

The following is an excerpt from *Google Hacks*, by Tara Calishain and Rael Dornfest and published by O'Reilly, 2003. (p 1 – 4).

Searching Google

Google's front page is deceptively simple: a search form and a couple of buttons. Yet that basic interface --- so alluring in its simplicity – belies the power of the Google engine underneath and the wealth of information at its disposal. And if you use Google's search syntax to its fullest, the Web is your research oyster.

But first you need to understand that the Google index isn't.

What Google Isn't

The internet is not a library. The library metaphor presupposes so many things – a central source for resource information, a paid staff dutifully indexing new material as it comes in, a well-understood and rigorously adhered-to ontology—that trying to think of the Internet as a library can be misleading.

Let's take a moment to dispel some of these myths right up front.

Google's index is a snapshot of all that there is online. No search engine—not even Google – knows everything. Then there's the content Google notices but chooses not to index at all: movies, audio, Flash animations, and innumerable specialty data formats.

Everything on the Web is credible. It's not. There are things on the Internet that are biased, distorted, or just plain wrong—whether intentional or not. Visit the Urban Legends Reference Pages (<http://www.snopes.com/>) for a taste of the kinds of urban legends and other misinformation making the rounds of the Internet.

Content filtering will protect you from offensive material. While Google's optional content filtering is good, it's certainly not perfect. You may well come across an offending item among your search results.

Google's index is a static snapshot of the Web. It simply cannot be so. The index as with the Web is always in flux. A perpetual stream of spiders deliver new-found pages, note changes, and inform of pages now gone. And the Google methodology itself changes as its designers and maintainers learn. Don't get into a rut of searching a particular way; to do so is to deprive yourself of the benefit of Google's evolution.

What Google is

The way most people use an Internet search engine is to drop in a couple of keywords and see what turns up. While in certain domains, that can yield some decent results, it's becoming less and less effective as the Internet gets larger and larger.

Google provides some special syntaxes to help guide its engine in understanding what you're looking for.

Within the page

Google supports syntaxes that allow you to restrict your search to certain components of a page, such as the title or the URL.

Kinds of pages

Google allows you to restrict your search to certain kinds of pages, such as sites from the educational (EDU) domain or pages that were indexed within a particular period of time.

Kinds of content

With Google, you can find a variety of file types; for example, Microsoft Word documents, Excel spreadsheets, and PDF files. You can even find specialty web pages like XML, XHTML, or RSS.

Special collections

Google has several different search properties, but some of them aren't as removed from the web index as you might think. You may be aware of Google's index of news stories and images, but did you know about Google's university searches? Or how about the special searches that allow you to restrict your searches by topic, to BSD, Apple, Microsoft, or the U.S. government?

These special syntaxes are not mutually exclusive. On the contrary, it's in the combination that the true magic lies. Search for certain types of pages in special collections or different page elements or different types of pages.

If you get one thing out of reading this, get this: the possibilities are (almost) endless. Experiment. Play. Keep your requirements in mind and try to bend the resources in Google to build a toolbox of search techniques that works specifically for you.

Google Basics

Generally speaking, there are two types of search engines on the Internet. The first is called the searchable subject index. This kind of search engine searches only the titles and descriptions of sites and doesn't search individual pages. Yahoo! is a searchable subject index. Then there's the full-index search engine, which uses computerized "spiders" to index millions sometimes billions, of pages. These pages can be searched by title or content, allowing for much narrower searches than searchable subject index. Google is a full-text search engine.

Whenever you search for more than one keyword at a time, a search engine has a default method of how to handle that keyword. Will the engine search for both keywords or for either keyword? The answer is called a Boolean default; search engines can default to Boolean AND (it'll search for both keywords) or Boolean OR (it'll search for either keyword). Of course, even if a search engine defaults to searching for both keywords (AND) you can usually give it a special command to instruct it to search for either keyword. But the engine has to know what to do if you don't give it instructions.

Basic Boolean

Google's Boolean default is AND; that means if you query words without modifiers, Google will search for all of them. If you search for :

snowblower Honda "Green Bay"

Google will search for all the words. If you want to specify that either word is acceptable, you put in an OR between each item:

snowblower OR snowmobile OR "Green Bay"

If you want to definitely have one term and have one of the two or more other terms, you group them with parentheses, like this:

snowblower (snowmobile OR “Green Bay”)

This query searches for the word ‘snowmobile’ or phrase “Green Bay” along with the word “snowblower”. A stand-in for OR borrowed from the computer programming realm is the | (pipe) character, as in:

snowblower (snowmobile | “Green Bay”)

If you want to specify that a query item must not appear in your results, use a – (minus sign or dash).

snowblower (snowmobile - “Green Bay”)

Simple Searching And Feeling Lucky

The I’m Feeling Lucky TM button is a thing of beauty. Rather than giving you a list of searchable results from which to choose, you’re whisked away to what Google believes is the most relevant page given to your search, a.k.a. the top first result in the list. Entering Washington post and clicking the I’m feeling lucky button will take you directly to <http://www.washingtonpost.com>. Trying president will land you at <http://www.whitehouse.gov/>

Just in Case

Some search engines are “case sensitive”; that is they search for queries based on how the queries are capitalized. A search for “GEORGE WASHINGTON” on such as search engine would not find “George Washington,” “george Washington” or any other case combination. Google is not case sensitive. If you search for Three, three, or THREE, you’re going to get the same results.