Released by ISC Community:- TickME

How To Google Accurately

By: Ketan Aagja

Basic Boolean

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 (OR).

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

XML Java "web Services"

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

XML OR Java OR "Web Services"

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

XML (Java OR "Web Services")

This query searches for the word "Java" or phrase "Web Services" along with the word "XML." A stand-in for OR borrowed from the computer programming realm is the | (pipe) character, as in:

XML (Java | "Web Services")

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

XML Java -"Web Services"

This will search for pages that contain both the words "XML" *and* "Java" but not

the phrase "Web Services."

The Special Syntaxes

In addition to the basic AND, OR, and quoted strings, Google offers some rather extensive special syntaxes for honing your searches. Google being a full-text search engine, it indexes entire web pages instead of just titles and descriptions. Additional commands, called special syntaxes, let Google users search specific parts of web pages or specific types of information. This comes in handy when you're dealing with 2 billion web pages and need every opportunity to narrow your search results. Specifying that your query words must appear only in the title or URL of a returned web page is a great way to have your results get very specific without making your keywords themselves too specific.

Released by ISC Community:- TickME

Here are some of the common keywords that you can add to your query in Google

intitle:

Restricts your search to the titles of web pages. The variation, allintitle: finds pages wherein all the words specified make up the title of the web page. It's probably best to avoid the allintitle: variation, because it doesn't mix well with some of the other syntaxes.

Eg: intitle:"george bush"

allintitle:"money supply" economics

inurl:

Restricts your search to the URLs of web pages. This syntax tends to work well for finding search and help pages, because they tend to be rather regular in composition. An allinurl: variation finds all the words listed in a URL but doesn't mix well with some other special syntaxes.

Eg: inurl:help

allinurl:search help

intext:

Searches only body text (i.e., ignores link text, URLs, and titles). There's an allintext: variation, but again, this doesn't play well with others. While its uses are limited, it's perfect for finding query words that might be too common in URLs or link titles.

Eg: intext:"yahoo.com"

intext:html

inanchor:

Searches for text in a page's link anchors. A link anchor is the descriptive text of a link. For example, the link anchor in the HTML code O'Reilly and Associates is "O'Reilly and Associates."

Eg: inanchor:"tom peters"

site:

Allows you to narrow your search by either a site or a top-level domain. AltaVista, for example, has two syntaxes for this function (host: and domain:), but Google has only the one.

Eg: site:loc.gov

site:thomas.loc.gov

site:edu

site:nc.us

link:

Returns a list of pages linking to the specified URL. Enter link:www.google.com and you'll be returned a list of pages that link to Google. Don't worry about including the http:// bit; you don't need it, and, indeed, Google appears to ignore it even if you do put it in. link: works just as well with "deep" URLs - http://www.raelity.org/apps/blosxom/ for instance-as with top-level URLs such as *raelity.org*.

Eg: link:www.google.com

cache:

Finds a copy of the page that Google indexed even if that page is no longer available at its original URL or has since changed its content completely. This is particularly useful for pages that change often. If Google returns a result that appears to have little to do with your query, you're almost sure to find what you're looking for in the latest cached version of the page at Google.

Eg: cache:www.yahoo.com

daterange:

Limits your search to a particular date or range of dates that a page was indexed. It's important to note that the search is not limited to when a page was created, but when it was indexed by Google. So a page created on February 2 and not indexed by Google until April 11 could be found with daterange: search on April 11. Note that daterange: works with Julian, not Gregorian dates (the calendar we use every day.)

Eg: "George Bush" daterange:2452389-2452389

neurosurgery daterange:2452389-2452389

filetype:

Searches the suffixes or filename extensions. These are usually, but not necessarily, different file types. I like to make this distinction, because searching for filetype:htm and filetype:html will give you different result counts, even though they're the same file type. You can even search for different page generators, such as ASP, PHP, CGI, and so forth-presuming the site isn't hiding them behind redirection and proxying. Google indexes several different Microsoft formats, including: PowerPoint (PPT), Excel (XLS), and Word (DOC).

Eg: homeschooling filetype:pdf

"leading economic indicators" filetype:ppt

related:

Finds pages that are related to the specified page. Not all pages are related to other pages. This is a good way to find categories of pages; a search for related:google.com would return a variety of search engines, including HotBot, Yahoo!, and Northern Light.

Eg: related:www.yahoo.com

related:www.cnn.com

info:

Provides a page of links to more information about a specified URL. Information includes a link to the URL's cache, a list of pages that link to that URL, pages that are related to that URL, and pages that contain that URL. Note that this information is dependent on whether Google has indexed that URL or not. If Google hasn't indexed that URL, information will obviously be more limited.

Eg: info:www.oreilly.com

info:www.nytimes.com/technology

Finding Technology Terminology

Specialized vocabularies remain, for the most part; fairly static words don't suddenly change their meaning all that often. Not so with technical and computer-related jargon. It seems like every 12 seconds someone comes up with a new buzzword or term relating to computers or the Internet, and then 12 minutes later it becomes obsolete or means something completely different - often more than one thing at a time. Maybe it's not that bad. It just feels that way. Google can help you in two ways; by helping you look up words and by helping you figure out what words you don't know that you need to know.

First things first: for heaven's sake, please don't just plug the abbreviation into the query box! For example, searching for XSLT will net you 900,000 results. While combing through the sites Google turns up may eventually lead you to a definition, there's simply more to life than that. Instead, add "stands +for" to the query if it's an abbreviation or acronym. "XSLT stands +for" returns around 29 results, and the very first is a tutorial glossary. If you're still getting too many results ("XML stands +for" gives you almost 1,000 results) try adding beginners or newbie to the query. "XML stands +for" beginners brings in 35 results, the first being "XML for beginners." If you're still not getting the results you want, try "What is *X*?" or "*X* +is short +for" or *X* beginners FAQ, where *X* is the acronym or term. These should be regarded as second-tier methods, because most sites don't tend to use phrases like "What is X?" on their pages, "X is short for" is uncommon language usage, and X might be so new (or so obscure) that it doesn't yet have an FAQ entry. Then again, your mileage may vary and it's worth a shot; there's a lot of terminology out there. If you have hardware or software-specific terminology as opposed to hardware or software related try the word or phrase along with anything you might know about its usage. For example, DynaLoader is software-specific terminology; it's a Perl module. That much known, simply give the two words a spin: DynaLoader Perl If the results you're finding are too advanced, assuming you already know what a DynaLoader is, start playing with the words beginners, newbie, and the like to bring you closer to information for beginners:

DynaLoader Perl Beginners

Google Groups

Usenet Groups, text-based discussion groups covering literally hundreds of thousands of topics, have been around since long before the World Wide Web. And now they're available for search and perusal as Google Groups (http://groups.google.com/). The Google Groups archive begins in 1981 and covers up to the present day. Over 200 million messages are archived. As you might imagine, that's a pretty big archive, covering literally decades of discussion. Stuck in an ancient computer game? Need help with that sewing machine you bought in 1982? You might be able to find the answers here. Google Groups also allows you to participate in Usenet discussions.

There are regional and smaller hierarchies, but the main ones are: *alt*, *biz*, *comp*, *humanities*, *misc*, *news*, *rec*, *sci*, *soc*, and *talk*. Most web groups are created through a voting process and are put under the hierarchy that's most applicable to the topic.

Browsing Groups

From the main Google Groups page, you can browse through the list of groups by picking a hierarchy from the front page. You'll see that there are subtopics, sub-subtopics, sub-sub-subtopics, and well, you get the picture. For example, in the *comp* (computers) hierarchy you'll find the subtopic *comp.sys*, or computer systems. Beneath that lie 75 groups and subtopics, including *comp.sys.mac*, a branch of the hierarchy devoted to the Macintosh computer system. There are 24 Mac subtopics, one of which is *comp.sys.mac.hardware*, which has, in turn, three groups beneath it. Once you've drilled down to the most specific group applicable to your interests, Google Groups presents the postings themselves, sorted in reverse chronological order. This strategy works fine when you want to read a slow (of very little traffic) or moderated group, but when you want to read a busy, free-for-all group, you may wish to use the Google Groups search engine. The search on the main page works very much like the regular Google search; your only clue that things are different is the Google Groups tab and each result has an associated group and posting date.

Google Groups and Special Syntaxes

You can do some precise searching from the Google Groups advanced search page. And, just as with Google web, you have some special syntaxes at your disposal. Google Groups is an archive of conversations. Thus, when you're searching, you'll be more successful if you try looking for conversational and informal language, not the carefully structured language you'll find on Internet sites. (well, some Internet sites, anyway.)

intitle:

Searches posting titles for query words.

Eg: intitle:rocketry

group:

Restricts your search to a certain group or set of groups (topic). The wildcard * (asterisk) modifies a group: syntax to include everything beneath the specified group or topic. comp.lang* or comp.lang.* (effectively the same) will find results in the group *comp.lang*, as well as *comp.lang.php*, *comp.lang.perl*, and so forth.

Eg: group:comp.lang*

group:alt*

group:comp.lang.perl.misc

author:

Specifies the author of a newsgroup post. This can be a full or partial name, even an email address.

Eg: author:fred

Released by ISC Community:- TickME

author:fred flintstone

author:flintstone@bedrock.gov

Google Images

Take a break from text-crawling and check out Google Images (http://images.google.com/), an index of over 390 million images available on the Web. Google's image search starts with a plain keyword search. Images are indexed under a variety of keywords, some broader than others; be as specific as possible. If you're searching for cats, don't use cat as a keyword unless you don't mind getting results that include "cat scan." Use words that are more uniquely cat-related, like feline or kitten. Narrow down your query as much as possible, using as few words as possible. A query like feline fang, which would get you over 3,000 results on Google, will get you no results on Google Image Search; in this case, cat fang works better. (Building queries for image searching takes a lot of patience and experimentation.) Search results include a thumbnail, name, size (both pixels and kilobytes), and the URL where the picture is to be found. Clicking the picture will present a framed page, Google's thumbnail of the image at the top, and the page where the image originally appeared at the bottom.

Searching Google Images can be a real crapshoot, because it's difficult to build multiple-word queries, and single-word queries lead to thousands of results. You do have more options to narrow your search both through the Advanced Image Search interface and through the Google Image Search special syntaxes.

Google Images Special Syntaxes

Google Images offers a few special syntaxes:

intitle:

Finds keywords in the page title. This is an excellent way to narrow down search results.

Eg: intitle: Great Wall of China

filetype:

Finds pictures of a particular type. This only works for JPEG and GIF, not BMP, PNG, or any number of other formats Google doesn't index. Note that searching for filetype:jpg and filetype:jpeg will get you different results, because the filtering is based on file extension, not some deeper understanding of the file type.

Eg: filetype:jpeg Albert Einstein

site:

As with any other Google web search, restricts your results to a specified host or domain. Don't use this to restrict results to a certain host unless you're really sure what's there. Instead, use it to restrict results to certain domains. For example, search for football.site:uk and then search for football.site:com is a good example of how dramatic a difference using site: can make.

Google News

Released by ISC Community:- TickME

Google search allows you to search for new stories from various sources in the Internet. The search form functions like Google web search, all searches are default AND. Search results group like news stories into clusters, providing title, source, date, and a brief summary (the link to the full story is included in the title). The searcher can sort their searches by relevance or date. The sort option appears on the right of the results page as you search.

Special Syntaxes

Google's News Search supports two special syntaxes.

intitle:

Finds words in an article headline.

Eg: intitle:miners

site:

Finds articles from a particular source. (Google News has over 4,000 sources).

Eg: miners site:bbc.co.uk

Making the Most of Google News

The best thing about Google News is its clustering capabilities. On an ordinary news search engine, a breaking news story can overwhelm search results. For example, in late July 2002, a story broke that hormone replacement therapy might increase the risk of cancer. Suddenly using a news search engine to find the phrase "breast cancer" was an exercise in futility, because dozens of stories around the same topic were clogging the results page. That doesn't happen when you search the Google news engine, because Google groups like stories by topic. You'd find a large cluster of stories about hormone replacement therapy, but they'd be in one place, leaving you to find other news about breast cancer. Some searches cluster easily; they're specialized or tend to spawn limited topics. But other queries like "George Bush" spawn lots of results and several different clusters. If you need to search for a famous name or a general topic (like crime, for example) narrow your search results in one of the following ways:

Add a topic modifier that will significantly narrow your search results, as in: "George Bush" environment, crime arson. Limit your search with one of the special syntaxes, for example: intitle:"George Bush". Limit your search to a particular site. Be warned that, while this works well for a major breaking news story, you might miss local stories. If you're searching for a major American story, CNN is a good choice (site:cnn.com). If the story you're researching is more international in origin, the BBC works well (site:bbc.co.uk).

Conclusion

As you have seen, Google is more than a text based search engine and by devoting a little more time in framing the query, you can narrow down your search and access your desired information from over 2 billion pages in the shortest possible time. These small techniques which you have learned now will surely help you to access the full power of Google and by doing this the whole Internet will become your playground.

Happy Googling.