Chapter 4

Advanced Search Page

Google's Advanced Search Page allows you to customize your search queries, making them much more specific than the ones we've looked at so far. You can restrict the search, expand it, and specify just what you want found and where. The Advanced Search Page is handy for the occasional user of custom queries because it offers a thorough set of pull-down menus offering many different choices. You can build your query a step at a time to frame it exactly the way you want. (For those who need advanced queries as a regular practice, Chapters 5 and 6 will show you how to include these query specifiers directly in the main search box.)

4.1 Automated Boolean

Boolean logic (named after English mathematician George Boole) involves the use of control words called logical operators that enable you to tell Google how to understand your query. The original logical operators are AND, OR, and NOT. On the Advanced Search Page, these have been expressed by descriptions of what you want to do, listed after the Find Results heading:

- With all the words. This setting is Google's default, which in effect places the Boolean logical operator AND between each word in your search string. Thus, for example, if you type in *grand unification theory* in either Google's main page search box or in the "with all the words" box on the Advanced Search page, Google understands this as a request to find pages that contain *grand* AND *unification* AND *theory* on them.
- With the exact phrase. This box creates a query the same as putting the phrase in quotation marks, requesting that pages be returned where the words appear in the order you typed them and directly adjacent to each other.
- With at least one of the words. Typing your query into this box requests the equivalent of a Boolean OR search. In other words, if you type *manic depressive bipolar* into the box, Google understands this as *manic* OR *depressive* OR *bipolar* and will look for pages with any one (or more) of the three words on them. Pages with two or all three of the words will be ranked higher than pages with only one of the words, but any page with at least one of the search words will be included in the results.

Tip 4.1 Using AND and OR

The important difference between an AND search and an OR search is that AND searches restrict your results with the more search terms you add, while OR searches expand your results with the more terms you add. It may at first seem counterintuitive, so repeat a few times: AND restricts, OR expands. To get fewer hits, use AND; to get more hits, use OR.

• Without the words. Any words typed in this box will cause Google to omit pages that contain them. For example, if you are searching for information about the fifteenth-century theologian Martin Luther, and you know that a search on those words will bring up many pages about the twentieth-century civil rights activist Martin Luther King, Jr., you can remedy the problem by typing *Martin Luther* in the "with the exact phrase" box and by typing *King* in the "without the words" box. Try the search first without and then with the omission of the word *King* and see the difference. (This box performs the Boolean NOT instruction, by the way.)

   One caution is in order with the use of the "without the words" box. It's sometimes difficult to know in advance what you don't want, and pages you would otherwise want might happen to include the word you have asked to serve as a page rejection filter. If you are sure that you can safely eliminate unwanted pages by using an exclusion word, then go ahead.

   As just indicated, you can enter search words into any or all of the logical operator boxes at the same time. You can ask for pages with a word or two, *and* with an exact phrase, *and* with at least this or that word, *and* without another word or two or three. As you can see, just these boxes allow you to construct carefully restricted (and expanded) searches.

4.2 Selecting Page Types

Google offers several ways to restrict your search results by the type of pages you want.

   Language. This choice allows you to choose pages in any language (the default), only in English, or in any one of about 35 languages. Need an article on Brie cheese written in French? What do the Germans really say about BMWs? When searching for information in a foreign language, remember that you must search in that language, since Google is looking for the words you type in.

Tip 4.2 Search in a Foreign Language.

If you want to search for a concept in a foreign language, but you don't know the language, you may be able to fake it with a little help from Google. Go the Language Tools page at `www.google.com/ language_tools` and you'll see a box where you can type in your search terms in English and have Google translate them into the language of your choice. There's no guarantee that the translated words will be the same as the words actually used in that language to describe that concept, but it may be worth a try.

The benefits of using Google's translation function include getting a sense of what is on a page in a foreign language and enjoying the continuous translation of pages as you click from one page to another on the site. There are several drawbacks, however:

• Translations are often rendered in quite broken English, with some words producing nearly meaningless sentences, as in "Page emergency not found."

• Graphics, such as banners and buttons, will not be translated, since they are not text. You might hold your mouse over the buttons and Google will translate the ALT tags (if any) that label the buttons.

• Complex pages that load from servers or in many pieces may defeat Google's ability to get text from the page and translate it.

File Format. Here's another evidence of Google's superiority in the search engine wars. Not only will Google return Web pages in HTML (hypertext markup language, the Web standard) like everybody else, but it has also indexed more than 35 million non-HTML documents connected to the Web. Google can find pages in Adobe Acrobat Portable Document Format (pdf), Adobe Postcript (ps), Microsoft Word document format (doc), Microsoft Excel spreadsheet format (xls), Microsoft PowerPoint format (ppt) and in cross-platform rich text format (rtf). By default, you will get all of these formats if relevant to your search query. But you can use this pull-down to specify a particular file type. For example, if you are looking for a spreadsheet, you can restrict your results to xls documents. Or if you want a PowerPoint presentation, look for ppt documents.

On Google's results page, you will be given the choice to view any non-HTML document as if it were an HTML document (by clicking on the hyperlink "View as HTML"), so you do not need to have the application (such as Microsoft Word) otherwise required by the document. However, to preserve formatting and see the document as it was created, you will need either the application or a free viewer (see Tip 4.3 below).

Tip 4.3 Get Acrobat Reader and PowerPoint Viewer

To view and print documents in Adobe Acrobat Portable Document Format, you will need to have Acrobat Reader installed on your computer. If you use a school computer, it may already be installed. If you don't yet have it on your own computer, you can get it free from `www.adobe.com`. To view PowerPoint presentations you will either need the PowerPoint program or the free viewer available from Microsoft at `www.microsoft.com/downloads`. (And in the event you do not have Microsoft Word or Excel, free viewers are available for them also. On the downloads page, just choose the program you want and enter "viewer" in the keyword box.)

Date. Here you can choose to get pages updated anytime (the default), within the past year, past six months, or past three months. If you are working on a current event or news story, or if you are researching a rapidly changing area in business, medicine, technology or other field, you might want to restrict your results by date. Some professors ask for sources within a certain time period, such as "this semester" or "the last year." If that's your situation, this is the tool you need.

4.3 Narrowing Your Search By Location

The next three options on the Advanced Search Page allow you to specify where the search words must appear in order to produce a hit.

Occurrences. Normally, Google includes pages where the search terms appear almost anywhere—in titles, text, hyperlinks and so on. However, you can tell Google to look only in the titles, or only in the text, or only in the page URLs themselves. For example, if you want an entire Web page or article to discuss the subject of your search terms, then you might try limiting the page hits to those with the search terms in the page titles.

Domain. You can restrict your results to pages that come from a certain top-level domain if you like. Here briefly are some of the choices:

.com   Commercial. These include businesses of all kinds, from tiny to huge multi-nationals, and some non-businesses. The most common top level.

.org   Organization. These include non-profit organizations, foundations, institutes, and associations. If the organization is reputable, the information is usually reliable. Of course, many organizations have political interests that may affect the objectivity of their information.

.gov   Government. Federal and state government sites, and nonmilitary government sites like NASA, NTSB, IRS, and so on use this. The information on these sites is the official word from these government agencies.

.mil   Military. Sites connected to the U.S. military.

.net   Network. Some Internet and other telecommunications providers use this, although it is not restricted to them. If a name is already taken in the .com or .org domain, an interested party might choose .net instead. For example, there are sites for `www.information.com`, `www.information.org`, and `www.information.net`.

.edu   Educational institution. Two- and four-year colleges and universities use this top level. Institutional pages with the .edu address reflect the official position of the college or university, but student and faculty pages may also be mounted on university servers and show this domain as well. Therefore, do not assume that any page with a .edu in the URL is the official word from the institution.

There are other top level domains, including .k12 for schools and school districts, .cc for community colleges, .lib for libraries, .state for state governments, and .ac for European colleges and universities. In addition, each country outside the United States has a national top-level domain, such as .uk for Great Britian (the .uk is for United Kingdom), .jp for Japan, and .de for Germany (the .de is for Deutschland). To see a list of all the top level domains, just take a trip to the Internet Assigned Numbers Authority at `http://www.iana.org/cctld/cctld-whois.htm`.

You can also limit your search to a single site. For example, you could look for articles on a particular topic such as *fat diet* and restrict your search to `www.usnews.com`, the site of U.S. News and World Report magazine. (Type *fat diet* in the "with all the words" box and `www.usnews.com` in the "domain" box.)

Safe Search. This selection turns on or off a filter to eliminate pornographic Web sites. In practice, the filter is not usually necessary. In the past (at other search engines), searching on a phrase such as *young girls' sports* would return many pages offering pornography. With Google's ranking technology, however, such pages are seldom seen in the top 100 or more results. Try searching on *breast bottle feeding* (to learn about the

debate between the two), and you'll see that even the word *breast* does not flood your results with porn. And, perhaps amazingly enough, even a search on *bare naked ladies* (a misspelling of the music group Barenaked Ladies) turns up only a few non-music related sites in the first 100 hits. If your search does result in a lot of porn, turn the filter on. Otherwise, it's usually safe to leave it off.

4.4 Page-Specific Search

Once you have found a page of value to you or if you already know a page you like, you can use this knowledge to find more pages like it. (Go to Advanced Search page and scroll down to Page-Specific Search.)

Similar. Typing a URL into the box here will cause Google to look for pages related to that page. The "Similar pages" choice is also available on the main search results screen whenever you perform a search. The true value of this choice is shown when your original search term or phrase can refer to many different topics. For example, if you simply type in *oxygen*, you'll get a mix of many different results. One is the Oxygen cable television channel for women, another is a "computers as common as oxygen" experiment at MIT, another a snowboard maker with the brand name Oxygen, another oxygen the chemical element, and so on. When you find the meaning of oxygen you are interested in, click on "Similar pages" and you'll get usually at least 20 or 30 pages closely related to that meaning. In the cases here, the pages similar to the cable channel cover women and women's health, the pages similar to the MIT page cover computers and smart environments, the pages similar to Oxygen Snowboards cover snowboards and other snowboard makers, and the pages similar to the one about oxygen the element concern that, other elements, and chemistry.

Note, then, that you can type in any word that seems even close to an idea you are interested in and use Google to help you narrow it down and focus on something specific.

Links. Typing a URL into this box will ask Google to locate pages that link to the page you typed in. Finding these links can be valuable for two reasons. First, you can see which pages (sites, organizations) have thought the page in question was valuable or important enough to link to it, together with the total number of links. This information can supply a rough estimate of the importance of the page to the Web community and may help you in your efforts at evaluating the page's quality. (Don't use popularity as your sole criterion, of course.) Second, pages that link to this page might cover the same or a similar subject, so you might find some valuable resources this way.

Tip 4.4 Webrings

Many groups of pages related to the same subjects have connected to each other in Webrings, so that you can move from site to site in the ring easily and see what else is available on the subject. If you want to see some of the Webring organizations (there are several), just type *webrings* into the Google search box. If you want to find out whether there is a Webring devoted to your topic or where in a Webring your topic occurs, type in your topic followed by the word, as in *jack the ripper webring*. In this case, you'll find information in the True Crime Webring.

4.5 Topic-Specific Searches

The choices in this section allow you to restrict your search to shopping, particular computer topics, U.S. Government sites, or particular universities. For example, you might choose University of Chicago from the menu and then type in *"art history" "faculty directory"* (using the two exact phrases) to get a faculty directory of the Art History Department.

Tip 4.5 100 Results Per Page

In the upper right hand corner of the Advanced Search Page is the option to have Google send you 10 results per page (the default) or 20, 30, 50, or 100. Unless you are using a very slow connection, you can save time and trouble by setting this to 100. If you want to save, print or email your results, this choice will also be more efficient.