document would have been more valuable in all of my search examples.

### Example: Eliminating Unwanted Matches

The exception to this is the **AND NOT operator** - it is essential to exclude unwanted but close matches when they outnumber the desired matches. An example of when to use this operator is given by the problem of finding information on growing apples, because you will be deluged by information on Apple computers. With enough work, you can start to see apples with stems, not cords, but it isn't easy. Using Alta Vista, "+apple -mac\* -comp\* -soft\* -hard\* -vendor" got me information on the Payson-Santaquin apple farming region and a federal apple agriculture database on the first page of results.

### Useful Non-Search Goodies

#### E-mail address books:

Most engines allow you to search for someone's name if you quote it "John Q. Webhead", but you have to be careful about exact spelling, use of initials, etc. Try the genuine special purpose [people finder](#) by Lycos (it only works for people on the Web or e-mail).

#### News Services:

Yahoo! has the best news, in my humble opinion, as they have [Reuters international news headlines](#). Most other news are ultra-brief summaries which read like "MacPaper." Another service that is very exciting is [Magellan's news](#), which provides links to learn more about subjects. A recent Israeli news item included links to a Likud party page and a Palestinian Information Center Page; very impressive concept!

### Useful Search Features

#### • Find Images to Steal (Alta Vista)

I bet you will all use this at one time or another, so I insist you credit this article and [webreference.com](#) for this goodie: With Alta Vista, you can limit your search to image titles by using the format:

image:title\_string

This was the only way I could find a useful picture of a nose for a physician's page - I had searched through jillions of clip art pages, and even contacted graphic artists, and they couldn't come up with anything as good as I found for free! USE THIS.

Try it now (replace ansel with your choice of image search string):

Alta Vista Search:

#### • Search for Strings in Titles (Alta Vista, OpenText) for faster results.

If applicable, this kind of search eliminates chaff by sticking to the pages that center on your subject, not ones that just mention a lexically related word. Use the syntax:

title:search\_string

in Alta Vista, or just use the simple pull-down menus in OpenText's "advanced search mode."

- **Find the Links to Your Own Site (Alta Vista)**

Alta Vista claims that you can get all the links to your own site by searching with the keyword construction: +link:http://mysite.com/-host:mysite in the Simple query

...I found that the most important link to one of my sites was missing from this search, so I was not impressed; however, my editor swears by this. Try it now (replace webreference below with your site name):

Alta Vista Search:

  
 

- **Find the Number of Links to Your Own Site (Alta Vista)**

For a more accurate estimate of the actual number of links to your site (or backlinks), use Alta Vista's advanced search, and display the results as a "count only." The above method will give you links, but approximates their number, this method more accurately estimates the number of backlinks. Try it now (replace webreference below with your site name) ABK-12-29-96:

Search  and Display the Results

**Selection Criteria:** Please use Advanced Syntax (AND, OR, NOT, NEAR).

## Which is the Best Search Engine?

(It's not just how big your data base is, it's how you use it.)

To decide which search engine I would choose as the best, I decided that nothing but useful results would count. [Previous articles](#) have emphasized quantified measures for speed and database sizes, but I found these had little relevance for the best performance in actual searches. By now, all engines have great hardware and fast net links, and none show any significant delay time to work on your search or return the results. Instead, I just came up with a few topics that represented, I felt, tough but typical problems encountered by people who work on the net: First, I tried a search with "background noise", a topic where a lot of closely related but unwanted information exists. Next, I tried a search for something very obscure. Finally, I tried a search for keywords which overlapped with a very, very popular search keyword. I defined a search as successful only if the desired or relevant sites were returned on the first page of results.

### Example - Search Terms Which Yield Too Many Matches

For the first type of search, I wanted to find a copy of Wusage to download, free software that lets you keep track of how often your server or a specific page is accessed, a common tool for HTML developers. This site is hard to find because output files are produced by the program on every machine running it that have the string "wusage" in their title and text. When I simply typed "wusage" into search page forms, Infoseek and Lycos were the only engines to find the *free version* of the software I wanted. (Note I gave no credit for finding the version for sale. A careful search of the sale version's page, did *not* produce any links to the free version's download site.) Infoseek's summaries were very poor, however, and all matches had to be checked.

### Always Search As Specifically As Possible

Most engines failed to find their quarry because the search was too broad. After all, how is the engine supposed to know I want the free version? After spending a long time to find out the *exact* name of what I wanted, "wusage 3.2", Infoseek, Excite, Magellan, and Lycos all found the site I was interested in. Alta Vista, Hotbot, and OpenText yielded nothing of interest on their first page. Magellan came out the clear winner on this search, as the site summary was by far the best. (Asking Alta Vista to display a detailed version of the results didn't change things at all!) Infoseek and Excite performed well, but Lycos listed a much older version of wusage (2.4) first.

### Think About Search Terms

It eventually occurred to me to search for "wusage AND free" to find the free copy of wusage. In some sense, Lycos was the winner this time because the free version was the first match listed; however, its summary was not very useful. While it did a better job than Infoseek, it didn't tell me whether each site was relevant or not. Magellan's response was very good, as it included a link leading to the software on the first page of matches, again with an excellent summary. Yahoo and Alta Vista also found it, but all these engines rated the fee version higher than the free version. OpenText did very well here, but only in advanced search mode where it was possible to specify

that wusage must be in the title, and "free" could be anywhere in the text. Wusage3.2 was listed as the second of only two entries - no digging here! Excite failed to find the site at all, and HotBot found only 10 matches for statistics of a server in Omaha.

Curiously, a search for "download wusage" did not improve the results over the single-word searches for any of the search engines! (It may be time for rudimentary *standardized* categories to be used on the Web: e.g. this is a download archive, this is an information only site, this is an authoritative site, etc.) The lesson here may just be "if at first you don't succeed..."

### Catalogs

Catalogs were not helpful. Yahoo!, under computers/software had nothing whatever to try for wusage: no http, no HTML, no wusage, not even servers. In Excite!, under computing/www/web ware, three more clicks got me to wusage, but -surprise!- I could not get to the free version. See why you don't want anyone *else* filtering your information?

The lessons from this search, which I have found repeated in other searches, are given in the "Examples: Summary . . ." box below.

### Example - Finding The Really Obscure

For this example, let's try to find out how to care for a "tegu", a South American lizard that is only moderately popular even among lizard enthusiasts. (If that's not an adequate example of obscure information, I don't know what is.) I know that a page exists called "TEGU INTRO" at <http://www.concentric.net/~tegu/tegu.html>, but we will simulate a blind search here. This search was full of surprises.

First I began by just searching for the string "tegu." Infoseek's first match was a tegu page I did NOT know about! Still, the one I wanted was not listed on the first page. Excite yielded nothing about tegus, only information on a vaguely related reptile, the "dwarf tegu." A search on the string "tegu care" yielded nothing relevant. (A search on their handy Usenet database did find the old tegu article I was looking for, three weeks old, which was no longer on my local news server. Other engines found this as well.) Lycos came up with the URL Infoseek found, plus two more, however, the additional listings were only pictures, not information. Searching for the string "tegu care" got nothing. Alta Vista found nothing useful either way, just ads for lizard food. OpenText found nothing, even when I searched for "tegu lizard." Hotbot found a picture of a tegu with "tegu care," but it did not return any relevant information with any search.

None of the searches I tried came up the URL I knew about. The lesson here is that you can really find new things on the Web with search engines, but if you need to find a specific page, it will always be a crap shoot. Advanced searches yielded nothing more with any engine ("tegu in title AND (care or lizard)", etc.) Some way to require that the searches were only among English language documents would have been much more helpful. Some northern-European sounding language apparently has the word tegu in it, not referring to a lizard, and many foreign language pages fouled my results on some engines. Another feature that would really have made a difference would be a filter for sales pages -- most of the mentions of tegu on the net are ads for "Monitor and Tegu Food", containing no care information. As expected, Yahoo! and Excite! Catalogs were useless here as well.

### Example - Selectivity: Apple Trees NOT Apple Computers

There are gobs of stuff on the net about Apple Computers, but what about growing apple trees? Surprisingly, this search was very easy! apple\* alone always yielded lots of stuff about the computers, and one often had to add as many as five excluded terms (apple\* -vendor\* -hard\* -soft\* -comp\* -mac\*) before receiving any matches for apples you can eat. Surprisingly, however, just apple\* tree\* usually yielded detailed information on growing apple trees on the first page of results. The poorer results required one to increase the search

### Examples Summary: How To Improve Your Searches

#### The most valuable search tool is specific information

on a search. (In the search for wusage, I had no problems when I knew that version 3.2 was what I needed.)

#### Think about your search terms - the next most important search tool

Obviously, since I wanted the free version of wusage in the example, I should have searched for "free AND wusage"; I got nothing with just "wusage" with most engines.

#### Good site summaries save you time by saving you surfing

Use Magellan or OpenText if possible. To research the example above, I had to pour through dozens of pages. Only Magellan's summaries really gave me any confidence that I did not have to check every site.

#### Specify a "title only" search if applicable

Title only searches are available only with OpenText and Alta Vista. In the examples, it yields more practical results than coming up with lots of search words, (as help pages suggest) or than forming logically complex search queries (as one might think). Adding more search words made the results above worse, not better. A Boolean search also did no better, e.g.. "wusage AND (free or download)" yielded nothing from Alta Vista.

#### Searches Can Yield New Information, but they are never complete

None of my searches ever found the good page on tegu care that I know exists.

command to apple\* tree\* grow\*.

**And The Winner Is . . .**

I don't really want to pick a winner. . . All right, if you insist: The "Search Test Results . . ." table, below, lists the engines in order of their ranking. **Lycos** is therefore the official heavy weight search engine champion of the universe, based on the tests above. However, I think this is missing the point. As shown in the table, "[Which Search Page . . . ?](#)", above, you should choose different engines for different tasks. None of the engines tested were able to limit their searches to images except for Alta Vista. This engine must therefore surely be the best *one* for graphics designers if they are allowed to use only one, but for most other purposes, the user will have to wade through the mountains of chaff and drek to find what they want. It is more beneficial to use different engines for different tasks; at most only a few are required.

Search Engine Test Results				
Engine	"One Item Among Many Related Pages" Test	"Obscure Item" Test	"Selectivity: Apple Trees Not Computers" Test	Comments
Lycos	Found item with broad search word and exact name. Found item first on results list with two search terms.	Found unknown item, but not known item.	Just apple\$ tree\$ yielded good results.	Returned the most relevant matches in the tests, but requires more time to check bad matches than Magellan.
Infoseek	Found item with broad search word and exact name. Found item with two search terms.	Found unknown item, but not known item.	Just apple\$ tree\$ yielded good results.	Poor Summaries.
OpenText	Found wusage in title search	Found Nothing.	Good results with 2 or 3 terms, most useful with 3 terms due to superior summaries.	Ability to specify title searches very useful and user-friendly. Summaries very good.
Alta Vista	Failed with approximate and exact words. Found item low on first page with two search terms.	Found nothing	Good results with apple* tree* grow*.	Keyword searches for images, titles, etc. are very useful in other searches.
Magellan	Found with exact name. Found item low on first page with two search terms.	Found nothing	Required three search terms: apple* tree* grow*	Superior summaries always save you surf time.
Excite	Found with exact name, failed with two word search.	Found nothing.	Required third search term: apple* tree* grow*, even then irrelevant results were first.	. . .
HotBot	Failed all searches	Failed all searches	Found only images, and did worse when grow* was	Poorest Performer (excluding catalogs).

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

## LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

## FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

## E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.