http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html [1]

Recommended Search Engines
UC Berkeley - Teaching Library Internet Workshops

Google is currently the most used search engine. It has one of the largest databases of Web pages, including many other types of web documents (blog [2] posts, wiki [3] pages, group [4] discussion threads and document formats (e.g., PDF [5]s, Word or Excel documents, PowerPoints). Despite the presence of all these formats, Google's popularity [6] ranking often places worthwhile pages near the top of search results.

Google alone is not always sufficient, however. Not everything on the Web is fully searchable in Google. Overlap studies [7] show that more than 80% of the pages in a major search engine's database exist only in that database. For this reason, getting a "second opinion" can be worth your time. For this purpose, we recommend Yahoo! Search [8] or Exalead [9]. We do not recommend using meta-search engines [10] as your primary search tool.

Table of features
Some common techniques will work in any search engine. However, in this very competitive industry, search engines also strive to offer unique features. When in doubt, look for "help", "FAQ", or "about" links.

Search Engine	Google [16] www.google.com [16]	Yahoo! Search [8] search.yahoo.com [8]	Exalead [9] www.exalead.com/search/ [9]
Links to help	Google help [17]	Yahoo! help [18]	Exalead help [19] and FAQ [20]
Size, type	IMMENSE. Size not disclosed in any way that allows comparison. Probably the biggest.	HUGE. Claims over 20 billion total "web objects."	LARGE. Claims to have over 8 billion searchable pages.
Noteworthy features	PageRank™ [21] system includes hundreds of factors, emphasizing pages most heavily linked from other pages. Many additional databases including Book Search, Scholar (journal articles), Blog Search, Patents, Images, etc.	Shortcuts [22] give quick access to dictionary, synonyms, patents, traffic, stocks, encyclopedia, and more.	Truncation lets you search by the first few letters of a word. Proximity search lets you find terms NEAR each other or NEXT to each other. Thumbnail page previews. Extensive options for refining and limiting your search.
Phrase searching what's this? [23]	Enclose phrase in "double quotes".	Enclose phrase in "double quotes".	Enclose phrase in "double quotes".
Boolean logic what's this? [24]	Partial. AND assumed between words. Capitalize OR. ( ) accepted but not required. In Advanced Search [25],	Accepts AND, OR, NOT or AND NOT. *Must be capitalized.* ( ) accepted but not required.	Partial. AND assumed between words. Capitalize OR. ( ) accepted. See Web Search Syntax [19] for more options.

	partial Boolean available in boxes.		
+Requires/ -Excludes what's this? [27]	- excludes + retrieves "stop words [28]" (e.g., +in)	- excludes + will allow you to search common words: "+in truth"	- excludes + retrieves "stop words [28]" (e.g., +in)
Sub-Searching what's this? [29]	The search box at the top of the results page shows your current search. Modify this (e.g., add more terms at the end.)	The search box at the top of the results page shows your current search. Modify this (e.g., add more terms at the end.)	The search box at the top of the results page shows your current search. Modify this (e.g., add more terms at the end.)
Results Ranking what's this? [30]	Based on page popularity measured in links to it from other pages: high rank if a lot of other pages link to it. Fuzzy AND [31] also invoked. Matching and ranking based on "cached" version of pages that may not be the most recent version.	Automatic Fuzzy AND [31].	Popularity ranking emphasizes pages most heavily linked from other pages.
Field limiting what's this? [32]	link: site: intitle: inurl: Offers U.S.Gov't Search [33] and other special searches [34]. Patent [35] search.	link: site: intitle: inurl: url: hostname: (Explanation of these distinctions [36].)	intitle: inurl: site: after:[time period] before:[time period] (For details, click on "Advanced search")
Truncation, Stemming (what's this?) [37]	No truncation. Stems some words. Search variant endings and synonyms separately, separating with OR (capitalized): *airline OR airlines*	Neither. Search with OR as in Google.	Use * example: messag*
Language	Yes. Major Romanized and non-Romanized languages in Advanced Search [38].	Yes. Major Romanized and non-Romanized languages.	Extensive language and geographic options. Use "Advanced Search".
Translation	Yes, in "Translate this page" link following some pages. To and sometimes from English and major European languages and Chinese, Japanese, Korean. Ues its own translation software with user	Available as a separate service [39].	Yes, in "Translate this page" link following some pages.

feedback.

You may also wish to consult "What Makes a Search Engine Good? [40]" - a table (PDF file) summarizing useful factors for evaluating search engines.

How do Search Engines Work?

Search engines do not really search the World Wide Web directly. Each one searches a database of web pages that it has harvested and cached. When you use 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 the current version of the page.

Search engine databases are selected and built by computer robot programs called spiders [41]. These "crawl" the web, finding pages for potential inclusion by following the links in the pages they already have in their database. They cannot use imagination or enter terms in search boxes that they find on the web.

If a web page is never linked from any other page, search engine spiders cannot find it. The only way a brand new page can get into a search engine is for other pages to link to it, or for a human to submit its URL for inclusion. All major search engines 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.

Many web pages are excluded from most search engines by policy. The contents of most of the searchable databases mounted on the web, such as library catalogs [42] and article databases [43], are excluded because search engine spiders cannot access them. All this material is referred to as the "Invisible Web [44]" -- what you don't see in search engine results.

 Quick Links

 Search Engines |Subject Directories [45] | Meta-Search Engines [10] | Invisible Web [44]

 Source URL (modified on 03/12/2018 - 08:50): http://www.ndj.edu.lb/projet-social-finding-information-internet-tutorial

Links
[1] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html
[2] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Blog
[3] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Wiki
[4] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Groups
[5] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#PDF
[6] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Popularity
[7] http://searchengineland.com/dogpile-search-results-on-major-engines-diverging-11364

[8] http://search.yahoo.com/

[9] http://www.exalead.com/search/

[10] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.html

[11] http://www.ndj.edu.lb/Glossary.html%23Phrase

[12] http://www.ndj.edu.lb/Glossary.html%23%2B/-

[13] http://www.ndj.edu.lb/Glossary.html%23Limit

[14] http://www.ndj.edu.lb/Glossary.html%23Truncation

[15] http://www.ndj.edu.lb/Glossary.html%23Case

[16] http://www.google.com/

[17] http://www.google.com/support/?ctx=web&hl=en

[18] http://help.yahoo.com/l/us/yahoo/search/basics/basics-04.html

[19] http://www.exalead.com/search/web/search-syntax/

[20] http://www.exalead.com/software/services/knowledgebase/websearchengine/faq/

[21] http://www.google.com/corporate/tech.html

[22] http://help.yahoo.com/l/us/yahoo/search/tips/basics-05.html

[23] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Phrase

[24] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#BooleanLogic

[25] http://www.google.com/advanced_search

[26] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Boolean.pdf

[27] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#+/-

[28] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#StopWords

[29] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#SubSearching

[30] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#ResultsRanking

[31] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#FuzzyAnd

[32] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Limiting

[33] http://www.google.com/unclesam

[34] http://www.google.com/options/specialsearches.html

[35] http://www.google.com/patents

[36] http://help.yahoo.com/help/us/ysearch/basics/basics-04.html

[37] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Truncation

[38] http://www.google.com/advanced_search.html

[39] http://babelfish.yahoo.com/

[40] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SrchEngCriteria.pdf

[41] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html#Spider

[42] http://www.lib.berkeley.edu/Catalogs/guide.html

[43] http://www.lib.berkeley.edu/find/types/articles.html

[44] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html

[45] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SubjDirectories.html

[46] http://www.lib.berkeley.edu/Comments/