

Search Engines

There are four main components to a search engine which will affect the results of your search.

1) The Webcrawler

This is the part of the search engine which combs through the pages on the internet and gathers the information for the search engine. Variable features which can affect your search results include:

- Included pages
Most search engines will find information by beginning at one page and then following all of the links on that page. It will then follow all of the links on these new pages, and so on. Therefore, if a page is not linked to from another page, it may never be found by a search engine. Authors can include unlinked pages in a search engine by submitting them to each specific search engine.
- Excluded pages
Some web administrators may choose to exclude their pages from search engines because they are internal pages or Intranets. Many web pages are also excluded because their content is dynamically generated from a database and a search engine cannot find it.
- Documents types
Different search engines will search different document types. All will search HTML documents, but some will also search PDF, PowerPoint, Word, Excel, and more.
- Frequency of crawling
An important part of a web crawler is how frequently it retrieves information from pages. Some sites it will visit more often than others.

2) The Database

The search engine's database is what you are actually searching. All of the information that a web crawler retrieves is stored in a database. Every time you use a search engine, it is this database you are searching, not the live internet. Variable features which can affect your search results include:

- Size of the database
Some search engines will have extremely large databases (~5 billion pages) for you to search, while others will have comparatively small ones (~150 million pages). A small database is not necessarily worse, however, since it may offer more focused and higher quality results than a larger database.
- Freshness of the database
The freshness of the database is a direct result of how frequently the web crawler retrieves new information. If the information in the database is fairly old, then your search results will suffer.

3) The Search algorithm

Each search engine interprets the terms you enter into the search box in different ways. Variable features which can affect your search results include:

- Operators
Most search engines allow you to use operators such as AND, OR, and NOT in order

to create complex search statements. The terms may need to be entered in upper case. Most also use the plus (+) sign to signal that a term must be included in the search results and a minus (-) sign to signal that a term must be excluded from the results.

– Phrase Searching

Search engines will generally search for words as phrases when quotation marks are placed around the phrase. Some search engines may use a drop down menu which offers the option of searching the terms as a phrase.

– Truncation

Some search engines will automatically truncate the terms you enter. This means that the search engine will not only search for the term exactly as you spelled it, but will also search on that term with alternate endings and as a plural. Some search engines will only search for variable endings on certain common words.

4) The Ranking algorithm

How a search engine ranks the results of your search is possibly the most important component of a search engine. Most searches will retrieve thousands of results. Since you probably will only look through the first 1-2 pages of results, you need the most relevant results to appear first. Variable features which can affect your search results include:

– Location and Frequency

All search engines look at the location and frequency of words in a page. If a term appears near the top of a web page, such as in the title or in the first few paragraphs of text, it is assumed that the page is more relevant than if the term is used at the bottom of the page. Pages where the words appear more frequently in relation to the other words on the page also qualifies the page as being more relevant than other web pages.

– Link Analysis

This feature analyzes how pages link to each other and then uses this information to determine the “importance” of each page. If a page is linked to from a large number of other pages, then it is ranked more highly.

– Clickthrough measurement

Some search engines also use clickthrough analysis. This means that a search engine might watch what results someone selects from a particular search, then eventually drop high-ranking pages that aren't attracting clicks, while promoting lower-ranking pages that do pull in visitors.

Searching Google

Automatic AND Between Words

When you enter multiple terms into Google, the search engine will return pages which include *all* of the terms you entered.

Example: anthrax smallpox botulism plague preparedness

Word Variations

At times, Google will search for variations on the terms you have entered. While helpful, this feature works inconsistently and you need to look at your results in order to determine exactly which terms Google has searched. You can turn off this feature by using a plus (+) sign before each term.

Example: Teen Pregnancy
+Teen Pregnancy
Example: Health Food
Healthy Food
Healthier Food

Common Words (where, how, you, etc.)

There are a number of common words that Google will ignore. In order to include common words in your search, you can use a plus (+) sign before each term.

Example: How do you write grants
Example: +write grants

Phrase Searching ("Quotes")

In order to search for phrases in Google, you need to place quotation marks around your phrase. Phrase searching will also turn off the exclusion of common words.

Example: "How do you" catch tuberculosis

Exclude Words (-)

If your search retrieves too many off-topic results, you can exclude words from your search results using a minus (-) sign before the excluded word.

Example: emergency preparedness -bioterrorism

Synonyms (OR)

When you need to search for synonyms or related terms, you need to use an upper case OR between each term or phrase. Be careful: Google will only search for the first ten terms that you enter.

Example: "Bad River" OR "Courte Oreilles" OR "Flambeau Bank" OR Sokaogon OR "St. Croix" Chippewa