

## The BEST Search Engines

UC Berkeley - Teaching Library Internet Workshops

[About This Tutorial](#) | [Table of Contents](#) | [Handouts](#) | [Glossary](#)

### Recommended Search Engines: Tables of Features

[Google](#) appears still to have the largest database of Web pages, including many other types of Web documents (e.g., [PDFs](#), Word, Excel, PowerPoint documents). Despite the presence of many advertisements and considerable clutter from [blog](#) sites and [newsgroups](#), Google's [popularity](#) ranking often makes pages worth looking at rise near the top of search results. Our new "[Googling to the Max](#)" course ([when offered?](#)) reflects our recognition that Google currently is the winning web search engine and so people need to learn to use it really well.

Google alone is often not sufficient, however. Less than half the searchable Web is fully searchable in Google. [Overlap studies](#) show that about half of the pages in any search engine database exist only in that database. Getting a second opinion is therefore often worth your time. For a second opinion, we recommend [Teoma](#) or [Yahoo! Search](#). We [no longer recommend using any meta-search engines](#) for web searching, although some of the features in [Dogpile](#) and Vivisimo's [Clusty](#) may prove useful for some searches.

**Features in common among the search engines we recommend.** Search engines have become a little bit standardized, allowing us to use some of the same search techniques in all of them:

Things You CAN Do in Google, Yahoo!, and Ask.com	Things NOT Supported in Google, Yahoo!, or Ask.com
<ul style="list-style-type: none"> <li>• <a href="#">Phrase Searching</a> by enclosing terms in double quotes</li> <li>• <b>OR</b> searching with capitalized OR</li> <li>• <a href="#">- excludes, + requires</a> exact form of word</li> <li>• <a href="#">Limit</a> results by language in Advanced Search</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Truncation</a> - use OR searches for variants (airline OR airlines)</li> <li>• <a href="#">Case sensitivity</a> capitalization does not matter</li> </ul>

### Some Ways the Recommended Search Engines Differ:

Search Engine	<a href="#">Google</a> <a href="http://www.google.com">www.google.com</a>	<a href="#">Yahoo! Search</a> <a href="http://search.yahoo.com">search.yahoo.com</a>	<a href="#">Ask.com</a> <a href="http://www.ask.com">www.ask.com</a>
Links to help	<a href="#">Google help pages</a>	<a href="#">Yahoo! help pages</a>	<a href="#">Ask help pages</a>
<b>Size, type</b> Size varies frequently and widely. See <a href="#">tests</a> and more <a href="#">charts</a> .	HUGE. Size not disclosed in any way that allows comparison. Probably the biggest. Biggest in <a href="#">tests</a> .	HUGE. Claims over 20 billion total "web objects."	LARGE. Claims to have 2 billion fully indexed, searchable pages. Strives to become #1 in size.
<b>Noteworthy features and limitations</b>	Popularity ranking using <a href="#">PageRank™</a> . Indexes the first 101KB of a Web page, and 120KB of PDF's. ~ before a word finds synonyms sometimes (~help > FAQ, tutorial, etc.)	<a href="#">Shortcuts</a> give quick access to dictionary, synonyms, patents, traffic, stocks, encyclopedia, and <a href="#">more</a> .	Subject-Specific Popularity™ ranking. Suggests broader and narrower terms.
<b>Phrase searching</b> ( <a href="#">term definition</a> )	Yes. Use " ". Searches common " <a href="#">stop words</a> " if in phrases in quotes.	Yes. Use " "	Yes. Use " ". Searches common " <a href="#">stop words</a> " if in phrases in quotes.
<b>Boolean logic</b> ( <a href="#">term definition</a> )	Partial. AND assumed between words. Capitalize OR. - excludes. No ( ) or <a href="#">nesting</a> . In <a href="#">Advanced Search</a> , partial Boolean available in boxes.	Accepts AND, OR, NOT or AND NOT, and ( ). <i>Must be capitalized</i> . You <b>must</b> enclose terms joined by OR in parentheses (classic <a href="#">Boolean</a> ).	Partial. AND assumed between words. Capitalize OR. - excludes. No ( ) or <a href="#">nesting</a> .
<b>+Requires/ -Excludes</b> ( <a href="#">term definition</a> )	- excludes + will allow you to retrieve " <a href="#">stop words</a> " (e.g., +in)	- excludes + will allow you to search common words: "+in truth"	- excludes + will allow you to retrieve " <a href="#">stop words</a> " (e.g., +in)
<b>Sub-Searching</b> ( <a href="#">term definition</a> )	Sort of . At bottom of results page, click "Search within results" and enter more terms. Adds terms.	Add terms.	Sort of . Add terms.

<b>Results Ranking</b> <a href="#">(term definition)</a>	Based on page popularity measured in links to it from other pages: high rank if a lot of other pages link to it. <a href="#">Fuzzy AND</a> also invoked. Matching and ranking based on "cached" version of pages that may not be the most recent version.	Automatic <a href="#">Fuzzy AND</a> .	Based on Subject-Specific Popularity™, links to a page by related pages. More <a href="#">info</a> .
<b>Field limiting</b> <a href="#">(term definition)</a>	link: site: intitle: inurl: <a href="#">Advanced Search</a> boxes for most of these. Offers <a href="#">Uncle Sam</a> for US federal pages and other <a href="#">special searches</a> .	link: site: intitle: inurl: url: hostname: ( <a href="#">Explanation of these distinctions</a> .)	intitle: inurl: site:
<b>Truncation Stemming</b> <a href="#">(term definition)</a>	No truncation. Stems some words. Search variant endings and synonyms separately, separating with OR (capitalized): <i>airline OR airlines</i>	Neither. Search with OR as in Google.	Neither. Search with OR as in Google.
<b>Case sensitivity</b> <a href="#">(term definition)</a>	No.	No.	No.
<b>Language</b>	Yes. Major Romanized and non-Romanized languages in <a href="#">Advanced Search</a> .	Yes. Major Romanized and non-Romanized languages.	Yes. Major Romanized languages. Use <a href="#">Advanced Search</a> to limit.
<b>Limit by age of documents</b>	In <a href="#">Advanced Search</a> .	In <a href="#">Advanced Search</a> .	In <a href="#">Advanced Search</a> .
<b>Translation</b>	Yes, in <a href="#">Translate this page</a> link following some pages. To and sometimes from English and major European languages and Chinese, Japanese, Korean.	Yes.	No.

You may also wish to consult "[What Makes a Search Engine Good?](#)" - a table (PDF file) summarizing the factors I use to evaluate search engines overall.

[Return to the top of this page](#)

## How do Search Engines Work?

Search Engines for the general web (like all those listed above) do not really search the [World Wide Web](#) directly. Each one searches a database of the full text of web pages selected from the billions of web pages out there residing on [servers](#). When you search the web using a search engine, you are always searching a somewhat stale copy of the real web page. When you click on [links](#) provided in a search engine's search results, you retrieve from the server the current version of the page.

Search engine databases are selected and built by computer robot programs called [spiders](#). Although it is said they "crawl" the web in their hunt for pages to include, in truth they stay in one place. They find the pages for potential inclusion by following the links in the pages they already have in their database (i.e., already "know about"). They cannot think or type a [URL](#) or use judgment to "decide" to go look something up and see what's on the web about it. (Computers are getting more sophisticated all the time, but they are still brainless.)

If a web page is never linked to in any other page, search engine spiders cannot find it. The only way a brand new page - one that no other page has ever linked to - can get into a search engine is for its URL to be sent by some human to the search engine companies as a request that the new page be included. All search engine companies offer ways to do this.

After spiders find pages, they pass them on to another computer program for "indexing." This program identifies the text, links, and other content in the page and stores it in the search engine database's files so that the database can be searched by keyword and whatever more advanced approaches are offered, and the page will be found if your search matches its content.

Some types of pages and links are excluded from most search engines by policy. Others are excluded because search engine spiders cannot access them. Pages that are excluded are referred to as the "[Invisible Web](#)" -- what you don't see in search engine results. The Invisible Web is estimated to be two to three or more times bigger than the visible web. For more information about the Invisible Web and how to find and use the web "hidden" in it, please see our [web page on this topic](#).

### Quick Links

More search tools: [Search Engines](#) | [Meta-Search Engines](#) | [Subject Directories](#) | [Invisible Web](#)

