

Chapter 6

Advanced Operators

6.1 Query Modifiers

Google supports a set of advanced operators (even more than are available on the Advanced Search page) that you can type in as part of a query in the regular Google search box. These operators provide you with excellent control over where you look and what results you get.

Site. The *site* operator restricts your search to pages within the site you name. The site can be any level of domain:

- The search *stolen kidneys site:com* will restrict your search for the words *stolen kidneys* to all the .com top-level domains, thus bypassing domains such as .edu and .org. (Note that you don't need the period in front of *com*.)
- The search *stolen kidneys site:about.com* will restrict your search for those words to the Web site about.com.
- The search *stolen kidneys site:urbanlegends.about.com* will restrict your search to the sub-site urbanlegends.about.com.

Note that you cannot restrict the search to specific directories with the site command (as in *site:urbanlegends.about.com/library*). And note also that there is no space between the *site* command, the colon (:) and the domain.

Tip 6.1 Get that URL

Whenever you retrieve a document that is not in HTML format and print it from a program other than your Web browser (such as Acrobat Reader or Word), there will usually not be a URL (the Web address of the page) on it. Documents in pdf, xls, doc and other formats are typically like this. However, the URL of the document is crucial for your bibliography. Therefore, always be sure to copy and paste the URL of the document into your bibliography whenever you retrieve one without a URL in it. For pdf documents, if the address is short, you can hand copy it onto the printout. For Word or Excel documents, you can copy and paste the URL into the document itself. Save yourself the trouble of locating the document a second time just to find the URL. Get it as soon as you print or save the document.

Intitle. The *intitle* operator instructs Google to look for a search word in the title of documents. It can be mixed with a search for other words anywhere in the document. Examples:

- The search *intitle:syllabus* will return all pages with the word *syllabus* in the title.

- The search *introduction literature intitle:syllabus* will return all pages with the words *introduction* and *literature* anywhere in the document and with the word *syllabus* in the title.
- The search *intitle:syllabus introduction literature* will return similar results, only with pages preferred where the word *syllabus* comes before rather than after the words *introduction* and *literature* (the *intitle* command covers only the word connected to it). Remember that your search words are taken by Google as a preferred order of appearance in the returned pages.

Allintitle. If you want to specify more than one word in a title, use the *allintitle* operator. This operator cannot be combined with a search for terms anywhere in the documents and thus must begin your query. Examples:

- The search *allintitle:drug abuse* will return all pages with the words *drug* and *abuse* in the title.
- The search *allintitle:drug abuse alcohol* will return all pages with those three words in the title, in any order.

Tip 6.2 Search the Archives

Many online magazines and news sites have archives that you can search for past articles. While many of these publications charge for their archives, many others do not. Both types usually have a search tool for the site. However, you can use Google's *site* command to search for an article or an author in one step. For example, try one of these queries:

- anthrax site:www.msnbc.com
- tax cut site:www.opinionjournal.com
- tax cut site:www.thenation.com

Inurl. The *inurl* operator tells Google to return pages that contain the attached word somewhere in the URL of the document. This operator can be combined with other search terms as well. For example: *acid rain inurl:cbs* will return pages with *cbs* somewhere in the address and with the search terms *acid* and *rain* on them.

One use of the *inurl* operator, then, is to check a favorite site for information about a relevant topic. For example, if you are a fan of urban legends and like to check up on some of the wild stories you hear, you may be familiar with the Urban Legends and Folklore guide at *about.com*. To check on a particular rumor, you can search the site quickly. For example, the query *Gerber savings bonds inurl:urbanlegends.about.com* will search the site for the words *Gerber savings bonds* and return the relevant pages.

Another use of the *inurl* operator is to find particular kinds of documents that use file names or directory names reflecting those documents. For example, the search *inurl:help* will return all pages with the

word help somewhere in the URL of the page. Some other possibilities include these:

- about [get information about the site or author or company]
- glossary [find a relevant discussion of terms related to the site]
- FAQ [read the frequently asked questions]
- tutorial [find a tutorial related to the site]

A third use for *inurl* is to find pages containing acronyms, such as those for the names of organizations (SPCA, NASA), medical conditions (OCD, ADHD), or other commonly abbreviated names or concepts (ROI, MBTI). For example, *inurl:mbti* will return all pages with those letters anywhere in the URL. This kind of search is useful when you want to locate a Web site, directory, or file name with the relevant acronym in it. It is especially useful for finding collections of information relating to a topic (several articles, for example).

A fourth use for *inurl* is to locate specific types of documents other than those in HTML. As mentioned in Chapter 4, Section 4.2, Google indexes several file formats other than HTML, including pdf, ps, doc, xls, ppt, and rtf. For example to locate a PowerPoint presentation about a marketing plan, you could use the query *inurl:ppt marketing plan*.

Allinurl. This operator can be used to search for documents where all the query words are located somewhere in a document's URL. For example, the query *allinurl:diabetes faq* will find Web pages with both *diabetes* and *faq* in the URL (site name, directory, or file name). The *allinurl* has a use similar to that of *inurl* because it allows you to find sites with directories or files that include specific search concepts.

Tip 6.3 Type File Extensions Accurately

If you are looking up a Web page you have found mentioned in a print source or otherwise need to type in a URL manually, be sure not only to get everything spelled right, but pay attention to the file extension. Originally, most pages ended in html, but now many end in htm. Adding or omitting that one letter can make the difference between finding the page and getting a "Page not found" message. And now, of course, many pages end in other extensions, such as asp, cfm, php, and so on.

6.2 Alternate Query Types

Cache. Using this operator in a query will return the page that Google has stored in its database. For example, *cache:www.cnn.com* will bring up the main page of the CNN Web site. Note that popular and frequently updated sites like this one are spidered much more often than other sites. In a test of this query, the cached page was only two days old. Other sites may have cached pages up to a month old.

When Google crawls the Web, it copies each page in order to make an index from it. This stored page may not match the page actually mounted on the Web today. Sometimes, if a Web site is not currently available, or if the site changes frequently (such as a weekly or daily update), you can view the earlier version of the page by using this command. You can also view the cached page by clicking on the “Cached” link on any results page.

The *cache* operator returns only one page at a time. If you enter a Web address (such as `www.chocolate.com`), you will get a cache of the site’s homepage. If you enter search terms in addition to the *cache* operator, Google will highlight the terms where they appear in the cached page.

Link. This operator allows you to find out what other pages link to the page you name. The number of links can be a measure of importance (or at least popularity). More valuable, though, is the fact that pages with similar content tend to link to each other, so this is a handy way of doing some cross searching. Example, suppose you like the art museum site `www.getty.edu`. The Google query `link:www.getty.edu` will return a list of pages linked to that page. After the internal page links (from other pages on the same Web site) you’ll find links from external sites, many of which also cover art topics.

Related. The *related* operator returns a list of Web pages that contain content similar to the content on the page named in the query. In the Getty Museum example above, the query `related:www.getty.edu` will return a list of other art museums’ Web sites. This command is also available as a hyperlinked “Similar pages” choice on the Google results page after each result.

The *related* operator is one of the most powerful and useful ways to locate other information about your topic quickly. Often you will find entire sites devoted to the topic of your query. This operator allows you to locate other focused pages in one step.

Tip 6.4 Watch the Screen to See Where You Are Going

When you see a hyperlink on a Web page, you can see the URL attached to that link by putting your mouse pointer on the link and then looking at the bottom of the screen (in both Internet Explorer and Netscape Navigator). You’ll see the address show up and you will know instantly whether the link takes you to another page on the same site or to another site altogether. If you mouseover all the links in a list, you can quickly get a feel for where you might choose to go.

Info. The *info* operator returns a small amount of information about the Web page specified. If the page is listed in the Google directory, the category is also listed. For example `info:www.whitehouse.gov` returns a hyperlink to the White House category of the directory. The *info* operator

is therefore useful if you have a relevant site but are not sure what category it belongs to and want to do some directory searching. (For more on the Google Directory, see Chapter 9.)

Stocks. This operator will cause Google to return stock information for the stock ticker symbol you enter. For example, *stocks:wmt* will return a page of stock market information for Wal-Mart, together with news and other links. If you don't know the ticker symbol for a company you're interested in, just type in the company name with the *stocks* command, as in *stocks:honda*. You'll get a "no such ticker symbol" notice and the option to look up the symbol of the name you typed in. You'll find the symbol in one more click.

Now you can perform advanced searches either with the Advanced Search Page or directly from the Google search box on the main page.

Tip 6.5 No HTTP Needed

As you've seen, when you type in a URL connected to a Google advanced operator, you don't need the *http://* in front, as in *glossary site:www.edmunds.com*. The fact is, you probably never need to type in that old *http://* anywhere. When you type a URL into the address box of your browser, just begin with the *www* or whatever starts the URL after the *http://* and save some keystrokes.

© 2003 by McGraw-Hill/Dushkin, Guilford, CT 06437, A Division of The McGraw-Hill Companies.

Copyright law prohibits the reproduction, storage, or transmission in any form by any means of any portion of this publication without the express written permission of McGraw-Hill/Dushkin, and of the copyright holder (if different) of the part of the publication to be reproduced. The Guidelines for Classroom Copying endorsed by Congress explicitly state that unauthorized copying may not be used to create, to replace, or to substitute for anthologies, compilations, or collective works.