

Chapter 1

Why Use Google?

Google (www.google.com) is the premier search tool on the Internet today, featuring not only the best Web search engine, but many additional features including a directory, image search, current news, user groups, and more. What has made Google the search tool of choice for so many users? This will become clear if we look at how search engines work and what makes a good one.

1.1 How a Search Engine Works

The World Wide Web consists of tens of millions of Web servers (computers hosting Web sites) and billions of Web pages, all interconnected. Each page is accessible by typing in an address (called a URL, for uniform resource locator). The problem the researcher faces is how to get the address of a page that may contain useful information. Printed directories may be a start, but no book could contain billions of addresses, and any printed material would be out of date before it could leave the printing plant.

Enter the search engine. Because all of the pages are interconnected by hyperlinks, where one page links to another (or usually several others), it is possible to move all around the Web just by following links. You may do this occasionally as you let your curiosity be your guide and click from page to page and site to site as you surf around. Search engines surf automatically by using programs called bots or spiders. The task of a search engine spider is to visit as many sites as possible by following links (both within the site and from site to site), index the pages, and use this index to help users of the search engine find the pages they want. Given this task, you can see that several factors determine how well a search engine performs, including the size of its index, the freshness of the content, and the ease of searching. Google excels in these areas and several more that are relevant to producing high quality results.

Tip 1.1 You Can Look It Up

Whenever you perform a search, Google gives you something in addition to the Web pages you need. It links each term in your query to a dictionary, so that by clicking on the word, you can find the definition of it. Look just under the tabs (Web, Images, Groups, Directory, News) and you'll see "Searched the Web for" followed by your search terms. Click on one and get the definition. You can

also use this feature just for the dictionary. Type in one of the following words, press the enter key or click on the Google Search button, and then instead of looking at the search results, click on the word after “Searched the Web for” to see the definition. (To return from Dictionary.com to Google, click on the back button.)

- *adventitious*
- *madrigal*
- *ergotamine*
- *teasel*

1.2 Size

Clearly, the search engine that indexes more pages and creates a larger index will have an advantage over search engines that have smaller indexes. You will be able not only to get a more thorough set of results from a larger index, but you are likely to find less common information, information that may be posted on only a few pages.

There is an ongoing competition between Google and Fast (www.alltheweb.com) for the title to the largest index, but with more than three billion Web and other document pages in its index, Google is enormous and usually considered the biggest. The more thorough your research needs to be, the more important a large-size index is, so that you can access a wide selection of articles and ideas about various aspects of your research topic. To look up what may seem to be an obscure term to you, like *caoutchouc*, and find tens of thousands of page hits, shows you the depth of content at your fingertips.

And while we are discussing size, remember that Google is more than just a search engine for Web pages. It’s a suite of search tools, each one with access to huge amounts of information. The Image Search contains an index of more than 425 million images (jpeg and gif files), including artwork, photographs, tables, graphs, drawings, and maps. The Google Groups database (an archive of Usenet postings) contains more than 700 million messages, discussing thousands of topics. (For the technical minded, Google is physically big, too, running on more than 10,000 small servers, hooked together.)

Tip 1.2 Finding Reference Materials

Google will find not only articles about a subject of interest but reference works and information tools. Try typing some of the following into Google’s search box and see what you get:

- *medical dictionary*
- *amortization calculator*
- *legal dictionary*
- *encyclopedia of philosophy*
- *Microsoft Word tutorial*

1.3 Freshness

Size by itself is not the only attribute of a high quality database. The content must be up to date, with outdated pages omitted and recent pages added. Here again, Google excels by crawling the Web anew about once a month, adding fresh pages and dropping expired or unavailable pages. The Google spiders (search robots) visit more than 4,000 news sources several times each day in order to offer up-to-date news around the clock. Directories that rely on humans to choose and update their sites often fall behind in eliminating old links that no longer work or that are outdated, and time is needed to add new entries. That's why most directories like Yahoo (www.yahoo.com) and LookSmart (www.looksmart.com) have partnered with search engine companies to supplement their directory entries.

1.4 Ease of Use

An ideal search tool will have a clean interface and be simple to use while still returning useful and pertinent results. Google fulfills this ideal by incorporating several defaults and methods of ranking pages. Even a simple search of one or two words usually produces excellent results. While we will discuss more details about building a query in Chapter 3, for now you might want to know that Google is not case sensitive, so you don't have to worry about capitalization, and a multiple-word query is automatically understood to be a search for all the words you type in. In other words, Google puts the Boolean logical operator AND in between the terms you type in. So if you type in *cyanoacrylate surgery*, Google looks for pages that contain both words. (Note: In many of the examples in this book, especially in the tips, we capitalize proper names out of courtesy. You don't have to.)

1.5 Full Text Indexing

Some search engines read only the titles of Web pages they index, or perhaps the first hundred words of the page, assuming that the words mentioned early on will be an accurate indicator of the page's content. Google goes beyond this by indexing the entire page. The benefit of this is clear: If the terms you are looking for are mentioned far down on a long page, Google has found them and will show you that page, whereas other search engines will miss that page.

Tip 1.3 Using Misspelling

Google is very helpful with misspelled words. If you enter a word in a search that does not match Google's spelling dictionary, you will not only get the results of the search for the term as you spelled it,

but Google will also ask you if you meant to spell the word correctly or at least differently.

- Type in the word *walmirt* into the Google search box and see what you get.
- Many people misspell the first word of Procter and Gamble as *Proctor*. Try a search on *proctor gamble* and notice that you will get the company's Web site (www.pg.com), pages that spell the name wrong, *and* pages that spell it correctly.

It may sometimes be the case that a page you want has spelled your search term incorrectly, especially if the word is a commonly misspelled one. For this reason, after you perform a search with all your words spelled correctly, you may want to enter a common misspelling or two to see what other results you get. Of course, pages with too many misspelled words should be viewed with special caution, but you might find something of use to you.

1.6 Superior Page Ranking

Even terms that may at first seem to be relatively restricted can be mentioned on a surprising number of pages. For example, the search phrase *ankylosing spondylitis* returns more than 38,000 results in Google. If these pages were listed in random order, the chances that you would find the really valuable discussions would be slim. How to order the results pages is a problem for every search engine. Some engines count the number of times a search term occurs on the page and put pages with more mentions higher than pages with fewer mentions. Some engines put pages higher up when the search term is closer to the beginning of the article, or if the term appears in a title.

Google has a much more effective way of ranking pages. Its PageRank™ technology uses an idea sometimes called collaborative filtering. The Google search engine looks at pages that link to other pages and counts those as votes of usefulness or quality. For example if Pages A, B, C, D, E, and F have created links to Page X while only Pages G and H have links to Page Y, Google assumes that Page X is more important than Page Y because so many more people have created links to X. More than that, Google checks to see how many people have linked to Pages A, B, C and so forth. The more links to those pages, the more important they are, and thus the more significant their votes for Page X or Y. This system might be thought of as Web democracy, where pages vote for other pages when links to those other pages are created.

But PageRank™ is only half the story. Google also uses a method of locating your search terms on the pages found, whether the terms appear in a title, several times, early on, and so forth. It weights the pages for relevance on the basis of these factors. These two evaluations (measure of importance and relevance to search terms) are combined to produce page results that are often truly astonishing. You will find that exactly the information you want often appears in the first hit or in the

first ten hits of your search results. For extended searches, the depth of relevance often extends far down into the results.

The good news here is that, even if you receive 100,000 hits for your search terms, the probability is very high that the first dozen or few dozen results will provide you with the highest quality and most relevant information.

Tip 1.4 Find a Company

Google's excellent page ranking technology allows you to find corporate Web sites by typing in just the company's name. You will usually find that the first hit gives you just what you want. Try some of the following or experiment by typing in any company name you choose:

- *Nike*
- *Honda*
- *K2*
- *McGraw-Hill*
- *Snapple*
- *Olympus*

1.7 Other Benefits

In addition to the strengths discussed above, Google offers several other benefits to the user. First, Google's use of many small servers divides the task of searching so well that most searches take less than a second, with many requiring less than half a second. Search speed is a great advantage for a society dedicated to instant results. Next, Google does not sell rankings to advertisers. Where some search tools rank their results in the order of how much they are paid by businesses, Google will not sell placements of keywords. The ranking you see on the screen is objective and commercial free. Google clearly separates the advertisements at the top or on the right and labels them "Sponsored Links." And third, Google offers an array of "power tools" that make searching both easier and more effective. We will cover these in later chapters.

Tip 1.5 Finding Books

Many full-text books are available online as well (and not just for sale by Amazon.com or Barnes & Noble). Books published before about 1900 are usually in the public domain (out of copyright) and can be posted legally on the Web. These include thousands of classics. (One site has more than 19,000 books online.) To find books of interest to you try some of the following searches or base your subject search on these:

- full text books
- full text classics online
- children's literature full text online
- "Democracy in America" full text online
- "The History of Rasselas" full text online

The more you use Google and become familiar and comfortable with its features, the more you'll enjoy this powerful and efficient tool. Feel free to experiment. You can't hurt anything. Often, the limitation in the search is the limitation in the imagination of the searcher. This book will give you many ideas for creating better searches, but you can be creative on your own. Have fun. Find what you want.

© 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.