REVISED AND EXPANDED

ALFRED AND EMILY GLOSSBRENNER



VISUAL QUICKSTART GUIDE

SEARCH ENGINES FOR THE WORLD WIDE WEB

Third Edition

Teach yourself how to search the Web the quick and easy way! This Visual QuickStart Guide uses pictures rather than lengthy explanations. You'll be up and running in no time!

ALTAVISTA
AOL SEARCH
GOOGLE
HOTBOT
MSN SEARCH
YAHOO
PLUS MORE

Petitioner has added letters to pages (a)-(e) and numbers to pages 349-353. Otherwise, it leaves the original page numbering.

MICROSOFT EXHIBIT 1057

VISUAL QUICKSTART GUIDE

SEARCH ENGINES

FOR THE WORLD WIDE WEB, 3rd Edition

Alfred and Emily Glossbrenner

Peachpit Press

(b)

Visual QuickStart Guide

Search Engines for the World Wide Web, 3rd Edition

Alfred and Emily Glossbrenner

Peachpit Press
1249 Eighth Street, Berkeley, CA 94710
510/524-2178 • 800/283-9444 • 510/524-2221 (fax)

Find us on the World Wide Web at http://www.peachpit.com

Peachpit Press is a division of Addison Wesley Longman

Copyright © 2001 by Alfred and Emily Glossbrenner

Editors: Nancy Davis, Whitney Walker, Becky Morgan
Production coordinators: Lisa Brazieal, Connie Jeung-Mills
Compositor: Owen Wolfson
Indexer: Emily Glossbrenner
Cover design: The Visual Group

Notice of Rights
All rights reserved. No part of this book may be reproduced or transmitted in any form by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. For information on getting permission for reprints and excerpts, contact Gary-Paul Prince at Peachpit Press.

Notice of Liability
The information in this book is distributed on an "As Is" basis, without warranty. While every precaution has been taken in the preparation of the book, neither the author nor Peachpit Press, shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the instructions contained in this book or by the computer software and hardware products described in it.

Trademarks
Visual QuickStart Guide is a registered trademark of Peachpit Press, a division of Addison Wesley Longman.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Peachpit Press was aware of a trademark claim, the designations appear as requested by the owner of the trademark. All other product names and services identified throughout this book are used in editorial fashion only and for the benefit of such companies. No such use, or the use of any trade name, is intended to convey endorsement or other affiliation with the book.

ISBN 0-201-73401-X

9 8 7 6 5 4 3 2

Printed and bound in the United States of America

Printed on recycled paper

(c)

025.04
G563s
2001

About the Authors

Alfred and Emily Glossbrenner are the authors of more than 60 books with combined sales of well over one million copies. They have been alerting people to the power and possibilities of online information since 1982—long before there was a World Wide Web. Their uncanny knack for explaining complex subjects in a way that anyone can understand, combined with their conversational style, has earned them praise from publications as diverse as *MacWorld*, *Forbes*, and *The New York Times*.

What information do you want to find today? The Glossbrenners will show you how to do it. Recently, they put their search skills to work to write *About the Author: The Passionate Reader's Guide to the Authors You Love, Including Things You Never Knew, Juicy Bits You'll Want to Know, and Hundreds of Ideas for What to Read Next* (Harcourt, 2000).

Packed with fascinating details about 125 popular novelists, plus lots of other book-related information—like how to find out-of-print books online, where to go for reading-group resources, what books have been featured on Oprah's Book Club and other major book lists—*About the Author* simply could not have been written without the Web.

The Glossbrenners live and work in a 200-year-old farmhouse on the Delaware River in Bucks County, Pennsylvania. You can write to them at `gloss@gloss.com`, or visit them on the Web at `www.trulyuseful.com`.

MILLS COLLEGE
LIBRARY

Mills College Library
Withdrawn

Praise from Readers and Critics

Here's a sampling of what readers and critics have said about earlier editions of *Search Engines for the World Wide Web: Visual QuickStart Guide.*

"[This book] is exactly what Web guides should be—concise, easily referenced, informative, and appropriate for all levels of expertise. If you have ever had the problems that most surfers experience when conducting a search—dredging up 100,000-plus hits on AltaVista, finding coffee pages when searching for Java, or pulling up the dreaded goose egg—you need to have this book close at hand."

—The Editors of Amazon.com

"I immediately liked *Search Engines for the World Wide Web*.... It's a small book, nicely organized, well written, and full of illustrations that make learning easier."

—Danny Sullivan, "Search Engine Report" Newsletter

"[The Glossbrenners' book] goes a long way toward bringing search newcomers, and even those who have been exploring the Net for awhile, to a higher skill level.... Those who read the book from cover to cover will feel they have 'been there' in all the nooks and crannies of each site before even going online to try them out."

—Wallys W. Conhaim, *Link-Up*

"If you've been stumped, overwhelmed, or just plain frustrated with your Internet searching, pick up this book and find the answers you've been looking for."

—*The Naked PC*

"This great little book covers all the major search engines Web users deal with every day.... There is information to get technophobes comfortable with the basics as well as advanced techniques that will help professional researchers and librarians. Highly recommended."

—Thom Gillespie, *Library Journal*

"At last, thanks to the lucid writing of the Glossbrenners and the ingenious format of Peachpit's *Visual QuickStart Guide* series, we have a work of reference that is clear, comprehensive, and concise.... With the Glossbrenners' new book in hand, you'll gain speed, confidence, and skills for searching the Web and finding precisely what you need."

—Michael Pastore, review for YouthTopia Web Site

"This [book] won me over in the first paragraph of the introduction: 'If you're like us, the fascination of *browsing* the Web wore off a long time ago. We want to sign on, get the information we need, sign off, and go about our business and our lives.'

"Hey, that's me.... This *QuickStart Guide* is easy to read without being condescending—perfect for beginners and advanced users alike."

—Krissy Harris, *Los Angeles Times*

"I just want to congratulate you on your excellent book, *Search Engines for the World Wide Web*. I thoroughly enjoyed it and will use it in my reference and training work."

—Margaret Smyth, Central Library, North York, Ontario, Canada

"The organization and design of this book are splendid.... The reader quickly learns how search engines work, and the importance of keywords. Useful tips are plentiful. The brief chapter on Basic Search Tools is helpful enough to have you slam your fist into your keyboard and exclaim, 'Why didn't someone tell me this stuff in the first place?' Ditto for the Seven Habits of Highly Effective Web Searchers....

"Perhaps my favorite part of [the book] is Appendix A, the Quick Reference pages, where all the important search engine commands are presented in alphabetical order by search engine name. This book is a winner for everyone who spends more than two minutes per month on the Web, with practical information at an affordable price."

—John Nemerovski, Book Bytes Reviewer, *My Mac Magazine*

"*Search Engines for the World Wide Web* shows readers the fastest and most efficient ways to find what they need with the least frustration. With easy-to-follow, step-by-step instructions accompanied by dozens of useful illustrations, [it] makes learning about search engines as quick and efficient as possible.

"[This book] is essential reading for the novice and has much to offer the experienced Internet information researcher as well.... It's also highly recommended for any Internet and World Wide Web information provider [who wants to] enhance and maximize accessibility and 'presence awareness' among Internet and Web users."

—The Computer Shelf, `alt.books.reviews`

"I would think that almost every library using the Internet would want at least one copy! We have had to rely on magazine articles and the Help sections at each search engine Web site in order to understand the rules for searching with each engine. [This] book will make it much easier to understand how to construct search strategies, since the information is all in one place."

—Richard Strauss, Branch Manager, Bucks County Free Library, Langhorne, PA

"[A] concise, thorough, readable guide.... Highly recommended."

—Windweaver Online Bookstore

(e)

TABLE OF CONTENTS

Introduction ix
How to Get the Most Out of This Book x
How the Book Is Organized xi

PART 1: SEARCH BASICS 1

Chapter 1: Search Engines and How They Work 3
The Perils of Internet Searching 4
Search Engines to the Rescue 5
The Big Six Search Engines 6
How Search Engines Work 9
More Information for Webmasters and Others ... 13

Chapter 2: Unique Keywords 17
The Challenge of Full-Text Searching 18
Choosing the Right Keywords 19

Chapter 3: Basic Search Tools 23
Searching in Plain English 24
Searching for Multiple Words and Phrases 26
AND Searches 27
OR Searches 28
NOT Searches 29
NEAR Searches 30
Parentheses and Nested Searches 31
Using Wildcards 32
Dealing with Stopwords 33

Chapter 4: Tips and Techniques 35
The Seven Habits of Highly Effective Web Searchers 36
Customizing Your Web Browser 43
Keyboard Shortcuts and Other Timesavers 46

PART 2: THE BIG SIX SEARCH ENGINES 51

Chapter 5: AltaVista 53
The AltaVista Home Page 55
Searching with AltaVista 60
Improving Results with Field Search 66

Using Advanced Search 69
Customizing AltaVista 74

Chapter 6: Google **79**
The Google Home Page 81
Searching with Google 84
Using Advanced Search 90
Customizing Google 95

Chapter 7: HotBot **97**
The HotBot Home Page 99
Searching with HotBot 104
Improving Results with Field Search 110
Customizing HotBot 114

Chapter 8: Lycos **117**
The Lycos Home Page 120
Searching with Lycos 125
Using Advanced Search 130
Customizing Lycos 135

Chapter 9: Northern Light **139**
The Northern Light Home Page 142
Searching with Northern Light 146
Improving Results with Field Search 152
Customizing Northern Light 154

Chapter 10: Yahoo! **157**
The Yahoo Home Page 159
Searching with Yahoo 163
Using Advanced Search 167
Using Yahoo's Web Directory 170
Customizing Yahoo 173

PART 3: OTHER GENERAL-PURPOSE SEARCH ENGINES **177**

Chapter 11: AOL Search **179**
The AOL Search Home Page 182
Searching with AOL Search 186

Chapter 12: Excite **193**
The Excite Home Page 195
Searching with Excite 198

Chapter 13: MSN Search **205**
The MSN Search Home Page 208
Searching with MSN Search 211

Chapter 14: Netscape Search 217
The Netscape Search Home Page 220
Searching with Netscape Search 223

PART 4: SPECIALIZED SEARCH ENGINES 229

Chapter 15: Deja for Newsgroup Searches 231
Understanding Newsgroup Names 234
Searching Newsgroups 236

Chapter 16: Topica for Mailing List Searches 241
Searching for Mailing Lists 243

Chapter 17: Argus Clearinghouse Subject Guides 247
Finding Subject Guides 249

Chapter 18: Infospace People-Finding Tools 253
Finding E-mail Addresses 255
Finding Phone Numbers and Addresses 258

Chapter 19: Zip2 for Business Searches 261
Finding Businesses with Zip2 264

Chapter 20: Other Search Tools from A to Z 269
Amazon.com 270
CDNOW 270
Centers for Disease Control and Prevention 271
CNET Shareware.com 272
Consumer World 273
CuisineNet Menus Online 274
Edmund's Automobile Buyer's Guides 275
Electric Library 275
Encyclopedia Britannica 276
Epicurious Food 277
Expedia Travel Information 277
FedEx Package Tracking 278
FindLaw Legal Resources 278
FirstGov 279
Information Please Almanacs 280
Internet Movie Database 280
MapQuest 281
Microsoft Product Support Services 281
Monster Career Center 282
New York Times Book Reviews 282
Parent Soup 283
Peterson's Education & Career Center 283

Project Gutenberg E-Texts 284
Tax Forms and Information 284
TV Guide Online 285
UPS Package Tracking 285
Wall Street Journal Online 286
ZDNet Computer Information 286
Zip Code Lookups 287

Appendix A: Search Engines Quick Reference 289
AltaVista 290
AOL Search 293
Argus Clearinghouse 294
Deja 295
Excite 296
Google 297
HotBot 298
InfoSpace 300
Lycos 301
MSN Search 303
Netscape Search 305
Northern Light 306
Topica 307
Yahoo 308
Zip2 309

Appendix B: Internet Domains and Country Codes 311

Appendix C: Usenet Newsgroup Hierarchies 317

Appendix D: Web Searcher's Toolkit 321
Essential Tools for Windows Users 324
Essential Tools for Macintosh Users 329

Index 333

Features Comparison Chart 348



INTRODUCTION

Late in the year 2000, the Internet passed the two billion mark—two billion Web pages, that is, all accessible via the World Wide Web. And with an estimated seven million new pages being created every day, it's entirely possible that as you read this, the Net may be approaching three, four, or even five billion total pages.

So how in blazes does anyone *find* anything? That's the million-dollar question. If you're like us, the fascination of "browsing the Web"—clicking on links to go aimlessly hither and yon—wore off years ago. We want to sign on, get the information we need, sign off, and go about our business and our lives.

Fortunately, some excellent tools are available on the Net to help you do just that. Dozens of them, in fact. They're called *search engines*, and, not surprisingly, the best of them consistently rank among the most popular sites on the Internet.

But the proliferation of search sites has created a new problem and a whole new set of questions. With so many search engines out there competing for our attention (one popular Internet directory currently lists 3,569), how do we find out what the really good ones are? Does it make any difference which one we choose? Will a search for, say, `digital camera product reviews` or `Ally McBeal episode guide` produce the same results whether we use AltaVista or Google or Yahoo?

How to Get the Most Out of This Book

This book will answer all of these questions and more. And like all the books in the *Visual QuickStart* series, it's designed to do so with a minimum of technical jargon and extraneous information. You'll find lots of step-by-step instructions and specific examples for using search engines in general, and the very best ones in particular.

In writing the book, we've made just a few basic assumptions about you:

- You understand the fundamentals of working with a computer, such as how to use a mouse and how to choose menu commands.
- You have access to the Internet—either through an Internet Service Provider (ISP) like AT&T WorldNet or EarthLink, or an online service like America Online (AOL) or Microsoft Network (MSN)—and you know how to sign onto the Net.
- You have some experience using a Web browser program like Microsoft Internet Explorer or Netscape Navigator to visit Web sites, and now you're ready to learn how to do more than just "surf the Net."

If you're not quite up to speed on one or more of these fronts, you may want to hold off on this book for the time being.

Instead, if you have a personal computer running Windows, check your favorite bookstore or local library for *The Little PC Book* by Larry Magid. If you have a Macintosh, look for *The Little Mac Book* or *The Little iMac Book*, both by Robin Williams. Published by Peachpit Press, all three of these books include chapters specifically aimed at people who are venturing onto the Internet for the very first time.

For in-depth coverage of Internet Explorer, try Steve Schwartz's *Internet Explorer 5: Visual QuickStart Guide*, also published by Peachpit Press and available in both Windows and Macintosh versions.

Once you're comfortable using your computer and Web browser software—and you've spent some time exploring Web sites on your own—the information presented in *this* book will make a lot more sense.

Companion Web Site

For book updates, links to search engine resources, and other goodies, visit the companion Web site for this book at `www.peachpit.com/vqs/search`.

How the Book Is Organized

The chapters in *Search Engines for the World Wide Web* are organized into four major parts, followed by four appendices. Here's what's covered in each one:

Part 1: Search Basics

In this part of the book we introduce you to the concept of search engines and how they work. You'll also learn about *keywords*—how to choose the right ones and the various methods of combining them for more effective searches. We round things out with some specific tips and techniques for using any search engine.

You should read the four chapters in this part of the book from start to finish, since each chapter builds on the information presented in the ones before. Also, you'll need this basic grounding in online searching to get the most out of the chapters on specific search engines in Parts 2, 3, and 4.

Part 2: The Big Six Search Engines

Here you'll find chapters on six of the best and most powerful search engines available today. You can read these chapters in any order. If you're somewhat familiar with one of the search engines covered here, you might start with that one and then branch out from there to learn about two or three others.

By the time you're finished reading about a particular search engine, you'll know its strengths and weaknesses and how to use the major features to create effective queries. Best of all, whenever you need a refresher, you can consult the Quick Reference guide included in each chapter.

Part 3: Other General-Purpose Search Engines

The Big Six search engines aren't the only game in town, of course. In this part of the book, we'll introduce you to four additional general-purpose search engines. You may very well be using one of these sites as your "portal" to the Internet—the place you visit first whenever you go online. But are you using the site's search features to maximum advantage? If you're not sure, read the relevant chapter in this part of the book, and refer to the Quick Reference guide for search rules and examples.

Part 4: Specialized Search Engines

Next we present some alternatives to the all-purpose "Swiss Army Knife" approach of the search engines covered in Parts 2 and 3. Just as cooks and carpenters need special tools from time to time, so too do Web searchers. This part of the book introduces you to some of the best of these special tools and helps you understand when to use them for faster, more efficient searches.

Appendix A: Search Engines Quick Reference

This is a collection of all the Quick Reference guides from throughout the book, organized alphabetically by search engine. Our thought is that when you're online and need a quick reminder of, say, the HotBot or Northern Light search rules, you may find it more convenient to turn to this appendix instead of going back to the individual search engine chapter.

Appendix B: Internet Domains and Country Codes

The information presented here will help you take advantage of power-searching techniques like zeroing in on a specific *type* of organization (based on the Internet domain designation in its Web address), or locating sites that originate in a particular country.

Appendix C: Usenet Newsgroup Hierarchies

This appendix explains how newsgroups are named and gives you the information you need to limit your queries to specific newsgroups, a feature offered by some search engines.

Appendix D: Web Searcher's Toolkit

In this part of the book, we'll tell you about a handful of Windows and Macintosh programs that we've found to be especially useful for getting the most out of the Internet. All of the programs are public domain or shareware, or they are offered in free demo versions, so that you can "try before you buy." We'll explain what each of these programs is designed to do and tell you how to get them.

PART 1

SEARCH BASICS

Chapter 1:	Search Engines and How They Work	3
Chapter 2:	Unique Keywords	17
Chapter 3:	Basic Search Tools	23
Chapter 4:	Tips and Techniques	35

Search Basics

If you're new to online searching, we suggest that you read the four chapters in this part of the book from start to finish. With this information under your belt, you'll get a lot more out of the specific search engine chapters in Parts 2, 3, and 4.

Chapter 1, **Search Engines and How They Work**, brings you up to speed fast on what search engines are all about and how they accomplish the mammoth task of collecting information from Web sites around the world.

Chapter 2, **Unique Keywords**, lays out specific steps for choosing the best search terms for your Web searches. Coming up with the right keywords is the essence of effective searching, and this chapter will show you how to do just that.

Chapter 3, **Basic Search Tools**, builds on the keyword concept by showing you how to enter and combine search terms. Each search engine has its own way of doing things, but the basic concepts are similar from one search engine to the next. Once you know the basics, you'll be well equipped to deal with almost any search engine.

Chapter 4, **Tips and Techniques**, lays out the "Seven Habits of Highly Effective Web Searchers," offers advice on customizing your Web browser, and presents some tried-and-true techniques for making your online sessions easier and more productive.

1 Search Engines and How They Work

Listening to television news shows and reading the daily paper, you could easily get the impression that the Internet and the World Wide Web hold the answers to virtually any question you could possibly ask. Sign on to your favorite search engine, tap a few keys, and all the world's knowledge is there for the taking. It couldn't be simpler.

But ask anyone who has actually tried to use a search engine to answer a specific question or track down a particular fact or figure and you're likely to hear a far different story: "I can't find *anything* on the Net." Or "The search engine I used turned up so *much* information, most of it totally useless, that I gave up."

The problem is at least twofold. First, the Internet is so vast and so lacking in organization that even longtime, experienced searchers express great frustration in using it for finding information.

Second, most people who use search engines don't really understand what they are and don't bother to learn how to take full advantage of their unique information-finding features. Falling prey to the hype, perhaps, they simply type one or two words into a search form and are then unpleasantly surprised when they are presented with thousands (or even millions) of Web pages.

The current state of search engines can be compared to a phone book that is updated irregularly, is biased toward listing more popular information, and has most of the pages ripped out.

—Steve Lawrence and Lee Giles, authors of a widely publicized 1999 study of search engines for the NEC Research Institute

The Perils of Internet Searching

One of our favorite descriptions of what it's like to search the Internet is offered by Clifford Stoll, veteran searcher and author of *Silicon Snake Oil* and *The Cuckoo's Egg*. By way of background, and to help you appreciate Mr. Stoll's formidable research skills, we should mention that *The Cuckoo's Egg* recounts how he used the Net to track down the notorious "Hanover Hacker"—a trail that also led to a KGB-backed spy ring. That was in the late 1980s, before most of the world had even heard of the Internet.

But despite his considerable research background and demonstrated ability to find information, here's how Mr. Stoll once described the Internet and his struggle to locate the answer to a specific question:

What the Internet hucksters won't tell you is that the Internet is an ocean of unedited data, without any pretense of completeness. Lacking editors, reviewers, or critics, the Internet has become a wasteland of unfiltered data. You don't know what to ignore and what's worth reading.

Logged onto the World Wide Web, I hunt for the date of the Battle of Trafalgar. Hundreds of files show up, and it takes 15 minutes to unravel them—one's a biography written by an eighth grader, the second is a computer game that doesn't work, and the third is an image of a London monument. None answers my question...



Figure 1.1 Google emphasizes *searching*—rather than browsing the Web by topic.



Figure 1.2 Yahoo's home page gives you two ways to look for information. You can type a query in the search form or click on one of 14 subject areas and work your way through the site's comprehensive topic directory.

Search Engines to the Rescue

It's not likely that the Internet will become less chaotic any time soon. In fact, it's a virtual certainty that it won't. After all, unlike America Online (AOL) and Microsoft Network (MSN), no one is "in charge" of the Internet. Once you understand and accept that, you'll be better able to deal with the chaos—and perhaps more forgiving when you can't find what you're looking for.

That said, let's move on to what you *can* do something about—developing a better understanding of search engines and how to use them to make sense out of the Internet's "ocean of unedited data."

Search engines defined

A *search engine* is a tool that lets you explore databases containing the text from hundreds of millions of Web pages. When the search engine software finds pages that match your search request (often referred to as *hits*), it presents them to you with brief descriptions and clickable links to take you there.

Google (**Figure 1.1**) and other leading search engines focus primarily on providing a powerful search capability and offering the largest possible database of Web pages. But in order to compete with Yahoo (**Figure 1.2**)—the most popular site on the Net—they also give you the option of using a multi-level *topic directory* to browse a much smaller collection of Web pages for information on a given subject.

Topic directories are prepared by human beings, who spend their days visiting, selecting, and classifying Web sites based on content. Yahoo maintains its own staff of "Yahoo Surfers" to perform this function and is widely considered to be the premier topic directory on the Net. Most of the other leading search engines rely on either LookSmart (`www.looksmart.com`) or the Open Directory Project (`dmoz.org`) for their topic directories.

The Big Six Search Engines

Of all the search sites available today, several are especially impressive: AltaVista, Google, HotBot, Lycos, Northern Light, and Yahoo. What sets them apart and earns them the distinction of being included in what we call the "Big Six" search engines?

- They've been around for two or more years.
- Their creators know the Net and the Web and have designed their sites and the underlying software with the primary goal of making searching as easy as possible.
- They all offer exceptionally large searchable databases, excellent help information, powerful search tools, and an overall "look and feel" that's conducive to searching.

Of course, like virtually all search engines these days, they also carry advertising of one form or another (**Figure 1.3**). That's how they finance their operations and avoid having to charge subscription fees for the use of their services.

Some have begun offering *portal features* like free e-mail, news, chat, stock quotes, and weather reports (**Figure 1.4**)—stuff that's completely unrelated to searching but presumably helps attract visitors (and thus advertisers). The idea is to get you to think of the site not just as a place to search but as your *portal to the Internet*—the first place you visit whenever you go online.

Fortunately, in the case of the Big Six, the advertising and ancillary features interfere very little (if at all) with your ability to search effectively. Thus, we strongly recommend that you get to know at least a couple of these sites and choose one as your primary search engine.



Figure 1.3 Advertising messages and corporate sponsorships, like this one for Network Solutions, are common on search engine sites.

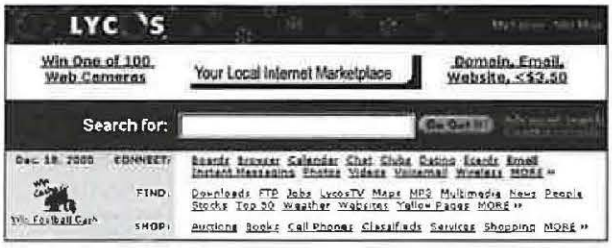


Figure 1.4 Extra features like news, weather, stock quotes, chat, and so forth get equal billing with searching at many search engine sites. Lycos is no exception.

If you already have a favorite Internet portal (AOL, Excite, MSN, and Netscape all have large followings), consider making one of the Big Six your *backup* search site—the place you go for a "second opinion" when you can't find exactly what you're looking for using your portal site's search feature.

✔ Tip

- One good way to quickly size up a search engine or portal is to look at its search tips and help information. Sites that are serious about searching will offer clear, detailed instructions for new and advanced users.

Where to find the Big Six

We'll have much more to say about each of the Big Six search engines in Part 2 of this book, but if you'd like to visit them now, here's where to find them:

AltaVista	`www.altavista.com`
Google	`www.google.com`
HotBot	`www.hotbot.com`
Lycos	`www.lycos.com`
Northern Light	`www.northernlight.com`
Yahoo	`www.yahoo.com`

✔ Tips

- The official address or URL (Uniform Resource Locator) for any Web site always begins with `http://`. However, with the latest versions of Web browsers like Microsoft Internet Explorer (IE) and Netscape Navigator, you don't have to include this portion of the address.
- You can often even omit the `www` and `com` portions of a Web address (**Figure 1.5**). It doesn't work for multi-word names separated by periods, but try it with any of the addresses in our search engine list.

  Just type the search engine name (`altavista`, `google`, `hotbot`, etc.) in your Web browser's Address or Location box and then hit Enter (or Ctrl Enter if you're using IE for Windows).
- For Web addresses ending in anything other than `com` (`edu`, `gov`, `net`, `org`, etc.), it's a good idea to type the complete address, like `www.whitehouse.gov`. IE and Netscape are working on enhancements to their methods of dealing with partial Web addresses, but both currently make the assumption that if you type, say, `whitehouse`, you're looking for `www.whitehouse.com`, which happens to be a pornography site instead of the official Web site for the home of U.S. presidents.



Figure 1.5 HotBot's official address is `http://www.hotbot.com`. But with the latest versions of Microsoft Internet Explorer and Netscape Navigator, you can just type `hotbot`.



Figure 1.6 Scooter is the spider responsible for building AltaVista's database of Web pages.

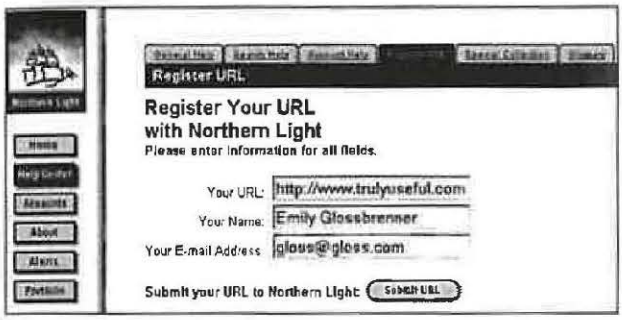


Figure 1.7 Here's a typical search engine submission form. Northern Light and other search engines try to make it as easy as possible for people to alert them to new sites.

How Search Engines Work

Search engines are designed to make it as easy as possible for you to find what you want on the Internet. But with hundreds of millions of Web pages stored on computers all over the world—and more being added all the time—how can search engines possibly collect them all? And what do they do with the information once they get it?

Spiders and indexes

Search engines do their data gathering by deploying robot programs called *spiders* or *crawlers*. Some even have names and personalities, like "Scooter," the AltaVista spider (**Figure 1.6**). Scooter and other such programs are designed to track down Web pages, follow the links they contain, and add any new information they encounter to a master database or *index*.

You don't really need to know the specifics. The key point to understand is that each search engine has its own way of doing things. Some have programmed their spiders or crawlers to search for only the *titles* of Web pages and the first few lines of text. Others snare every single word, ignoring only the graphics, video, sound, and other multimedia files.

The spiders' work is supplemented with information supplied directly to the search engines by professional and amateur Web developers. With millions of new Web pages being created every day, search engines welcome this kind of help and make it easy for anyone to suggest a site by simply clicking on a link on the home page and completing a simple form (**Figure 1.7**).

continues on next page

It may take weeks or even months for a form to be processed using these free submission services, however. So some search engines have begun offering "express service" to customers willing to pay a fee for having a Web site added to the database within a couple of days. Others are experimenting with charging customers to have their Web pages visited more frequently by the search engine's spider program.

These services go by different names, but they're often referred to generically as *paid-inclusion* or *pay-for-submission* services. As they begin to catch on, such services should help to improve the freshness and comprehensiveness of search engine databases.

✔ Tips

- Yahoo currently charges $199 for adding a commercial Web site to its topic directory within 48 hours. (The free service can take as long as six months.)
- Inktomi, the search engine that powers HotBot and a number of other leading search sites, charges $6 to $20 per page for a one-year subscription. (The more pages you submit, the lower the cost per page.) The service promises inclusion in the database within two days and a return visit by the Inktomi spider every 48 hours.
- Don't confuse paid-inclusion and pay-for-submission services with the more controversial *pay-for-placement*, which guarantees preferential treatment in search engine rankings. The GoTo search engine (`www.goto.com`) pioneered the concept of pay-for-placement. Instead of charging a set fee for inclusion in its database, GoTo allows Web site owners to bid on search terms that are relevant to their sites. The higher the bid, the higher that Web site appears in the search results (**Figure 1.8**).

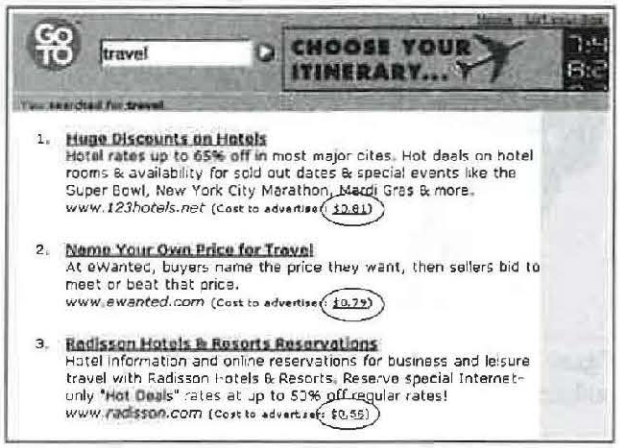


Figure 1.8 With GoTo's pay-for-placement approach to presenting search results, these Web sites paid the most for the keyword *travel* and therefore get the top three slots.

<META NAME="KeyWords" CONTENT="free, free, free, free, free, free, free, free, money, money, money, money, money, money, money, money, money, money, money, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, sex, web, pages, Mall, online, shopping, shop, mall, malls, store, stores, birthday, christmas, holiday, gift, gifts, art, collectibles, automotive, books, clothing, computers, electronics, food, beverage, restaurant, health, fitness, housewares, money, finance, bank, music, professional services, real estate, specialty shops, sports, recreation, toys, games, travel">

Figure 1.9 One technique for spamdexing or spoofing is to insert a block of text like this in the Web page. The words may or may not have anything to do with the actual content of the page.

Relevancy formulas

Every search engine uses a different method for calculating *relevancy*—how well a particular Web page matches your search request. As mentioned above, at GoTo, the method is quite straightforward: the higher the bid, the higher a site's ranking.

For most search engines, however, the relevancy formulas are a closely guarded secret. And they are constantly being changed in an effort to stay one step ahead of crafty Web developers, some of whom engage in a practice called *spamdexing* or *spoofing*—doctoring a Web page to fool search engines into putting it high on the list of search results.

One of the most common tricks is to load a Web page with words—like *free* or *money* or *sex*—that may or may not have anything to do with the site (**Figure 1.9**). The Web page may even be designed so that the words are hidden from view when you look at the page with your browser.

A search engine can respond to tactics like this by rewriting its software, changing the relevancy formula to give a *lower* ranking to pages with a lot of repeated words.

How good are the spiders?

The beauty of spiders and crawlers is that they operate around the clock. But some take longer than others to make their rounds. And, with one exception, even the best of them have located and indexed less than a third of the estimated 2.1 billion pages that are out there. (See **Table 1.1** for a comparison of several of the leading engines.)

That's why you may be successful finding a specific Web site using one search engine but not another. It's also a good reason to get into the habit of trying the same search on multiple engines.

continues on next page

Table 1.1

How Much of the Web Is Indexed?

Search Engine	Number of Web Pages Indexed	Percentage of Web Pages Indexed
Google	1.2 billion	57%
Netscape Search (powered by Google)	1.2 billion	57%
Yahoo (powered by Google)	1.2 billion	57%
Lycos (powered by FAST)	575 million	27%
HotBot (powered by Inktomi GEN3)	500 million	24%
AltaVista	350 million	17%
Northern Light	330 million	16%
Excite	250 million	12%
AOL Search (powered by Inktomi)	110 million	5%
MSN Search (powered by Inktomi)	110 million	5%

Note: The percentage figures have been calculated based on search engine sizes reported by Search Engine Watch (www.searchenginewatch.com) as of January 2001 and an estimate that the World Wide Web contains 2.1 billion pages.

At this writing, Google holds the record for the largest database—some 1.2 billion pages. (The Google database is used by both Netscape Search and Yahoo for their Web results.) FAST (the search engine that powers Lycos) is second, with 575 million pages. And Inktomi's GEN3 database (the one used by HotBot) is a close third, with 500 million pages.

But remember, even though Google is more than twice as big as FAST and Inktomi, there are still close to a *billion* Web pages that it doesn't cover. You can count on the fact that a large percentage of those Web pages *have* been found by and included in the databases of one or more of the other leading search engines.

✔ Tips

- Keep in mind that when you use a search engine to "search the Web," what you are actually searching is that particular site's database of information, collected by a spider program and supplemented with entries submitted by site creators. Each search engine you use will produce a different set of results, because each creates and maintains its own database.
- When you explore a subject using a search engine's topic directory, you're likely to get far fewer hits. But the sites you find will all have been selected and classified by human beings.

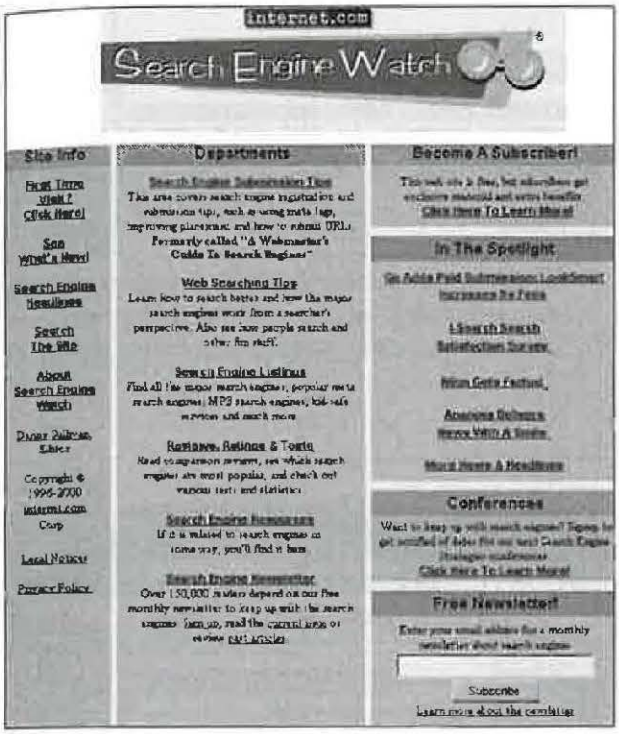


Figure 1.10 The Search Engine Watch Web site at www.searchenginewatch.com is the definitive online source for information about search engines and how they operate.

More Information for Webmasters and Others

Many search engines provide basic information about how their spiders operate—what they search for, how many sites they visit in a day, how often the master database is updated, and so forth. You may have to dig a bit to find it, but it's probably there.

Start by checking the search engine's home page for an About link. Or try the search tips and help information, usually accessible via a Help link on or near the search form.

For even more detail, look for the site's instructions to Webmasters on how to submit a URL to the search engine. To learn more about the AltaVista spider, database, and indexing methods, for example, go to the AltaVista home page (www.altavista.com) and click on Submit a Site and then on How AltaVista Works.

Resources for Webmasters

If you're responsible for designing Web pages and making sure that people can find them on the Internet, you'll definitely want to track down the information provided by each of the major engines on how their spiders operate and how their indexes are developed. But you'll also want to pay a visit to one of the Net's premier resources for Webmasters:

Search Engine Watch. This Web site (**Figure 1.10**) is hands down the best source of information we've encountered on how search engines work. Created several years ago by Internet consultant and journalist Danny Sullivan, Search Engine Watch is now owned by Internet.com (www.internet.com). But Mr. Sullivan continues to maintain the site, which is located at www.searchenginewatch.com.

continues on next page

Among its many offerings is a section called "Search Engine Submission Tips," with all the latest information on how to design a site for maximum visibility by the major engines. For people who are simply curious about search engine technology, there's a wealth of interesting information—facts and figures about search engine sizes, comparative reviews and ratings, tutorials, a glossary, and more.

✔ Tips

- If you don't see what you want on the Search Engine Watch home page, try using the site's search feature.
- Search Engine Watch publishes a twice-monthly newsletter issued via e-mail. You can sign up for free, but regular readers are encouraged to pay $79 for an annual subscription ($49 for six months), which helps to support the site and entitles you to additional reports and information not available to the general public.

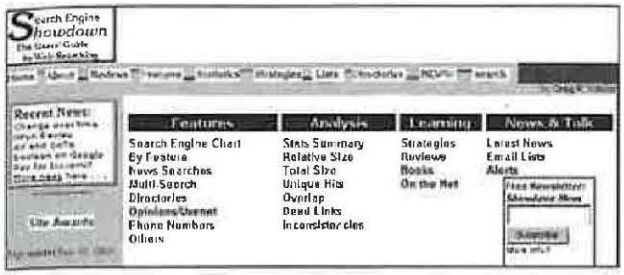


Figure 1.11 Search Engine Showdown at www.searchengineshowdown.com is aimed at *users* of search engines, rather than Webmasters.



Figure 1.12 The Web Search site at websearch.about.com is a good place to look for beginner's guides and tutorials on search engines.

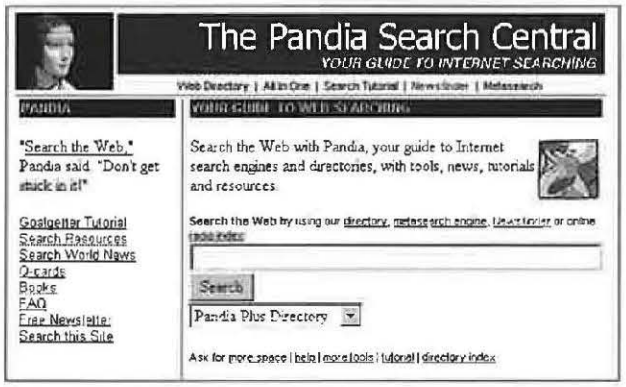


Figure 1.13 Pandia (www.pandia.com) is another excellent source of beginner's guides, tutorials, and search tools.

Other great resources

Search Engine Watch is aimed largely, though not exclusively, at Webmasters. If you're primarily interested in learning about search engines so that you can become a more effective *searcher*, we recommend these sites:

Search Engine Showdown. Created and maintained by researcher and writer Greg R. Notess, this award-winning site (**Figure 1.11**) compares and evaluates search engines from a user's perspective. Check here for regularly updated feature lists, detailed reviews, and performance reports. The site is located at www.searchengineshowdown.com.

Web Search. This site (**Figure 1.12**), hosted by Web consultant Chris Sherman, is a good source of search tips, tutorials, and feature articles. One of the About.com subject guides, it's located at websearch.about.com.

Pandia Search Central. Named after a Greek goddess of light and enlightenment, the Pandia Web site (www.pandia.com) is beautifully designed and full of great information (**Figure 1.13**). Its creators, Per and Susanne Koch, who happen to live in Norway, are dedicated to helping Internet users become more effective searchers. They have produced a truly remarkable resource for doing just that.

2

UNIQUE KEYWORDS

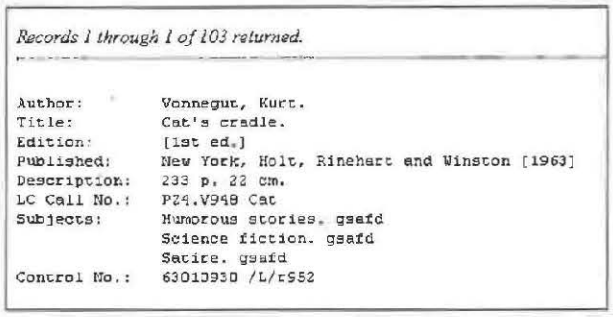
The biggest challenge with any type of online searching is choosing the right search terms or *keywords*. This goes double for Web searching because, when you use tools like AltaVista, Google, Lycos, and other search engines, what you are typically searching is the *full text* of Web pages collected automatically by spider or crawler programs.

That sounds great in theory. But ask any librarian or professional searcher—anyone who makes a living using high-powered, high-cost databases like DIALOG, Dow Jones/News Retrieval, and LEXIS-NEXIS—about full-text searching. What they'll tell you is that it's actually the most difficult type of searching in the online world.

To help you understand why, we'll take a look in this chapter at another type of database searching that most everyone is familiar with—using the electronic card catalog at the local library. Then, with that as background, we'll show you how to meet the challenge of full-text searching by choosing the right keywords for your Web searches.

The Challenge of Full-Text Searching

Traditional databases like the electronic card catalog at your local library let you do *field searches*. If you're interested in the writer Kurt Vonnegut, for example, you can search for Vonnegut in the database's *author* field to produce a list of his books. Click on a title and all the relevant information about that book will be presented in familiar, card-catalog format (**Figure 2.1**).



Records 1 through 1 of 103 returned.

```
Author:         Vonnegut, Kurt.
Title:          Cat's cradle.
Edition:        [1st ed.]
Published:      New York, Holt, Rinehart and Winston [1963]
Description:    233 p. 22 cm.
LC Call No.:    PZ4.V948 Cat
Subjects:       Humorous stories. gsafd
                Science fiction. gsafd
                Satire. gsafd
Control No.:    63010930 /L/r552
```

Figure 2.1 Search for author Kurt Vonnegut at your local library and you'll get information like this for each of his books.

If, on the other hand, you're looking for Kurt Vonnegut's biography or other information *about* him, you can search for his name in the *subject* field. Or maybe what you really want is a recent magazine article that features Kurt Vonnegut, in which case you'd use the periodicals database instead of the one for the library's collection of books.

Web search engine databases, in contrast, make no distinction among the various types of information available about Kurt Vonnegut. Do a search on his name and you'll get a hodge-podge of Web sites, many of them personal home pages created by Vonnegut fans (**Figure 2.2**).



Top 10 Displayed: Full Description | Show Titles Only | View by URL

- KURT VONNEGUT: The most comprehensive Vonnegut site in the cosmos
  URL:
- Kurt Vonnegut Index
  URL:
- Marek Vit's Kurt Vonnegut Corner
  URL:
  Excite Directory Match:
- Kurt Vonnegut -- Home Page
  URL:
  Excite Directory Match:
- Kurt Vonnegut
  URL:
  Excite Directory Match:
- Books by Kurt Vonnegut
  URL:
- Untitled
  URL:
- Kurt Vonnegut - Quotes, Bokonon, and more
  URL:
  Excite Directory Match:
- Stories, Listed by Author
  URL:
- Kurt Vonnegut Essay Collection
  URL:
  Excite Directory Match:

Next >>

Figure 2.2 Search the Web for Kurt Vonnegut and there's no telling what you'll come up with, as these results show.

Web page field searches

Field searching with Web search engines, if it's available at all, is limited to fields having to do with the Web page itself—its title, Web address or URL (Uniform Resource Locator), the date the page was created, and so forth—not the information *on* the page.

Searching these Web page fields can be quite useful, but it doesn't come close to matching the precision and sophistication of traditional database field searching. Which brings us back to our original point: To be an effective Web searcher you'll need to get quite good at coming up with the *right* keywords.

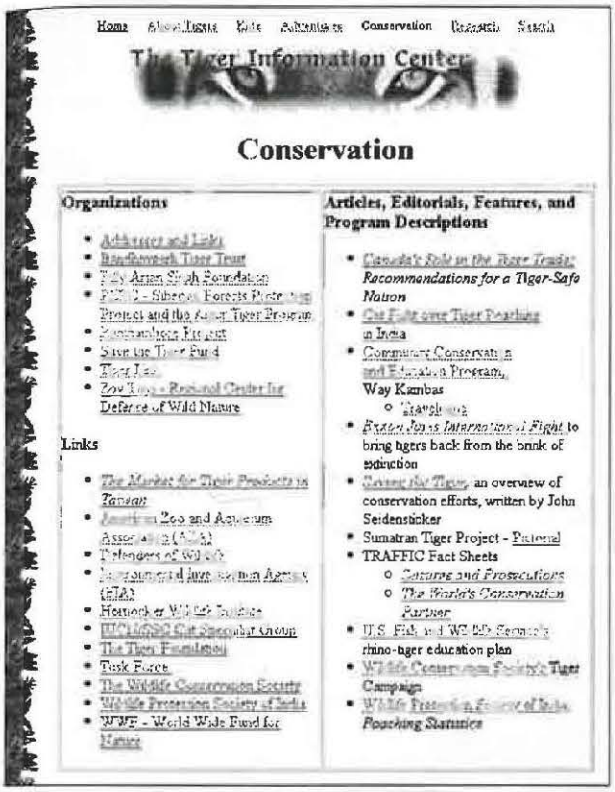


Figure 2.3 Searching for a unique phrase like `Bengal tiger` is more likely to produce a site like this near the top of your search results list than would a search for the far more common `tigers`.

Choosing the Right Keywords

Learning to come up with the most effective keywords for your Web searches will take time and practice. Don't be surprised or disappointed if you're not successful every time. After all, even the best online searchers find Web searching to be quite a challenge.

We don't pretend to have all the answers ourselves, but here are some points to keep in mind that can help you improve your Web searching success rate.

To get the best search results:

1. **Use the most unique keyword you can think of.** Take the time to think about the words that will almost certainly appear on the kind of Web page you have in mind. Then pick the most unique or unusual word from that list.

   If you're looking for information about efforts to save tiger populations in Asia, for example, don't use `tigers` as your search term. You'll be swamped with Web pages about the Detroit Tigers, the Princeton Tigers, and every other sports team that uses the word *tigers* in its name.

   Instead, try searching for a particular tiger species that you know to be on the endangered list—`Bengal tiger` or `Sumatran tiger` or `Siberian tiger`. (With some search engines, you'll need to enclose the words in quotation marks to let them know you're doing a *phrase search*: `"Bengal tiger"`.) Chances are, you'll find sites like the one in **Figure 2.3** near the top of the list.

continues on next page

2. **Make it a multi-step process.** Don't assume that you'll find what you want on the first try. Take your best shot. Then review the first couple of pages of results, paying particular attention to the sites that contain the *kind* of information you want. What unique words appear on those pages? Make a few notes and then do another search using those words.

   You might even try changing the *order* in which you type your keywords in the search form's text box. Some search engines give more weight to the first word in your query than to those that appear second or third, on the assumption that you will typically put your most important search word first.

   Consequently, a search for `endangered tiger species` might very well produce different results than a search for `tiger endangered species`.

3. **Narrow the field by searching just your previous results.** If your chosen keyword returns relatively good information but too much to review comfortably, try a second search of just those results. This is sometimes referred to as *set searching.*

   Most of the leading search engines make set searching quite easy by providing an option on the search results page that allows you to "Search within these results" (**Figures 2.4** and **2.5**). But even if the search engine doesn't provide this option, you can accomplish essentially the same thing by simply adding another keyword to your search request and submitting it again.

4. **Look for your keyword in the Web page title.** Often the best search strategy is to look first for your unique keyword in the *titles* of Web pages. All the Big Six search engines allow you to search Web page titles, but the rules for doing so vary slightly depending on which search engine you're using.

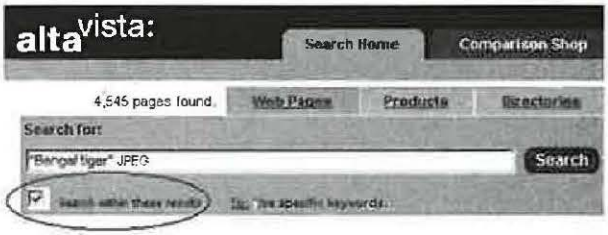


Figure 2.4 Set searching is easy with AltaVista. Once you've located sites offering information about Bengal tigers, for example, you can look for JPEG images of Bengal tigers by doing a second search of just those results.

View, Copy or Download *Bengal Tiger*



Figure 2.5 Here's the JPEG image we found using AltaVista's set-searching option.



Figure 2.6 Notice that the phrase *Middle Ages* appears in all the Web page titles shown here. That's because we specified that we wanted to search for `title:"Middle Ages"`.

With AltaVista, HotBot, Lycos, and Northern Light, you can put a `title:` field-search term right in your query. For example, if you're looking for information about marriage customs in the Middle Ages, start with a search of Web pages that have the words *Middle Ages* in the title, like this: `title:"Middle Ages"` (**Figure 2.6**). Then do a second search of just those results, looking for `"marriage customs"`.

With Google, follow the same steps, but instead of `title:`, use the search term `allintitle:"Middle Ages"`.

With Yahoo, use `t:"Middle Ages"` to find the phrase *Middle Ages* in topic categories and Web page titles.

Notice that in each of these examples, there is no space between the colon and the phrase that we are searching for.

5. **Find out if case counts.** It's important to know whether the search engine you're using pays attention to uppercase and lowercase letters in your keywords. Will a search for `Java`, the Sun Microsystems program, also find sites that refer to the program as `JAVA`?

   Some search engines ignore case completely. But with others (AltaVista and HotBot, for example), if you use *any* uppercase letters, the search engine will assume you want to look for only that combination of upper- and lowercase.

   Consequently, for the broadest possible search, it's usually a good idea to use all lowercase letters for your queries.

continues on next page

6. **Use initial caps to search for proper names.** As we've said, it's usually a good idea to put your queries in lowercase because that gives you the broadest possible search. The keyword `java` will find references to *java*, *Java*, *JAVA*, and even *jaVa*.

   But when you're looking for a person's name, a geographical location, a book or movie title, or anything else that you might reasonably expect to be presented with initial caps, by all means type your query that way. Doing so will greatly reduce the number of *false drops*, pages that contain the words in your query but are completely off the mark.

7. **Check your spelling.** If you've used the best keyword you can think of and the search engine comes back with the message "No results found" or some such, check your spelling before you do anything else.

   Nine times out of ten, the reason a search engine comes up empty-handed is because of a spelling or typing error (**Figure 2.7**). At least, that's been our experience.

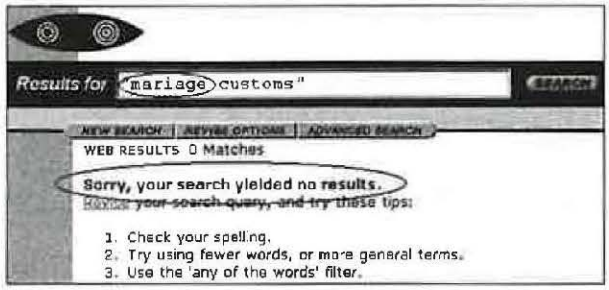


Figure 2.7 Spelling errors—typing `mariage` instead of `marriage`, for example—are virtually certain to stump the search engine and produce a "No results found" message.

✔ Tip

- The Quick Reference guides in Parts 2, 3, and 4 of this book (and repeated in Appendix A) are a good place to look for specifics on field searching and case-sensitive searching with different search engines.

3

BASIC SEARCH TOOLS

As we've said before, choosing the right keywords is the essence of effective searching. It's especially important for *Web* searching, because there's so much information out there—some of it worthwhile, much of it totally useless—and there's no single company or organization responsible for organizing and making sense of it.

But coming up with good, unique keywords is actually just half the battle. You also need to know how to *enter and combine* keywords, taking advantage of the basic search tools offered by the major search engines. Your success rate will improve dramatically once you learn what these tools are and when and how to use them.

Specific procedures vary from one search engine to the next, of course. In fact, some search engines have a different set of instructions depending on whether you're using their simple or advanced versions. But certain general concepts apply across the board. That's what we'll cover here.

Once you have this basic understanding of the tools that are available, you'll be better equipped to deal with just about any search engine you encounter on the Web. The chapters in Parts 2, 3, and 4 will fill in the details for some of the best of them.

Search Tools at a Glance

Here are the basic search tools you'll read about in this chapter:

- Plain-English Searches
- AND Searches
- OR Searches
- NOT Searches
- NEAR Searches
- Nested Searches
- Wildcards
- Stopwords

Searching in Plain English

For new Web searchers, one of the best tools going is *plain-English* or *natural-language* searching. Most of the leading search engines have developed techniques that are amazingly good at finding what you want based on a simple question—especially if the question includes at least one unique keyword or phrase.

One problem with plain-English searches like the one shown in **Figure 3.1** is that they typically produce a very large number of hits. But as long as the information you're looking for shows up at or near the top of the list (**Figure 3.2**), it doesn't really matter.

Another problem you're likely to encounter when you type your first plain-English search is fitting it into the tiny search box that's provided by many search engines. Don't be misled into thinking that you're limited to typing a very short question. The box in Figure 3.1 may look small, but at most search sites, you can actually type as long a question as you like. The first part will simply scroll out of view on the left-hand side of the box.



Figure 3.1 Plain-English searches like this often produce excellent results. Many search engines encourage you to simply "ask a question."

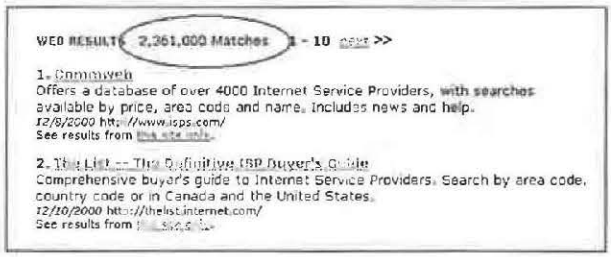


Figure 3.2 The question posed in Figure 3.1 identified more than two million sites, but the first two look extremely promising, so who cares?

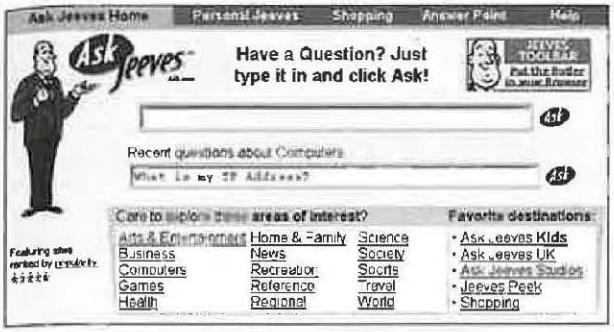


Figure 3.3 The Ask Jeeves Web site at `www.ask.com` specializes in providing high-quality answers to plain-English questions.

✔ Tips

- Except for Yahoo, all of the Big Six search engines do a pretty good job of handling plain-English searches. But if you really like this method of looking for information on the Web, be sure to try Ask Jeeves (`www.ask.com`). Its Web database is very small and carefully selected by a staff of researchers. And unlike the Big Six, when you "ask Jeeves" a question (**Figure 3.3**), he'll attempt to provide you with the best *single* answer—rather than every possible match in the database.

- Word order and phrasing can make a big difference with plain-English searches. If you don't find what you're looking for with a direct question, try rephrasing your query, putting the most unusual word or phrase first:

  Query 1: `What companies offer "parental control software"?`

  Query 2: `"parental control software" companies`

  Think of it as a two-step process. The direct-question approach of Query 1 may not work. But it will help you clarify your search request and zero in on the most unusual term, which you can place first in Query 2.

Searching for Multiple Words and Phrases

The most important thing to find out about any search engine is how to look for *multiple* words and phrases. Plain-English queries are great, and they can probably handle many of your search requirements. But sometimes you need more precision than you can get from a plain-English search.

Here are the features to look for:

- **AND Search**. How do you tell the search engine that you want to find Web sites that include references to Keyword A *and* Keyword B?
- **OR Search**. How do you specify that it's not necessary for *both* keywords to appear in the results, as long as one *or* the other is present?
- **NOT Search**. How do you look for one keyword while specifically *excluding* another?
- **NEAR Search**. How do you find two words or phrases in close *proximity* to one another?

The technical term for this type of searching is *Boolean* searching, and AND, OR, NOT, and NEAR are among the traditional *Boolean operators*.

Over the next few pages, we'll explain how you can do AND, OR, NOT, and NEAR searches with the leading search engines. We'll also show you how to use parentheses to combine words and phrases into more complex queries.

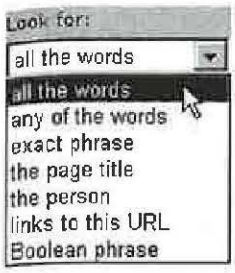


Figure 3.4 AND searches can sometimes be done by simply choosing `all the words` from a menu like this one.

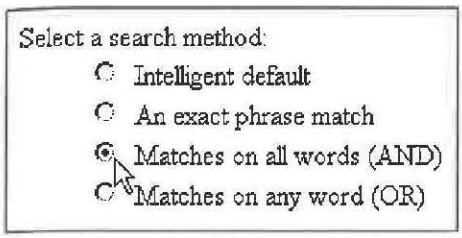


Figure 3.5 Some search engines let you specify an AND search by clicking on a radio button.

Common Methods for Doing AND Searches

- Search for `Monet AND Renoir`
- Search for `+Monet +Renoir`
- Search for `Monet Renoir` and select the `all the words` option

AND Searches

Searching for two or more keywords—both of which *must* appear in the results—is an excellent strategy for doing precise searches and greatly reducing the number of search results you have to consider. Some search engines offer more than one method for doing AND searches, so you can choose the one you like best.

The most common way to tell a search engine that you want to do an AND search is to put a plus sign in front of each word or phrase that *must* appear in the results: `+Renaissance +sculpture`.

Other search engines require that you actually use the Boolean operator AND to combine words or phrases: `needlepoint AND supplies`, for example, to look for references to both terms. You may even be required to use full caps for AND. To avoid having to remember whether or not full caps are necessary, it's a good idea to get into the habit of using full caps all the time for AND.

Some search engines make AND searching extremely easy by offering it as a menu option or radio button. You simply type two or more unique keywords or phrases and then choose `all the words` from a drop-down menu (**Figure 3.4**) or click on the radio button with a similar label (**Figure 3.5**).

✔ Tips

- You'll find that some search engines default to doing an AND search, while others default to doing an OR search. HotBot, for example, assumes that you want to do an AND search unless you tell it differently. AltaVista, on the other hand, defaults to OR searching.
- Whenever you do an AND search, it's a good idea to put the most unusual search term first. That's because some search engines pay attention to word order and factor it into their relevancy rankings.

OR Searches

OR searches cast a much broader net than AND searches and may result in a very large number of hits for you to consider. They make good sense, though, when there's more than one way that the person, object, or thing you're looking for might be referred to in a Web page or other document. If you're doing research on the Clinton presidency, for example, you'd want to look for references to *President Clinton* or *Bill Clinton* or *William Jefferson Clinton.*

Some search engines do an OR search by default. In other words, if you type several words or phrases, leaving a space between each one, the search engine assumes that you want to find references to *any one* of them: `"President Clinton" "Bill Clinton" "William Jefferson Clinton"`.

Other search engines require that you actually type the word OR between the words or phrases: `Greenspan OR "Federal Reserve Chairman"`. Full caps may or may not be required for the word OR, but it's a good habit to get into.

Finally, some search engines offer OR searching as a menu option or radio button. You type your search words and phrases and then choose the option labeled `any of the words` or some such (as opposed to `all the words`). See **Figures 3.6** and **3.7** for examples.

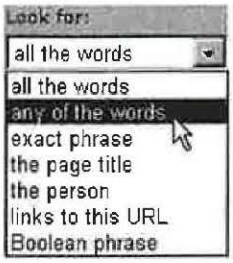


Figure 3.6 Choosing any of the words from a menu like this is the way you tell some search engines to do an OR search.

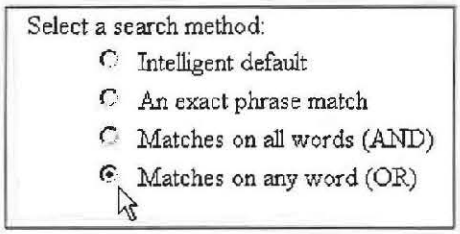


Figure 3.7 Here's another approach to OR searching. Just click on a radio button to indicate that the search engine should look for any of the words you've typed in the search form.

Common Methods for Doing OR Searches

- Search for `UPS U.P.S.`
- Search for `UPS OR U.P.S.`
- Search for `UPS U.P.S.` and select `any of the words` option

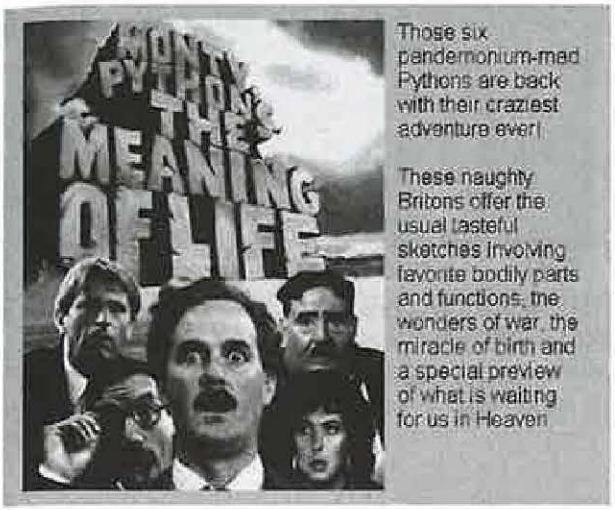


Figure 3.8 A NOT search like `python -monty` or `python NOT monty` can help you avoid information like this and find snakes instead.

NOT Searches

Here's the example you'll come across again and again to illustrate when to use a NOT search: You're looking for information on snakes and key in a search for `python`. Much to your dismay, your search results are dominated by Web pages aimed at devoted Monty Python fans (**Figure 3.8**). What to do?

The answer, of course, is to *exclude* Monty Python pages with a NOT search. Usually that means putting a minus sign in front of the word you want to avoid: `python -monty`. (Notice that there's no space between the minus sign and the word that's being excluded.)

Some search engines allow (or require) the use of the word NOT (or AND NOT) to exclude a word: `python NOT monty` or `python AND NOT monty`.

✔ Tips

- NOT can be a very powerful operator. But it also makes it quite easy to unwittingly throw the baby out with the bath water. Suppose the definitive Web site on pythons—"The Master Python Page"—happens to have been created by someone named Monty Shields. The NOT search in our example could very well prevent you from ever finding it.
- By all means learn to use NOT searches. Just be aware of the potential for inadvertently excluding good material.

Common Methods for Doing NOT Searches

- Search for `python -monty`
- Search for `python NOT monty`
- Search for `python AND NOT monty`

NEAR Searches

Sometimes you're not just interested in finding multiple keywords that are *mentioned* in the same document. You want to be able to specify that they appear in *close proximity* to one another. That's what NEAR searching is all about.

AltaVista is one of the few search engines that offers NEAR searching. And it's only available as part of the Advanced and Power Search capabilities, not on the main AltaVista search form. The format is: `Constitution NEAR "Electoral College"`.

See **Figures 3.9** and **3.10** for an example of NEAR searching with AltaVista.

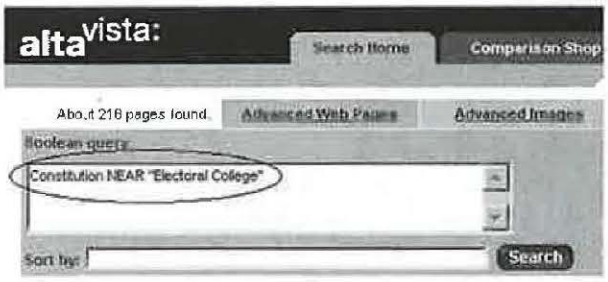


Figure 3.9 With AltaVista, you can do a NEAR search like this to find references to the word *Constitution* within 10 words of the phrase *Electoral College*.

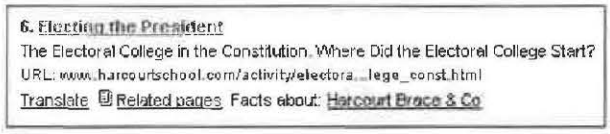


Figure 3.10 Among the first six sites found by the NEAR search in Figure 3.9 was this one, which looks like it might be right on the money.

✔ Tip

- Just how near is NEAR? AltaVista defines it to mean that the terms must appear within 10 words of each other.

NEAR Searches

- Use **AltaVista** (Advanced Search and Power Search only)
- Search for `Japan NEAR climate` to find the two terms within **10 words** of each other.

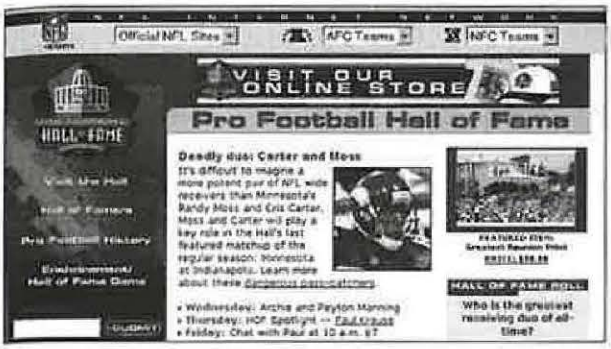


Figure 3.11 It doesn't take a complicated nested search to find the Web site for the Pro Football Hall of Fame in Canton, Ohio.

Parentheses and Nested Searches

Some search engines allow you to create more complex queries by grouping AND, OR, NOT, and NEAR statements using parentheses. For example, you could use a search like `Canton NEAR (Ohio OR OH)` to find any reference to Canton, Ohio, whether the state name is spelled out in full or abbreviated.

You can even create what are called *nested searches*, which can get quite complex, with one search statement "nested" within another: `Canton NEAR ((Ohio OR OH) AND ("Pro Football" NEAR "Hall of Fame"))`.

Our advice is to use parentheses in your searches sparingly, and only when you really need them. Complex nested searches are fine for the professionals. But for most search tasks, and especially for new online searchers, they're not worth the `mental NEAR ((energy OR effort) AND aggravation)`!

You'll be far better off concentrating on coming up with good, unique keywords and sticking to simple combinations like AND, OR, NOT (and possibly NEAR) searches. `Canton AND "Pro Football" AND "Hall of Fame"`, for example, is much more straightforward (and less error-prone) than the complex nested search presented above, and will very likely lead you to a page like the one shown in **Figure 3.11.**

Tips for Using Parentheses and Nested Searches

- Use parentheses sparingly if at all.
- Unique keywords combined with AND, OR, NOT, and NEAR often work just as well and are less error-prone.
- Complex nested searches are best left to the professionals.

Using Wildcards

Searching with *wildcards*—sometimes referred to as *truncation*—means using a special character, typically an asterisk, to indicate that you want to look for variations on a particular word, like `medic*` to find references to *medical, medicine, medicinal,* and *medication.*

Rules differ from one search engine to the next, but most require that you use at least three other characters along with the asterisk. And it's usually a good idea to also include at least one other unique keyword in your query (**Figure 3.12**). Otherwise, your search may take painfully long and return too many hits to be truly useful.

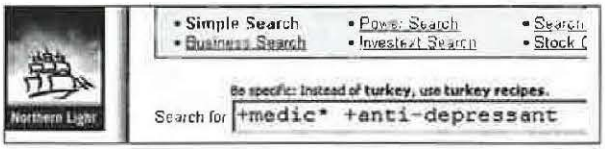


Figure 3.12 A wildcard search is likely to be excruciatingly slow unless you include at least one other term in your search request, as shown here.

✔ Tips

- Some search engines allow you to place wildcard characters in the middle (or even at the beginning) of a word.
- You may also be given a choice between two different wildcard symbols, depending on whether you want to search for *multiple* characters or a *single* character.
- For specifics on using wildcards with a particular search engine, see the Quick Reference guides in Parts 2, 3, and 4 of this book.

Tips for Using Wildcards

- Use wildcards to look for spelling variations and alternate word endings.
- The wildcard character is usually (but not always) an asterisk: `medic*`.
- Use at least three letters along with the wildcard character.
- For best and fastest results, combine the wildcard search term with at least one other unique keyword.

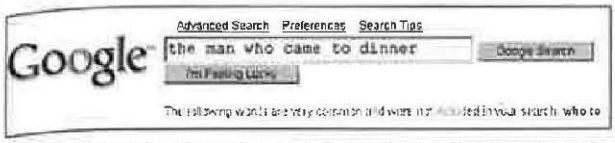


Figure 3.13 The Google search engine tells you exactly what words have been ignored from your query, in this case *who* and *to*.

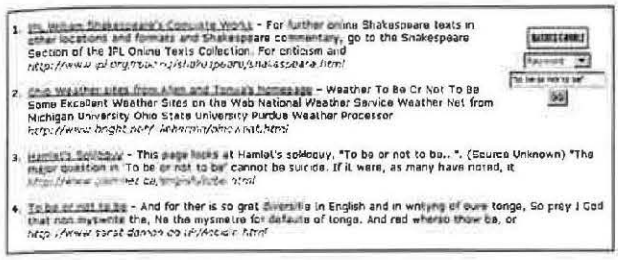


Figure 3.14 A Lycos search for "to be or not to be" finds Hamlet's soliloquy, even though the phrase is composed entirely of common stopwords. The secret is to enclose the words in quotes.

Stopword Tips

- Every search engine ignores certain very common words (*stopwords*) like *the, to, with,* the Boolean operators, etc.
- To look for words that might be confused with Boolean operators, put them in double quotes: `Portland NEAR "OR"`.
- Some search engines will recognize stopwords that are included in phrases enclosed in double quotes: `"to be or not to be"`.

Dealing with Stopwords

You may come across the term *stopwords* at search engine sites, often in the help information or search tips. Stopwords are words that search engines ignore because they are too common, or because they are reserved for some special purpose.

The list varies from one search engine to the next, but it typically includes words like *a, an, any, the, to, with, from, for, of, that, who,* and the Boolean operators AND, OR, NOT, and NEAR. We're not aware of any major search engine that publishes its complete list. But Google does the next best thing—it tells you if it has ignored one or more of the words in your query (**Figure 3.13**).

Should you need to use a stopword as part of a search, you can sometimes signal the search engine *not* to ignore it by setting it off in double quotes: `Portland NEAR "OR"`.

Some (though not all) search engines also pay attention to stopwords that are included as part of a phrase: `"The Man Who Came to Dinner"` or `"to be or not to be"` (**Figure 3.14**). Of the Big Six search engines, AltaVista, Lycos, Northern Light, and Yahoo all do a pretty good job of recognizing and acting on stopwords that are included in phrases.

✔ Tips

- Most major search engines offer all of the basic tools described in this chapter. But keep in mind that the specifics vary from one to the next. In some cases, there's even a different set of rules depending on whether you're using the search engine's simple or advanced search form.
- The chapters in Parts 2, 3, and 4 of this book will introduce you to many of the specifics. Be sure to also check the search tips and help sections for any search engine you decide to use on a regular basis to learn about new features that may have been added.

4

Tips and Techniques

So far we've covered search engines and how they work, the importance of choosing unique keywords, and the basic search tools offered by most search engines on the World Wide Web. We'll wrap up this part of the book with some specific tips and techniques that will help you get the most out of the time you spend online—no matter what search engine you decide to use on a regular basis.

We'll start with what we call "The Seven Habits of Highly Effective Web Searchers." Then we'll offer suggestions for customizing your Web browser so that you can access your favorite search tools automatically. We'll conclude with some advice on keyboard shortcuts and other timesavers.

The Seven Habits at a Glance

1. Develop the Internet habit.
2. Use the best tool for the job.
3. Read the instructions.
4. Choose unique keywords.
5. Use multiple search engines.
6. Consider the source.
7. Know when to look elsewhere.

The Seven Habits of Highly Effective Web Searchers

We're not professional searchers, but we've logged a lot of hours over the years searching the Internet and other online systems. We have a pretty good idea of what works and what doesn't. Here are our recommendations for effective Web searching, organized into seven steps or "habits" that can make you a better searcher.

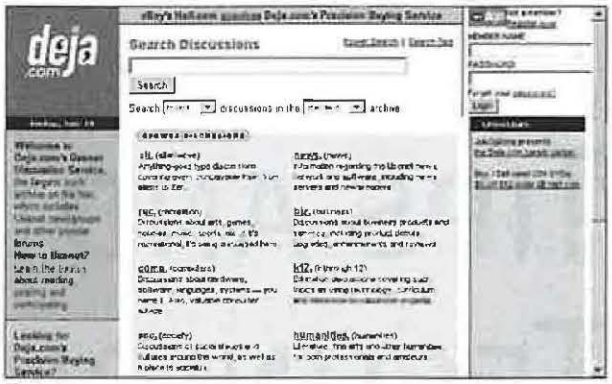


Figure 4.1 Deja (www.deja.com) is the ideal tool for searching newsgroups. No other site even comes close.

1. Develop the Internet habit.

When you have a question about anything—and we mean *anything*—your first step in nearly every case should be to check the Net. The answer may lie deep within a company-sponsored Web site, or in a newsgroup posting from two years ago, or among the millions of listings in a white or yellow pages directory, or somewhere else. But with the right search tool and search strategy, chances are you can find it.

2. Use the best tool for the job.

For day-to-day searching, you can't go too far wrong with any of the Big Six search engines (AltaVista, Google, HotBot, etc.). Don't forget, though, that each has its strong points—summarized in a sidebar on page 37 and covered in more detail in Part 2 of this book—that you'll want to consider when choosing the one to use for a particular job.

Keep in mind, too, that as good as these search engines are, they're not the best tools for every job. For example, some search engines give you the option of searching Usenet newsgroups as well as the Web. But to take full advantage of newsgroups, the best tool is Deja (www.deja.com), a search engine designed and optimized for newsgroup searches (**Figure 4.1**).



Figure 4.2 Search Engine Guide (`searchengineguide.com`) is a great place to look for special-purpose search engines. Browse by category or use the search feature, as we've done here to look for sports-related search engines.

Most general-purpose search engines that offer a newsgroup search feature use Deja to "power" it. But they typically provide only a *subset* of the search capabilities you'll find at the Deja site.

Part 4 of this book will introduce you to Deja and some of our other favorite special-purpose search engines, like Topica (`www.topica.com`) for locating Internet mailing lists and the Argus Clearinghouse (`www.clearinghouse.net`) for finding subject guides to the Net.

To track down other special-purpose engines using a searchable directory, visit Search Engine Guide at `searchengineguide.com` (**Figure 4.2**).

Big Six Search Engines and What They Do Best

- **AltaVista** is a good choice for finding obscure facts and figures. It's one of the few search engines to offer full Boolean and case-sensitive searching, as well as a variety of field-search options to help target a search.
- **Google** covers more of the Net than any other search engine. When thoroughness counts, be sure to check Google. Its method of ranking Web sites based on *link popularity* (the more links to a particular site, the higher its ranking) works especially well for general searches.
- **HotBot** makes it exceptionally easy to search for multimedia files and to locate Web sites by geography. If you want the power of AltaVista with a much simpler interface, go with HotBot.
- **Lycos** is another good choice for doing multimedia searches. Use it as well for finding phrases containing common stopwords (for example, `"to be or not to be"`). Unlike some search engines, Lycos won't ignore stopwords in phrases.
- **Northern Light** lets you simultaneously search both the Web and a "special collection" database of articles, transcripts, and other documents not readily accessible on the Net.
- **Yahoo** has the best, most detailed Web directory, making it an excellent choice for exploring a subject to find out what's available on the Net.

✔ Tip

- Usenet newsgroups (or *newsgroups* for short) are freewheeling "global conversations" on virtually every subject imaginable. They've been around far longer than the World Wide Web and are an excellent source of advice, personal opinions, and commentary. Think of them as expanding your circle of acquaintances when you're looking for things like recipes, travel tips, software fixes or workarounds—you name it.

3. Read the instructions.

No two search engines work exactly the same way, so it pays to read the online help and search tips provided at the site. For example, some search engines default to doing an AND search—type several words in the search box and the engine will assume you want to find *all* of them in your search results. Other search engines take the same request and by default perform an OR search—returning pages that contain any one of the words but not necessarily all of them.

At a minimum, you need to learn how to do AND, OR, and phrase searches at any site you're going to use on a regular basis. Another important point to zero in on is *case-sensitivity*—whether the search engine recognizes and acts upon uppercase and lowercase letters in your search request. Some search engines offer case-sensitive searching, others ignore case completely.

4. Choose unique keywords.

Before launching a search, take the time to think about what unique words or phrases are likely to appear in the information you want to find and try them first. To locate sites devoted to impressionist painters, for example, you might try `Monet AND Renoir AND Degas`. That's sure to produce better results than a search for `impressionists` or `painters`.



Figure 4.3 A few words of a poem enclosed in double quotes are often all it takes to locate the complete work, as shown here.

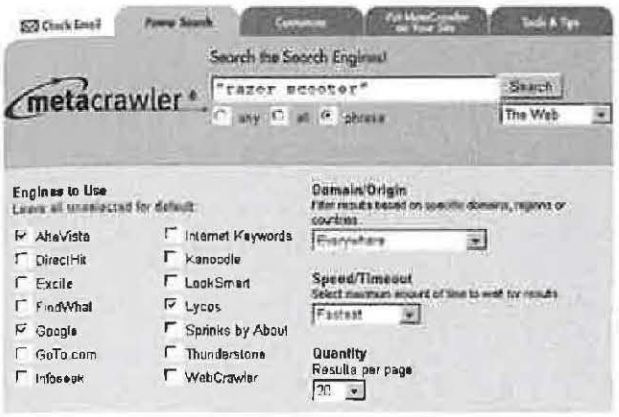


Figure 4.4 With MetaCrawler and other metasearch services, you can direct your query to multiple search engines—in this case AltaVista, Google, and Lycos.

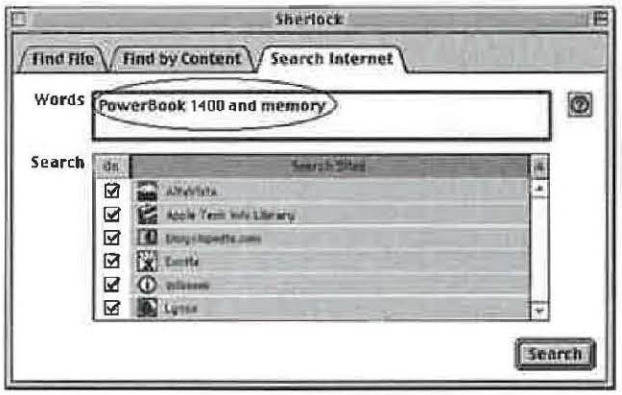


Figure 4.5 Mac users have a built-in metasearch capability called Sherlock. Type a word or phrase in the Words text box as shown here, select the engines you want to query, and click on Search.

What about the complete text of a famous quotation, literary work, or even a joke you heard at the office? Try a phrase search on some small portion that you remember: `"tiger tiger burning bright"` to locate the William Blake poem (**Figure 4.3**). Or `"man walks into a bar"` for that joke you'd like to add to your repertoire.

5. Use multiple search engines.

Every search engine has its own way of doing things and none of them covers *everything*. In fact, as you may recall from our discussion of spiders in Chapter 1, even Google—which currently holds bragging rights to the largest database of Web pages—has yet to find and index more than 40 percent of the Web. So when thoroughness counts, you should plan on using several search engines.

✔ Tips

- You can automatically submit a query to multiple search engines using what are called *metasearch* services. MetaCrawler (`www.metacrawler.com`) is one of the oldest and most popular (**Figure 4.4**). Others with large followings include Dogpile (`www.dogpile.com`), InFind (`infind.com`), ProFusion (`www.profusion.com`), and CNET Search.com (`www.search.com`). The idea has caught on with many professional and casual searchers, who find that it can be a real timesaver.
- Macintosh users running Mac OS 8.5 and later versions have a *built-in* metasearch capability. Part of the operating system's Sherlock application, the Internet Search tab lets you specify a word or phrase as well as the sites you want to search (**Figure 4.5**).

continues on next page

- The major drawback to using a metasearch service (or the Mac's Sherlock metasearch feature) is that it takes your query and reduces it to its *simplest* form—stripping out quotation marks, for example, and any other punctuation and search terms that can't be handled by *all* the engines on the metasearch service's list.

  That may or may not be a problem, depending on the nature of your search. But it's not the same—and won't produce the same results—as going yourself to, say, three different search engines and crafting a query for each one that follows the prescribed format and takes advantage of each engine's unique search features.

- We've also found metasearch engines to be more prone to timeouts than regular search engines. And they typically return only a small number of results (10-50) from any given search engine.

6. Consider the source.

Just because it's on the Net doesn't mean the information is either accurate or true. After all, virtually anyone can "publish" anything on the Internet and the World Wide Web. So be skeptical at all times.

If the information is on a Web site, try to determine:

- What person or organization created the information?
- What's the motivation behind it?
- When was the material last updated?

The same goes for newsgroup postings, where unscrupulous marketers sometimes plant positive comments about their own products and negative ones about their competitors'—making it appear as though the comments were made by actual users of the products.

Multiple Search vs. Metasearch

It might take a bit more time to visit two or three search engines and conduct separate queries at each one. But you're likely to get better results in the long run.

Metasearch engines and the Mac's Sherlock feature have several major limitations:

- Phrase and Boolean searching are generally not available.
- Most return only 10-50 hits from any given search engine.
- They are highly subject to timeouts.

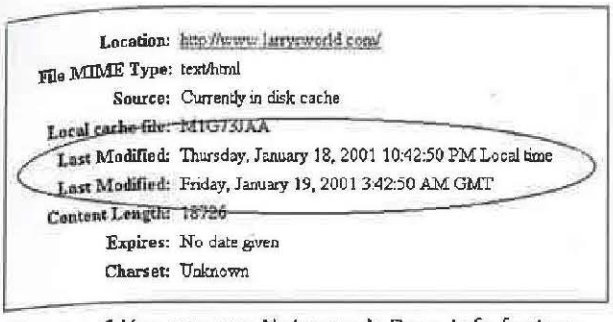


Figure 4.6 You can use Netscape's Page Info feature to find out when a Web page was last modified.

✔ Tips

- Many Web sites include an About Us or Company Info link that you can click on to get more information about the company behind the site.
- Netscape's Page Info feature can sometimes help in determining the currency of a specific Web page. Click on View and then on Page Info. The resulting screen (**Figure 4.6**) will often include the date the Web page was last modified. (Microsoft Internet Explorer doesn't currently offer a Page Info feature.)

7. Know when to look elsewhere.

Don't assume that the Internet contains the sum total of human knowledge. The Net will always surprise you, both with the information that it *does* contain and with its lack of information on some specific topic.

Part of being a good online searcher is knowing when to stop. The information you want may or may not be available online. And if it *is* online, it may be buried so deep that it's not worth the time and trouble to locate it.

Your efforts may be better spent getting the information, fact, figure, or whatever you need using conventional printed reference works: almanacs, dictionaries, encyclopedias, and so forth. Start with the reference section at your local library, or ask the reference librarian for help.

Avoiding Search Rage

Based on the results of a survey reported in a recent issue of Danny Sullivan's "Search Engine Report" newsletter, it takes about 12 minutes for the average person to start feeling anger and frustration at not finding what they are looking for on the Net. Consequently, Sullivan recommends that you follow the "10-minute rule."

If you haven't found what you're looking for in 10 minutes, says Sullivan, it's time to try more traditional alternatives—like contacting your local librarian, or picking up the telephone and calling directory assistance to get the phone number of a company that might be able to help you.

✔ Tips

- For really tough search assignments, your best bet may be to find a recognized expert on the subject. Start by consulting the "Sources and Experts" list compiled by news researcher Kitty Bennett of the *St. Petersburg Times*. It's available at `www.ibiblio.org/slanews/internet/experts.html`.
- You might also consider hiring a professional searcher (or *information broker*, as they are known in the trade). For recommendations, contact the Association of Independent Information Professionals. Their Web site at `www.aiip.org` (**Figure 4.7**) includes information about the organization's member-referral program.
- To learn more about becoming a professional searcher yourself, check your local library or favorite out-of-print book source for *The Information Broker's Handbook: Third Edition* by Sue Rugge and Alfred Glossbrenner. Although it's no longer in print, this book is still considered by information professionals to be the definitive work on the subject.

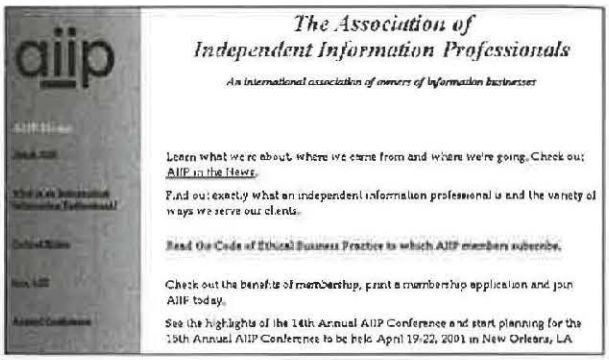


Figure 4.7 To hire a professional searcher, visit the AIIP Web site (`www.aiip.org`) and look for information about their member referral program.



Figure 4.8 This is the MSN home page, the place that Microsoft hopes Internet Explorer users will begin all their online sessions.



Figure 4.9 Netscape's home page includes a drop-down menu that lets you choose the search engine you use for your queries. The default is Netscape Search, but we've chosen Google instead.



Figure 4.10 Once you've customized your Web browser with your preferred home-page location, clicking on the Home button will take you there instantly.

Customizing Your Web Browser

When you install Microsoft Internet Explorer (IE), it will be set up so that whenever you go online, you'll be taken automatically to the Microsoft Network (MSN) home page at `www.msn.com` (**Figure 4.8**). Similarly, when you install Netscape Navigator, it assumes that you want to begin all of your online sessions at the Netscape home page, located on the Web at `home.netscape.com` (**Figure 4.9**).

We prefer, instead, to have our Web browser take us directly to our favorite search engine. After all, when we go online, it's usually to search for something. And even though both MSN and Netscape offer a selection of search tools on their home pages, neither site offers the range of features that you'll have access to if you go directly to one of the Big Six search engines.

That being the case, we've customized our Web browser so that the search engine we use most often comes up automatically when we sign on. We can also get to it quickly at any time by clicking on our browser's Home button (**Figure 4.10**). The other Big Six search engines are all included in our IE Favorites and Netscape Bookmarks lists, so that we can easily consult them when we need a "second opinion."

If you'd like to customize your Web browser so that it's optimized for searching, here are the steps to follow.

To make your favorite search engine appear automatically:

- **Internet Explorer 5 (Windows Users).** Click on Tools and then choose Internet Options. In the home page Address box, type the complete Web address (including http://) of your favorite search engine. Then click on OK.

  Note: If you are running an older version of IE for Windows, click on View (instead of Tools) to change your default home page.

- **Internet Explorer 5 (Macintosh Users).** Click on Edit and then choose Preferences. Then click on Browser Display in the list on the left side of the screen. In the home page Address box, type the complete Web address (including http://) of your favorite search engine and click on OK.

  Note: If you're running IE 4.5 for Mac, the link to choose from the list on the Preferences page is Home/Search (instead of Browser Display).

- **Netscape 4 and 6 (Windows and Mac Users).** Click on Edit and then choose Preferences. On the Navigator page, make sure the Home Page radio button is selected. In the Location box, type the complete Web address (including http://) of your favorite search engine. Then click on OK.

Complete Web Addresses for Big Six Search Engines

- AltaVista
  http://www.altavista.com
- Google
  http://www.google.com
- HotBot
  http://www.hotbot.com
- Lycos
  http://www.lycos.com
- Northern Light
  http://www.northernlight.com
- Yahoo
  http://www.yahoo.com

To add other search engines to your Favorites or Bookmarks:

- With Microsoft Internet Explorer, go to the search engine site, click on Favorites, and then choose the option for adding a page to your Favorites.

  Or use the keyboard shortcut Ctrl D if you're a Windows user, ⌘ D on the Macintosh.

- With Netscape Navigator, go to the search engine site, click on Bookmarks, and then choose Add Bookmark (if you're using Netscape 4) or Add Current Page (Netscape 6).

  Or use the keyboard shortcut Ctrl D (Windows) or ⌘ D (Macintosh).

✔ Tip

- Once you've set your browser's default home page location as described here, that's the site that will greet you whenever you sign on. It's also the one you'll be taken to when you click on your browser's Home button or select the menu option that tells the browser you want to "Go Home."

Keyboard Shortcuts and Other Timesavers

Customizing your Web browser so that it starts up automatically with your favorite search engine is one way to save time and get the ball rolling faster when you need to search for information on the Net. Now let's take at look at some other techniques for making your online sessions easier and more productive.

Keyboard shortcuts

Searching for menu commands can be a nuisance with any application, Web browsers included. We find that it's often faster and easier to perform common tasks with a couple of keystrokes. These keyboard shortcuts (**Table 4.1**) for Windows and Macintosh systems work with both Microsoft Internet Explorer and Netscape Navigator.

Making short work of Web addresses

We've mentioned this shortcut before but it bears repeating here: With the latest versions of Microsoft Internet Explorer (IE) and Netscape Navigator, you don't have to type `http://` in your browser's Address or Location box to get to a Web site. You can even leave out www and com to access your favorite search engine—or any site that begins and ends with those letters.

For example, to go to the Google Web site (`http://www.google.com`), just type `google` in your browser's Address or Location box (**Figure 4.11**) and press Enter (Ctrl Enter if you're using IE for Windows).



Figure 4.11 With Internet Explorer and Netscape Navigator, you don't have to type `http://www` or com to get to Google's Web site. Just `google` is enough.

Table 4.1

Keyboard Shortcuts

Task	Windows	Macintosh
Go to the next Web page.	Alt →	⌘]
Go to the previous Web page.	Alt ←	⌘ [
Add current Web page to Bookmarks/Favorites.	Ctrl D	⌘ D
Organize Bookmarks/Favorites.	Ctrl B	⌘ B
Copy highlighted text from a Web page.	Ctrl C	⌘ C
Paste highlighted text from a Web page.	Ctrl V	⌘ V
Find text on a Web page.	Ctrl F	⌘ F
Print current Web page.	Ctrl P	⌘ P
Open History folder.	Ctrl H	⌘ H
Open new Web page.	Ctrl L	⌘ L
Open new Browser window.	Ctrl N	⌘ N

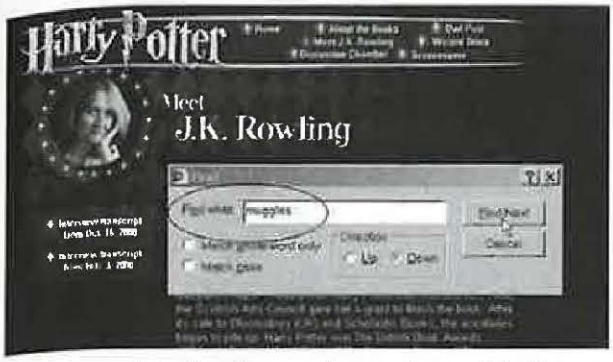


Figure 4.12 Locating the word *muggles* on this Harry Potter Web page is a snap (or a click!) with your browser's Find feature.



Figure 4.13 If your browser offers a Print Preview feature, it will be on the File menu. You can save time and paper by using Print Preview to selectively print just the pages you need from a multi-page Web page.

More Timesavers and Productivity Tools

In this section we focus on shortcuts and timesavers that are built into IE and Netscape. For recommendations on Web browser add-ons and other productivity-enhancing tools that you might want to add to your "Web searcher's toolkit," see Appendix D.

Searching for text in Web pages

Search engines often return pages that are loaded with text, and it's not immediately clear that the information you want is there. To avoid the time and trouble of scrolling and reading through page after page, let your Web browser's Find feature do the work for you.

With IE and older versions of Netscape (prior to Version 6), the Find feature is located on the Edit menu. With Netscape 6, it's on the Search menu. Alternatively, with both IE and Netscape, you can access the Find feature by using the keyboard shortcut Ctrl F (⌘ F on the Mac).

When the Find dialog box appears (**Figure 4.12**), simply enter the word or phrase you want to locate on the Web page and then click on your browser's `Find` (or `Find Next`) button.

Avoiding long print jobs

Imagine hitting your Print button and then discovering that the "Web page" you're printing is actually 28 pages long. And all you wanted was a couple of telephone and fax numbers that appear about two-thirds of the way through the document.

The secret to avoiding a problem like this is to activate your browser's Print Preview feature *before* hitting the `Print` button.

If you're using IE 5 (Windows or Mac) or Netscape 4 for Windows, you'll find the `Print Preview` link (**Figure 4.13**) on the File menu. (Unfortunately, at this writing Netscape 4 for Macintosh and Netscape 6 for Windows and Mac don't offer Print Preview.) Using Print Preview, find the page that contains the information you're after. (You'll probably have to "zoom in" to read the text and the page numbers that appear in the footer.) Then print just the page (or pages) you need.

Alternatively, with both IE and Netscape, you can highlight the text you're interested in and then copy and paste it to another application, like Word or Notepad, using the keyboard shortcuts Ctrl C to copy and Ctrl V to paste (⌘C and ⌘V on the Macintosh). Then print it from there.

✔ Tip

- From time to time, you'll encounter a Web page that won't display properly in the Print Preview window. The header and footer may be there, for example, but the main text—the portion you're most interested in printing—will be blank. When that happens, try sending the page to yourself as an e-mail message using the Send (or Send Page) link on the File menu.

Opening a second Web browser

Sometimes the Web is painfully slow and a complicated search can seem to take forever to finish. When that happens, consider opening a *second* browser window so that you can go about your business and check back in a couple of minutes.

With IE, the command sequence is `File` and then `New`, `Window` (**Figure 4.14**). With Netscape, click on `File` and then `New Navigator Window` (`New Navigator` on the Mac).

Alternatively, with both IE and Netscape, you can use the keyboard shortcut Ctrl N (⌘N on the Mac).

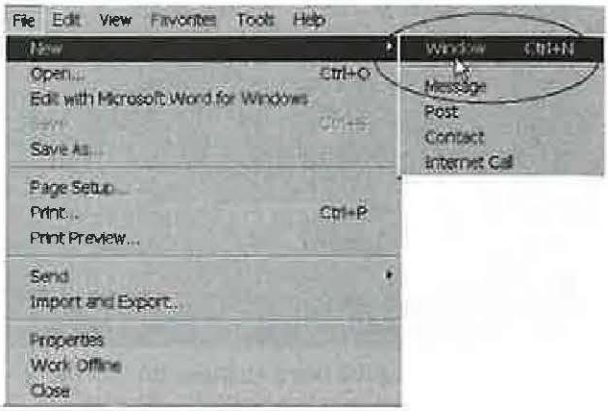


Figure 4.14 You can open a second Web browser window from the Internet Explorer File menu, as shown here. Netscape Navigator works much the same way.

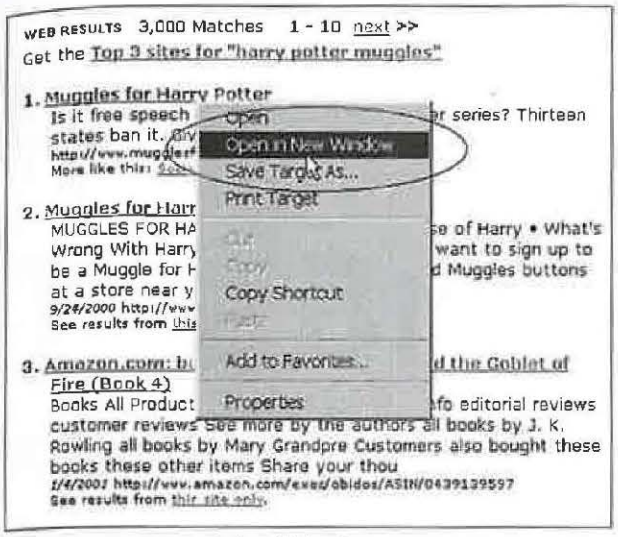


Figure 4.15 We've right-clicked on Muggles for Harry Potter to bring up the menu that includes an option for opening the link in a new browser window.

Viewing search results in a new window

How many times has this happened to you? You've initiated a search and produced a list of results that looks quite promising. You begin clicking on individual items, and the links on those pages, to see if they contain the information you're after. Eventually, you want to get back to your original results list. But it's buried so far down in the Back button's list that you can't easily find it. Or it may have even disappeared completely.

You can avoid this problem by keeping your search results on the screen and viewing individual results in a *new* window. Here's the technique: Instead of *clicking* on a link, do a *right-click* (Control-click on the Macintosh) and choose the menu option for opening it in a new window (**Figure 4.15**).

Dealing with broken links and "Page Not Found" messages

Some percentage of your searches are bound to lead you to Web pages that no longer exist. When that happens, try deleting everything after the last slash in the URL that appears in your browser's Address or Location box. Then press Enter. It may very well be that the Web site you're after still exists—it's just a particular file that's been eliminated or renamed.

PART 2

THE BIG SIX SEARCH ENGINES

Chapter 5:	AltaVista	53
Chapter 6:	Google	79
Chapter 7:	HotBot	97
Chapter 8:	Lycos	117
Chapter 9:	Northern Light	139
Chapter 10:	Yahoo!	157

The Big Six Search Engines

Once you've got a basic understanding of search engines and searching (either through direct experience or by reading Part 1), you'll be ready for these chapters on six of the very best search engines:

- **AltaVista**
- **Google**
- **HotBot**
- **Lycos**
- **Northern Light**
- **Yahoo!**

As you can see, the search engines are presented in alphabetical order. Feel free to skip around, perhaps starting with a search engine you've used before or heard the most about.

And remember, it's not necessary to master all six of these search engines. Use the information presented here to learn about the particular strengths of each one. Work through the step-by-step examples, and try some sample searches of your own with the book opened to the appropriate Quick Reference guide.

Once you've identified the search engine you like best, consider making it the default starting location for your Web browser, as suggested in Chapter 4. Add the others to your browser's Bookmarks or Favorites list so that you can get to them quickly when you need to consult a second (or third) search engine.

5

AltaVista



Good to Know

- AltaVista is a favorite of professional searchers because of its large database, powerful search features, and search customization options.
- It's one of the few search engines to offer full Boolean and case-sensitive searching, along with a variety of field-search and language options to help you target your searches.
- Developed in 1995 at Digital Equipment Corporation in Palo Alto, California, AltaVista has been awarded more search-related patents than any other company in the world.
- AltaVista is majority-owned by CMGI, an Internet development and operating company. But Compaq (which had acquired the search engine with its 1998 purchase of Digital) holds a small stake as well.

AltaVista is one of the most powerful and comprehensive search engines available today. Launched in 1995 as a pioneering effort to index the full text of the World Wide Web, it has evolved over the years from a pure search engine to a portal service offering the usual assortment of news headlines, shopping, free e-mail, and advertising. But AltaVista's primary emphasis is still very much on *searching*.

If you've never used a search engine before, you'll be pleased to discover how quickly—and precisely—you can find information with AltaVista. All it takes is a little practice and the mastery of a few simple techniques that we'll cover in this chapter.

Experienced searchers will appreciate the fact that AltaVista offers full Boolean search capabilities, case-sensitive searching, and a range of field-search options that is second to none. AltaVista is also one of the few search sites to provide a customization feature that goes beyond mere cosmetics. With AltaVista, you can tailor the search form to limit your queries to pages written in a particular language. And you can specify the amount of detail presented in your search results, as well as the number of items presented on each page.

To take full advantage of AltaVista's power and sophistication, you'll need to spend some time familiarizing yourself with its various search forms. (The rules for entering a query differ depending on which form you're using.) Once you've learned your way around, you may well decide to join the legions of professional searchers who've made AltaVista their search engine of choice.

✔ Tips

- The AltaVista database is created by a spider program known as Scooter (**Figure 5.1**). Scooter constantly roams the World Wide Web, collecting some 10 million Web pages each day and delivering them to AltaVista, where special software indexes every word. Scooter's work is supplemented by information provided to AltaVista by Web site creators.
- If you'd like to know more about the AltaVista spider, database, and indexing methods, click on `Submit a Site` on the AltaVista home page. Then click on `How AltaVista Works`.
- Although the AltaVista spider software is apparently quite good at finding new Web pages, it falls short when it comes to weeding out dead ones. In the most recent "Dead Links Report" from Search Engine Showdown, AltaVista ranked worst among the major search engines.

  For current rankings and information on how the dead links survey is conducted, go to `www.searchengineshowdown.com`.
- As you explore the Web, you may come across sites that offer search capabilities "powered by AltaVista" (**Figure 5.2**). Many of the search skills and strategies presented in this chapter—particularly the ones summarized in the Quick Reference tables—are applicable at these sites as well.



Figure 5.1 The AltaVista spider, affectionately known as Scooter and represented by this smiling image, is actually a very sophisticated software program that constantly scours the Web for new and updated pages.



Figure 5.2 Web sites that display this logo have a built-in search capability provided by AltaVista, so you can apply much of what you learn in this chapter to those sites as well.

Contact Information

AltaVista Company
Palo Alto, California
650-320-7700 (voice)
650-320-7720 (fax)
www.altavista.com

The AltaVista Home Page

AltaVista is constantly tinkering with its home page (`www.altavista.com`), so we can't guarantee that it will look exactly like the one shown in **Figure 5.3**. Features come and go, especially those not directly related to searching.

What we'll focus on here are the most important search-related features on the AltaVista home page—the ones we think you'll find most useful for day-to-day searching. As you'll soon discover, AltaVista offers a range of options for finding information on the Web. So whether you're a complete novice, a seasoned researcher, or somewhere in between, chances are you'll find that AltaVista has you covered.

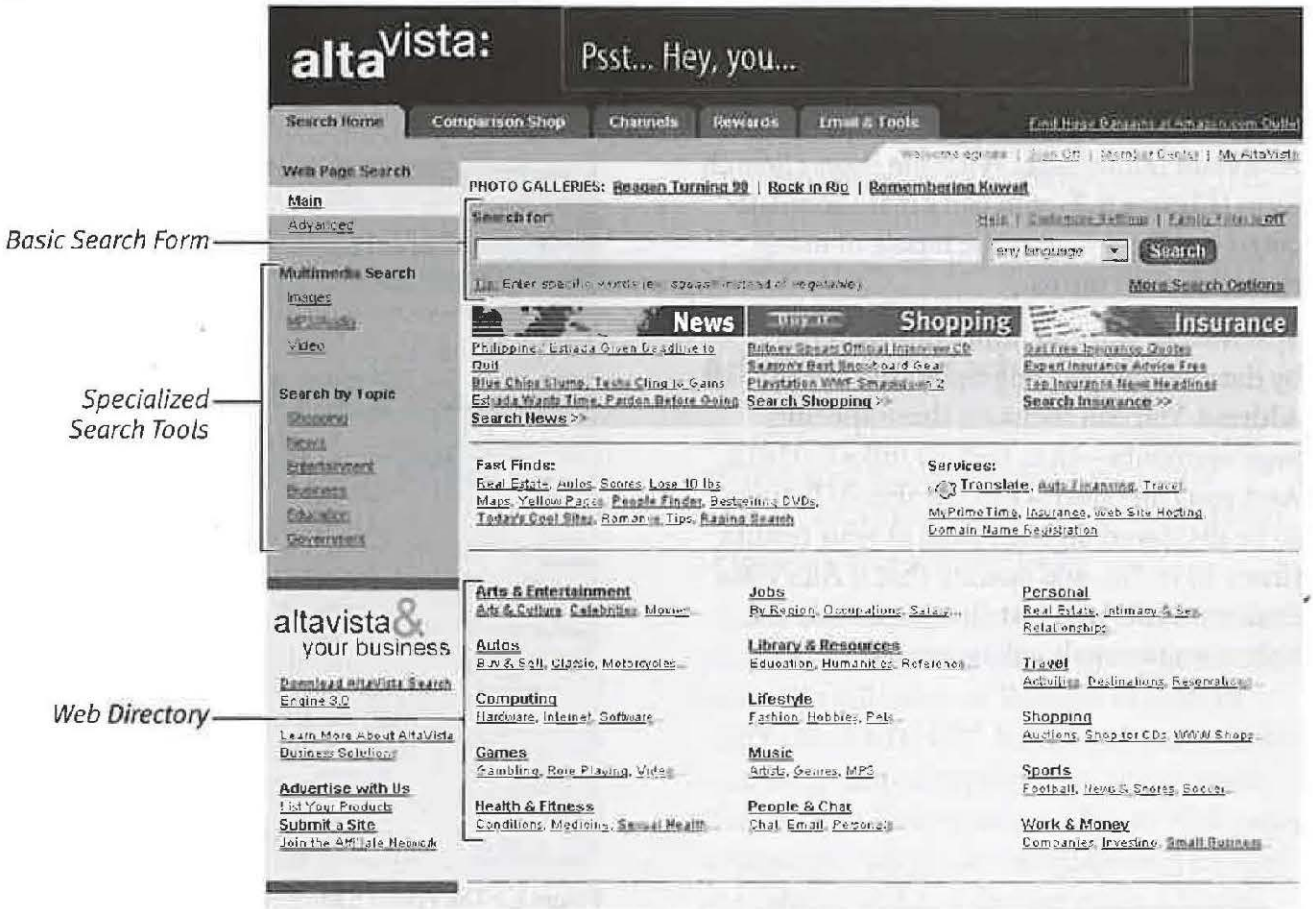


Figure 5.3 The AltaVista home page is crammed with features. From a searcher's perspective, the most important are the Basic Search form, the Web directory, and the specialized search tools accessible via links on the left side of the page.

Basic Search form

Many people find that AltaVista's Basic Search form (**Figure 5.4**) is the only one they ever need. It works just fine for most occasions, and even allows you to limit your searches to pages presented in a particular language. Just use the drop-down menu (**Figure 5.5**) to make your selection.

You can search for a phrase by enclosing the words in quotes, and use plus and minus signs to include (+) or exclude (-) words. You can also zero in on a specific part of a Web page by incorporating one of AltaVista's many field-search terms in your query.

We'll cover field searching and other aspects of using the Basic Search form in more detail later in this chapter.

Power Search

If you'd like a bit of guidance as you create your queries, click on More Search Options (**Figure 5.6**) near the Search button on the AltaVista home page. With the Power Search form (**Figure 5.7**), you can create sophisticated queries without the hassle of using special search terms.

Specifically, you can limit your searches by date, language, geography, and Web address. You can focus on three specific page elements—title, text, or links to URLs. And you can control the number of listings to be displayed on each page of your results (from 10 to 50), and specify that if AltaVista finds multiple pages at the same Web site, only a single result will be presented.



Figure 5.4 You will probably do most of your AltaVista searches using the Basic Search form.



Figure 5.5 Choose English from the drop-down Languages menu if you want to limit your Web search to English-language pages.



Figure 5.6 Clicking on this link on the AltaVista home page takes you to the Power Search form.

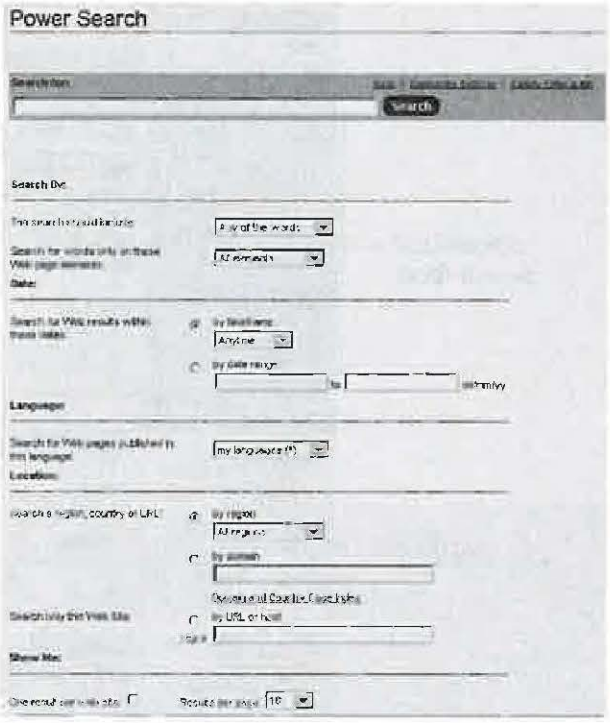


Figure 5.7 The Power Search form walks you through the process of building a complex query. Just make your selections from the drop-down menus and radio-button options.

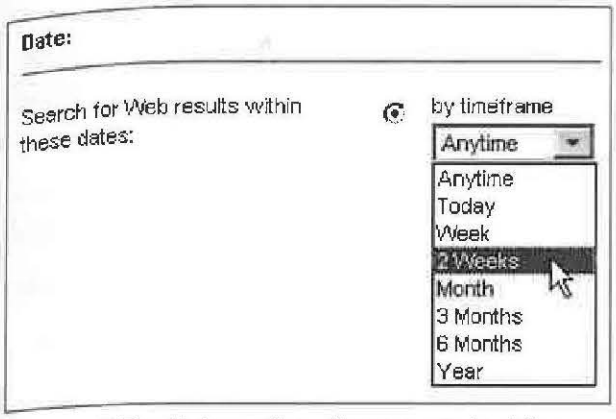


Figure 5.8 The Date section allows you to look for Web pages created or modified in the last two weeks, as shown here. You can also specify a range of dates.



Figure 5.9 This link on the AltaVista home page takes you to the Advanced Search form, favored by research professionals.

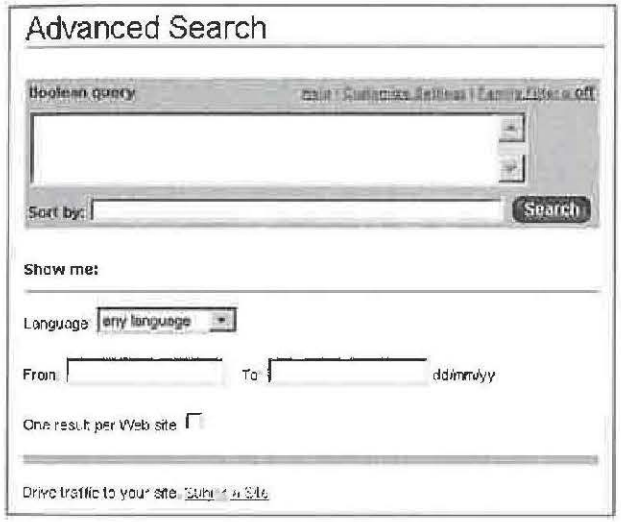


Figure 5.10 The Advanced Search form is the one to use if you prefer using Boolean expressions to search the Web. It also gives you more control over your search results than either Basic or Power Search.

Best of all, you're prompted through the entire process. All you have to do is make your selections by clicking on menu options and radio buttons (**Figure 5.8**).

✔ Tip

- If you like Power Search, try these specialty search options listed on the left side of the AltaVista home page. They apply the same menu/radio-button approach to finding specific types of information:

  Multimedia Search. For locating images, MP3/audio, and video files on the Web.

  Shopping. For comparison shopping at over 700 stores on the Web.

  News. For finding news articles that have appeared within the last two weeks.

  Entertainment. For locating entertainment news, audio and video clips, image files, radio broadcasts, and Web sites offering entertainment content.

  Business. For searching AltaVista's database of 1.8 million company fact sheets by name, ticker symbol, or domain name.

  Education. For limiting your search to Web sites with edu domain names.

  Government. For limiting your search to sites with gov domain names.

Advanced Search

For experienced searchers and others who prefer using Boolean operators in their queries, AltaVista's Advanced Search option (**Figure 5.9**) is the way to go.

On the Advanced Search form (**Figure 5.10**), you can create as complex a query as you like, using the full range of Boolean operators—AND, OR, AND NOT, NEAR, parentheses for enclosing nested expressions, and so forth. You can also limit your searches to Web pages written in a particular language and created or updated within a certain range of dates.

continues on next page

Finally, you can control the sort order for your results, giving more weight to certain words so that documents containing them appear higher on the list. And you can specify whether you want to see single vs. multiple results from each Web site.

For more on using this search form, see "Using Advanced Search" later in this chapter.

Web directory

When you're not exactly sure what you're looking for, or you simply want to get an idea of what the Web has to offer on a particular topic, try AltaVista's Web directory (**Figure 5.11**), provided by LookSmart.

To browse the directory, simply choose a main category and then "drill down" through the various subcategories.

To *search* rather than browse, click on a main category. (Any one will do.) Then type one or more keywords in the search form (**Figure 5.12**) and specify whether you want to search just that category or the entire directory. LookSmart automatically performs an AND search, looking for items that contain all the words in your query.

✔ Tip

- Phrase searching, plus and minus signs for including and excluding words, and Boolean operators are *not* allowed when searching the AltaVista/LookSmart directory. Consequently, when it comes to directory searches, Yahoo (`www.yahoo.com`) is a far better choice. For more on Yahoo, see Chapter 10.



Figure 5.11 AltaVista's Web directory is organized into 15 major categories and dozens of subcategories. If you don't see the topic you're looking for here, click on any link to access the directory's search form.



Figure 5.12 A search like this would find references to the words *song* and *lyrics* in the Music category of AltaVista's Web directory.

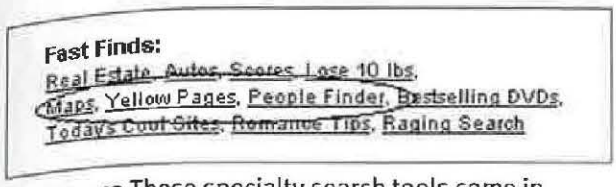


Figure 5.13 These specialty search tools come in handy for quick look-ups of maps and driving directions, business and residence addresses, and phone numbers.



Figure 5.14 Try Raging Search if you like AltaVista but prefer a "pure search engine" experience, without *any* portal features.

Other search-related tools

If you decide to make AltaVista your primary search site, you'll want to become familiar with its collection of specialty search tools, included in a section of the home page called Fast Finds (**Figure 5.13**). The ones you're likely to find most useful are `Maps`, `Yellow Pages`, and `People Finder`.

✔ Tips

- To save keystrokes, you can reach the AltaVista home page using the Web address `www.av.com`.
- For "pure AltaVista" searching without any of the usual portal features or advertisements, click on `Raging Search` on the AltaVista home page, or go directly to `www.ragingsearch.com`. Designed for hard-core search enthusiasts, Raging Search (**Figure 5.14**) uses the same database of Web pages as AltaVista and offers a full range of customization options.
- To make AltaVista your default home page, follow the instructions for "Customizing Your Web Browser" in Chapter 4.
- You can get back to the AltaVista home page at any time by clicking on `Search Home` or on the AltaVista logo at the top of most pages.

Searching with AltaVista

Now that you know your way around the AltaVista home page, let's do some searching. Despite its name, the Basic Search form is actually quite powerful. If you take full advantage of the tools that AltaVista makes available to you for use with Basic Search, you'll be amazed at how easily you can zero in on the information you're looking for on the Web.

In this section, we'll show you in step-by-step fashion how to enter search terms and special punctuation on the Basic Search form. For a summary of the material covered here, see the AltaVista Basic Search Quick Reference (**Table 5.1**).

Table 5.1

AltaVista Basic Search Quick Reference

For this type of search:	Do this:	Example:
Plain-English Question	Simply type the question in the search form text box. Use as many words as necessary.	What is the date of the Battle of Trafalgar?
Phrase Search	Type the phrase as a sequence of words surrounded by **double quotes.**	`"Battle of Trafalgar"`
AND Search (multiple words and phrases, each of which *must* be present)	Use a **plus sign (+)** in front of each word or phrase that must appear in the results.	`+London + "art museum"`
OR Search (multiple words and phrases, any one of which may be present)	Type words or phrases separated by spaces, without any special notation.	`Stratford Shakespeare`
NOT Search (to exclude a word or phrase)	Use a **minus sign (–)** in front of a word or phrase you want to exclude from the results.	`+python –monty`
Case-Sensitive Search	Use lowercase to find any *combination* of uppercase and lowercase. Use capital letters (initial caps or a combination of uppercase and lowercase) to force an exact match of your search term. The first example would match *Bath* (but not *bath* or *BATH*). The second example would match *BATH* only.	`Bath` `BATH`
Date Search	Not available on the Basic Search form. Use Advanced Search or Power Search.	
Field Search	Type field-search keyword in lowercase, followed by a colon and your search word or phrase. (See **Table 5.2** for a complete list of field-search keywords.)	`title:"Victoria and Albert Museum"` `host:cambridge.edu` `domain:com`
Nested Search	Not available on the Basic Search form. Use Advanced Search or Power Search.	
Proximity Search	Not available on the Basic Search form. Use Advanced Search or Power Search.	
Wildcard Search	Use an **asterisk (*)** at the end of or within a word, along with at least three letters at the beginning of the search term.	`Brit*` `col*r`

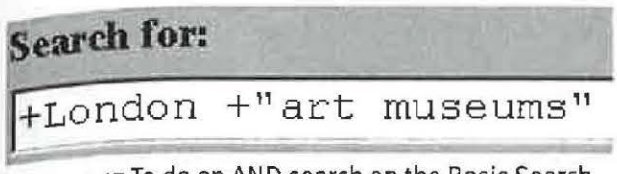


Figure 5.15 To do an AND search on the Basic Search form, put a plus sign (+) in front of each term.



Figure 5.16 A minus sign (-) in front of a search term tells AltaVista to exclude Web pages containing that word or phrase.



Figure 5.17 To do an OR search, type the words (or phrases) separated by a space.

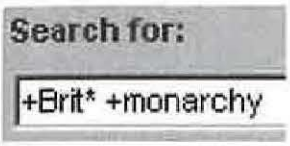


Figure 5.18 A search like this would locate references to both *British monarchy* and *monarchy in Great Britain*.

To enter a Basic Search query:

1. Go to `www.altavista.com` and type several unique keywords, a phrase in quotes, or a combination of words and phrases in the Basic Search form. Use a plus sign (+) in front of each word or phrase to tell AltaVista that they must *all* appear in your search results (**Figure 5.15**).
2. Use a minus sign (-) in front of any word or phrase you want to *exclude* from the search results. To find London museums *other than* the Victoria and Albert, use a search like the one shown in **Figure 5.16**.
3. To look for Web pages with references to *any* (but not necessarily *all*) of the terms in your query, type the words or phrases with a space between them. A search like the one in **Figure 5.17**, for example, would find all pages that have references to either *Trafalgar* or *Admiral Nelson*.
4. Use an asterisk (*) at the end of (or within) a word to do a wildcard search. Be sure to use at least three letters at the beginning of the search term or AltaVista will ignore the request. Using `Brit*` as a search term will find references to *Britain*, *British*, and *Britannia*, among others.

   For best results, combine a wildcard search term with at least one other unique keyword or phrase (**Figure 5.18**).
5. When you've finished entering your query, click on the `Search` button or press the Enter key to submit it.

continues on next page

✔ Tips

- You'll typically get much better results with AltaVista if you search for a phrase, or a combination of words and phrases, rather than a single word. If you're not sure what words and phrases to use, try entering your search as a question in plain English: `What is the line of succession for the British monarchy?`.
- AltaVista uses an automated process to detect commonly used phrases in your queries, even if you haven't enclosed the words in quotation marks. Consequently, you'll sometimes get the same results, with or without quotes (**Figure 5.19**). Still, it's a good idea to get into the habit of always enclosing phrases in quotation marks on the Basic Search form. After all, there's no way of knowing for sure exactly what AltaVista considers a "common phrase."

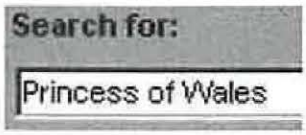


Figure 5.19
Common phrases like this are included in AltaVista's phrase dictionary, so you'll get the same results with or without quotation marks.

To make your search case-sensitive:

- Use initial caps or all caps as appropriate to search for a proper name or an acronym: `Bath` to search for the town of that name; `BBC` to find sites featuring the British Broadcasting Company.

✔ Tips

- Keep in mind that if you use any capital letters in your search request, AltaVista will perform a case-sensitive search. If, on the other hand, you enter your search terms in lowercase, you'll find all references to the term, regardless of case. Searching for bath, for example, will find references to bath, Bath, and BATH.
- When in doubt, enter your AltaVista search terms in lowercase. (If you're a two-finger typist who routinely turns on Caps Lock and pecks away, you're not likely to have much success with AltaVista.)
- Case-sensitive searching comes in handy when you're looking for a company name that also happens to be a common word or phrase. Searching for `next` will return millions of hits. But a case-sensitive search for `NeXT` (or better yet `"NeXT Inc"`) will help you zero right in on sites dealing with the company founded by Steve Jobs.



Figure 5.20
If a particular search returns a lot of foreign-language pages, you can eliminate them by using the languages menu.

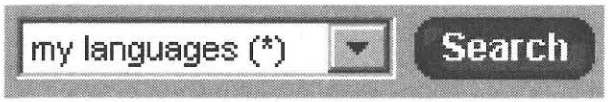


Figure 5.21 Once you've customized the Language Options, AltaVista will automatically search for pages in your chosen language(s).



Figure 5.22 For an instant translation to English, click on the `Translate` link in your search results.

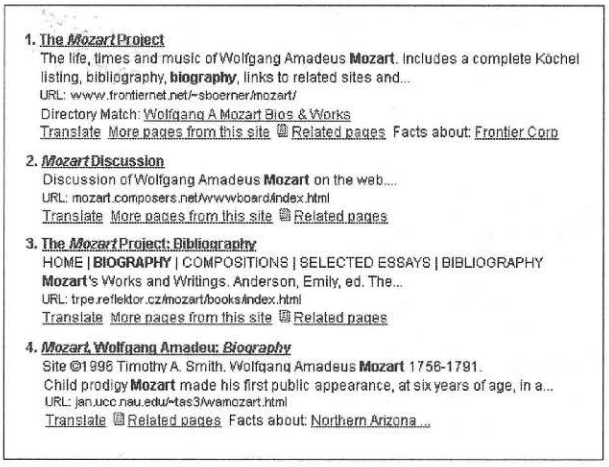


Figure 5.23 Here's an example of how AltaVista presents search results for a query that includes the terms *Mozart* and *biography*.

To limit your search to a particular language:

1. Use the Basic Search form's drop-down menu (**Figure 5.20**) to make a language selection. Choosing `English`, for example, will ensure that only Web pages written in English will be included in your search results.
2. To permanently change the default setting to `English` (or whatever languages you prefer), click on `Customize Settings` and then on `Language Options`. Make your language selections and then click the `Submit` button.
3. The default language setting for your searches will now be `my languages(*)` (**Figure 5.21**).

✔ Tip

- Another way of dealing with unfamiliar languages is to use AltaVista's translation tool, Babel Fish. If your results include a Web site in French, German, Italian, Portuguese, Russian, or Spanish, click on the `Translate` link (**Figure 5.22**) for an instant translation to English. It won't be perfect, but it could give you enough information to go on.

To view your search results:

1. Click on the `Search` button to the right of the search form text box and AltaVista will present your results, 10 to a page, with the best matches listed first (**Figure 5.23**). Your search terms will be highlighted in bold to make them easy to spot in the site descriptions. In Web page titles, which are bold to begin with, they'll be italicized.

continues on next page

2. Zero in on a typical entry (**Figure 5.24**) on the results page and you'll see that it includes the title of the document, a brief description provided by the person who created it (or, if a description has not been provided, the first couple of lines of text on the page), and the URL.

3. If your query matches multiple pages at a single Web site, AltaVista "clusters" the results and presents only one page from that Web site—a neat feature that allows you to get more variety in your results. To look at other pages from the site that match your query, click on the link labeled `More pages from this site` (**Figure 5.25**).

4. AltaVista also tells you exactly how many matches it found in its database (**Figure 5.26**). And if you scroll down to the bottom of the page, you'll find a word count for each of the keywords and phrases in your query (**Figure 5.27**), as well as the words (if any) that AltaVista ignored because they appear too frequently in its database.

   This information can sometimes help you refine a search or identify problems. For example, if AltaVista found no instances of a particular term, it may be because you misspelled it.

5. If your search results look promising but you don't find exactly what you're looking for in the first couple of pages of results, you may want to modify your query and try again. For cases like this, AltaVista gives you the option of searching within the first set of results (**Figure 5.28**) instead of the entire database.

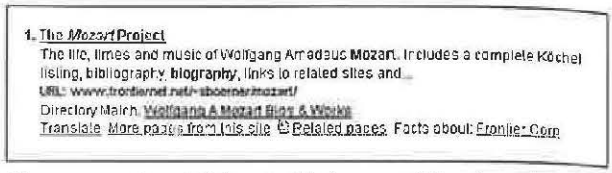


Figure 5.24 In addition to Web page title, description, and URL, this listing includes several links, including one for a match in the Web directory.

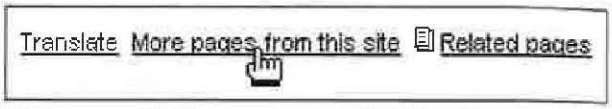


Figure 5.25 AltaVista presents a link like this whenever it finds your search terms on multiple pages at a single Web.

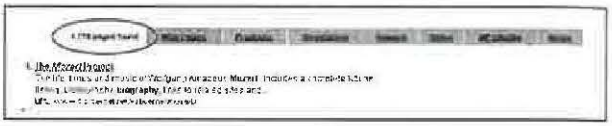


Figure 5.26 The first page of your search results will include information on the total number of Web pages found by AltaVista.



Figure 5.27 Near the bottom of your search results page, you'll find a word count for each of your search terms.



Figure 5.28 If you decide that your initial query was too broad, you might try adding another word or phrase and searching again *within* your original results.



Figure 5.29 Another possibility for a follow-up search is to click on one of these options. Choose MP3/Audio, for example, to find sound files that match your search terms.

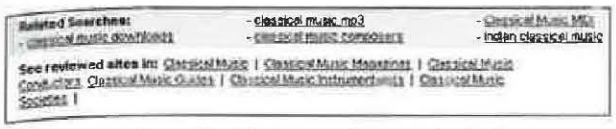


Figure 5.30 Your AltaVista results may include sections like these that can help clarify your search or locate sites that have been reviewed and approved for inclusion in the Web directory.



Figure 5.31 Don't get your hopes up when you find this section in your search results. It's often just an advertising gimmick.

6. With a single click, you can also direct your query to one of AltaVista's other search options (Products, Directories, Images, etc.) located above the search form on the search results page (**Figure 5.29**).

✔ Tips

- If your original search is quite general, your results may include a section labeled "Related Searches" to help you narrow the focus and home in on the information you're looking for. And if it matches a category in the AltaVista Web directory, you'll find them in a section labeled "See reviewed sites in" (**Figure 5.30**).
- You may also find a section of your search results labeled "AltaVista Recommends" (**Figure 5.31**). If so, don't be taken in by the fact that one of your search terms may be incorporated into the "recommendations." In most cases, following these links will simply lead you to generic promotional offerings by AltaVista advertisers.
- AltaVista allows you a great deal of control over the amount of detail included in your search results, as well as the number of items presented on each page. You can also turn bold highlighting of your search terms on or off. For more information, see "Customizing AltaVista" later in this chapter.
- For additional guidance in using the Basic Search form, click on the Help link above the Search button. AltaVista's help information is better than most—nicely organized and well written. You can even print out a "Cheat Sheet" with search examples and suggestions for getting better results from your AltaVista queries.
- Our only complaint about AltaVista's help information is that it doesn't always keep up with the changes being made at the site. From time to time, you'll encounter a reference to a feature that's no longer offered, which can be a bit disconcerting.

Improving Results with Field Search

It's not uncommon to enter a query on the AltaVista Basic Search form that returns hundreds, if not thousands, of Web pages—many of which have little or nothing to do with the information you're looking for.

When that happens, try adding one of AltaVista's *field-search* terms to your query. Field-search terms allow you to limit your search to specific parts (or *fields*) of Web pages.

You'll find a complete list of field-search terms that you can use on both the Basic and Advanced Search forms in the AltaVista Field Search Quick Reference (**Table 5.2**). The ones you are likely to find most useful are Title, Domain, Host, and Link.

✔ Tip

- For point-and-click field searching, use AltaVista's Power Search form. It gives you the ability to incorporate some of the more popular field-search terms into your queries by making selections from drop-down menus and radio-button options.

Table 5.2

AltaVista Field Search Quick Reference

Search Term	Description	Examples
anchor:	Searches for Web pages that contain the specified hyperlink.	`anchor:"free product samples"`
applet:	Searches for Java applets. If you don't know the name of the applet, try combining an applet wildcard search with some other search term.	`applet:beeper` `+applet:* +Java`
domain:	Searches Web addresses for a specific domain (**com, edu, gov, net, org**, etc.) or two-letter Internet country code. (See Appendix B for a complete list.)	`domain:edu` `domain:uk`
host:	Searches just the *host name* portion of Web addresses.	`host:beatlefest.com` `host:oxford.edu` `host:BBC`
image:	Searches Web pages for the filenames of images matching your search term.	`image:ringo.gif` `image:*.gif`
like:	Searches for Web pages similar (or related in some way) to the specified URL.	`like:www.beatle.net`
link:	Searches for hypertext links (URLs) embedded in a Web page.	`link:www.songlyrics.com`
text:	Searches for text in the body of the Web page.	`text:"Strawberry Fields"`
title:	Limits search to the part of the Web page that the author labeled as the title.	`title:"John Lennon"`
url:	Searches for text in complete Web addresses (URLs).	`url:beatles.html`

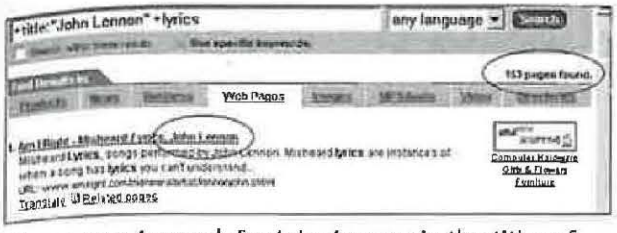


Figure 5.32 A search for *John Lennon* in the *titles* of Web pages reduced the number of hits from several thousand to 153.



Figure 5.33 A search like this would help you locate Beatles-related sites in the U.S. that are commercial in nature.

To search Web page titles:

- Add a Title field search to your query. Let's say your initial search for "+"John Lennon" +lyrics results in too many hits. Try restricting your query to Web pages that include the singer/songwriter's name in the title: +title:"John Lennon" +lyrics (**Figure 5.32**). In most cases, you'll find that your search returns a much more manageable list of results.

To limit your search by Internet domain:

- Include a Domain field search in your query. For example, to focus on Web pages of commercial establishments, include +domain:com in your search request (**Figure 5.33**). You can also limit your search to Web pages for a *specific* domain name by including the entire name with the Domain search term. For example, using +domain:beatlefest.com in your query would limit your search to the Beatlefest Web site.

  Here are the field-search terms for the most common Internet domains:

  domain:com (commercial establishments)

  domain:edu (educational institutions)

  domain:gov (government agencies)

  domain:mil (U.S. military)

  domain:net (networks like AT&T and Sprint)

  domain:org (non-profit organizations)

To limit your search by country:

- Do a Domain search for the country's two-letter country code. Adding `+domain:uk` to a Beatles search would limit the query to sites based in the United Kingdom (**Figure 5.34**).

✔ Tips

- You can also use a Domain search to eliminate foreign-language sites from your search results. If you can't read French and want to avoid French-language pages, for example, you could add `-domain:fr` to your query. (Note the use of the minus sign.).
- Remember, though, that AltaVista gives you the option of selecting your preferred language from a drop-down menu on the search form. You'll probably find that approach to be easier and more convenient than doing a Domain search.
- See Appendix B for a complete list of Internet domains and country codes.

To limit your search by host name:

1. Use a Host field search to find all the Web pages "hosted" on a particular computer. A search for `+host:harvard.edu`, for example, would find Web pages hosted by a computer at Harvard University.
2. To exclude all such pages, put a minus sign (-) in front of the field-search term. A search like the one in **Figure 5.35**, for example, would help you locate sites that mention Harvard University while avoiding Harvard's own Web site.

To search for embedded links:

- Use a Link field in your query. The search term `+link:www.songlyrics.com` would help you identify Web pages that include embedded links to the Song Lyrics Web site.

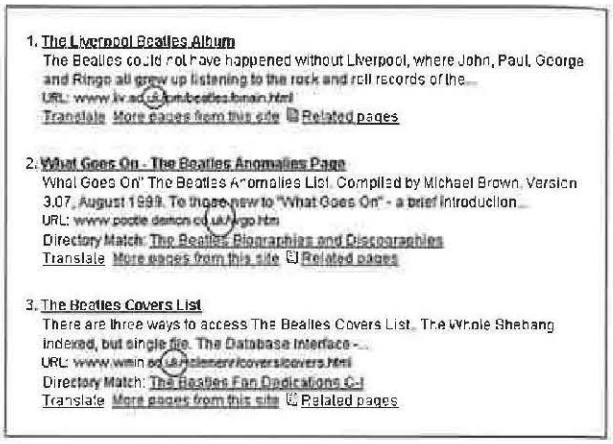


Figure 5.34 If you include `+domain:uk` in your query, your search will be limited to Web sites in the United Kingdom, as shown here.



Figure 5.35 To exclude Web pages from a particular host system, use a Host field search preceded by a minus sign (-).

✔ Tip

- Webmasters and Web site creators often use a Link search to find out who is sending traffic to their Web sites. It also provides a rough indication of a site's popularity—the more links, the more popular the site.

Using Advanced Search

Many AltaVista fans never venture beyond the search engine's main page. They master the Basic Search form and use it exclusively for most of their Internet searches. If they can't find what they're looking for with Basic Search, they just move on to another engine.

But power users and professional searchers will tell you that AltaVista's Advanced Search form (**Figure 5.36**) is worth taking the time to learn. In fact, some of what you've learned already about Basic Search applies to Advanced Search as well. The techniques for using wildcards, performing a case-sensitive search, and field searching are the same on both the Basic and Advanced Search forms.

continues on next page

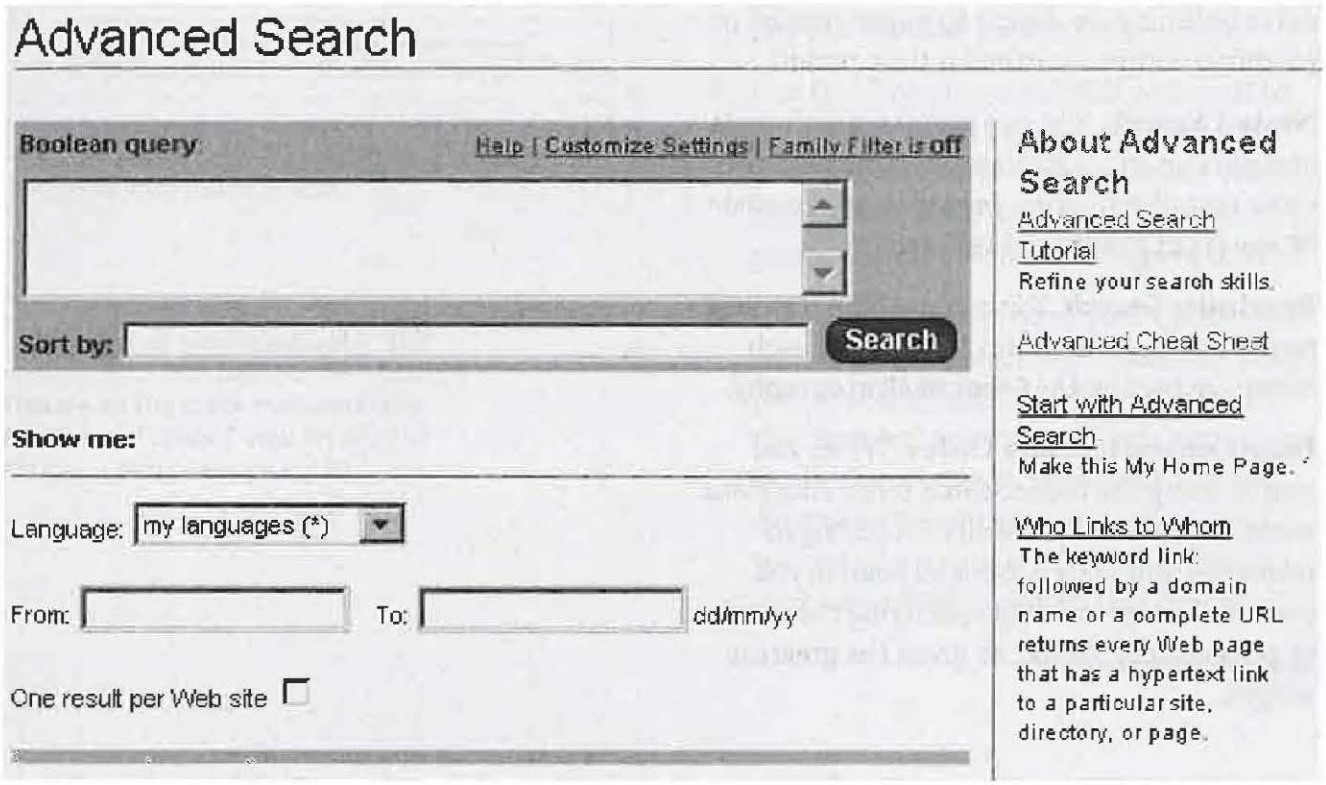


Figure 5.36 AltaVista's Advanced Search form offers a number of search options not offered on the Basic Search form, including use of Boolean operators and date searching.

The rules for specifying phrases and for combining words and phrases are different, however.

Specifying Phrases. On the Advanced Search form, it is *not* necessary to enclose phrases in quotation marks. Whenever two or more words appear together without a Boolean operator between them, AltaVista interprets the words as a phrase.

Combining Words and Phrases. Instead of using plus and minus signs to require and exclude keywords and phrases, you must use the traditional Boolean operator AND (to require a term) and the somewhat less traditional AND NOT (to exclude a term). To search for several words or phrases, any of which *may* be present in your search results, you must link the terms with the Boolean operator OR.

Advanced Search also offers several very useful features not available with Basic Search:

Date Search. You can specify a range of dates to limit your search to pages created or modified within a particular time period.

Nested Search. You can group search words and phrases in parentheses in order to form more complex queries: `(mystery AND (author OR novelist)) AND bestseller`.

Proximity Search. You can use NEAR to find terms that appear within 10 words of each other: `Catherine the Great NEAR biography`.

User-Controlled Sort Order. When you search using the Basic Search form, AltaVista sorts and ranks your results according to relevance. But with Advanced Search, *you* control the sort order by specifying the words or phrases that should be given the greatest weight.



Figure 5.37 Use this link on the AltaVista home page to access the Advanced Search form.

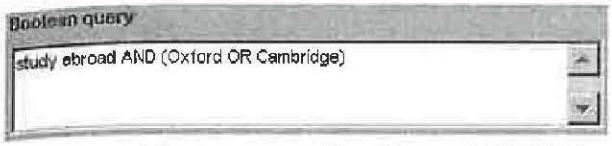


Figure 5.38 A Boolean query like this would find Web pages with references to the phrase *study abroad* (no quotes required) and either *Oxford* or *Cambridge*.



Figure 5.39 Words typed in the "Sort by" box are given greater weight when AltaVista organizes your search results. Pages containing those words will be listed first.

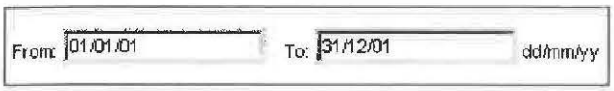


Figure 5.40 Specifying a range of dates like this would limit your search to Web pages created or modified during the 2001 calendar year.

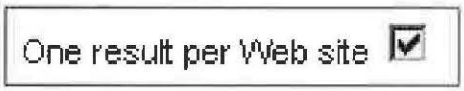


Figure 5.41 The check mark here tells AltaVista to "cluster" your results and show you only one page per Web site.

To search with Advanced Search:

1. Go to the AltaVista home page (`www.altavista.com`) and click on Advanced (**Figure 5.37**).
2. Type your search request in the area labeled "Boolean query" (**Figure 5.38**). Use AND, OR, and NEAR to combine words and phrases. To exclude a word, be sure to use AND NOT: `Oxford AND NOT shoes`. You can also use parentheses in your query to group words and phrases.
3. If there's any particular word or phrase that should be given greater weight in the presentation of search results, enter it in the "Sort by" area of the form (**Figure 5.39**). Boolean expressions are not allowed in this part of the search form.
4. To limit your search to documents created within a range of dates, enter that information in the form `DD/MM/YY`. Be sure to use a two-digit number for day, month, and year (**Figure 5.40**). And note that day (rather than month) comes first.
5. Use the drop-down menu if you want to limit your search by language, and click on the box labeled "One result per Web site" if you prefer not to see multiple pages from the same site (**Figure 5.41**).
6. Click on Search to enter your query. Your results will be presented in random order, unless you entered a word or phrase in the "Sort by" text box, as explained above. In all other respects, search results for Advanced Search are identical to what you get when you use the Basic Search form, described earlier in this chapter.

continues on next page

✔ Tips

- You can use uppercase or lowercase for Boolean operators on the Advanced Search form.
- Boolean operators are also allowed on the Power Search form, as long as you choose `Boolean` from the drop-down menu (**Figure 5.42**).
- To do a Boolean search for images, MP3/audio, or video files, you can use special versions of the Advanced Search form accessible by clicking on the appropriate multimedia search link (**Figure 5.43**) on the left side of the Advanced Search page.
- The AltaVista Advanced Search Quick Reference (**Table 5.3**) includes a summary of all the Advanced Search features and how to use them.
- For more tips and search examples, as well as a tutorial and "Cheat Sheet," click on `Help` on the Advanced Search form.

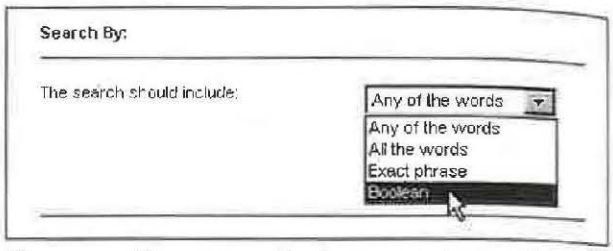


Figure 5.42 You can use Boolean operators on the Power Search form as long as you specify `Boolean` on the drop-down menu.



Figure 5.43 Use these links on the Advanced Search page to access special versions of the Advanced Search form tailored for finding image, MP3/audio, and video files.

Table 5.3

AltaVista Advanced Search Quick Reference

FOR THIS TYPE OF SEARCH:	DO THIS:	EXAMPLES:
Plain-English Question	Not recommended on Advanced Search form. Use Basic Search instead.	
Phrase Search	Just type the phrase in the search box (*without* quotation marks). AltaVista interprets as a phrase any words that appear together without a search operator between them.	Tower of London
AND Search (multiple words and phrases, each of which *must* be present)	Use AND between words or phrases to specify that both must be present in the results.	Oxford AND Cambridge
OR Search (multiple words and phrases, any one of which may be present)	Use OR between words or phrases to specify that you want to find references to either or both terms.	Oxford OR Cambridge
NOT Search (to exclude a word or phrase)	Use AND NOT in front of the word or phrase you want to exclude from the query.	Oxford AND NOT Cambridge
Case-Sensitive Search	Use lowercase to find any *combination* of uppercase and lowercase. Use capital letters (initial caps or a combination of uppercase and lowercase) to force an *exact match* of your search term, as shown in the example.	Round Table
Date Search	Type the range of dates you want to search in the From and To boxes, using the form DD/MM/YY. The example would search for dates between January 1, 2001, and December 3, 2001 (not March 12, 2001).	From: 01/01/01 To: 03/12/01
Field Search	Type field-search keyword in lowercase, immediately followed by a colon and your search word or phrase. (See **Table 5.2** for a complete list of field-search keywords.)	title:Castle Howard domain:com host:cambridge.edu
Nested Search	Use **parentheses** to group search expressions into more complex queries. The example would find *Queen Mother* as well as *Queen Mum*.	Queen (Mother OR Mum)
Proximity Search	Use NEAR to find words or phrases that appear within 10 words of each other. The example would find *bed and breakfast* as well as *bed & breakfast* and *breakfast in bed*.	bed NEAR breakfast
Wildcard Search	Use an **asterisk (*)** at the end of or within a word with at least three letters at the beginning of the search term.	bicycl* col*r

Customizing AltaVista

If you use AltaVista a lot, you'll want to take advantage of its customization features. As we mentioned earlier in this chapter, AltaVista gives you more control than most search engines over how you search and how your results are displayed. You can also create your own personalized version of the AltaVista home page.

Here's a brief summary of the AltaVista customization features:

Results Options. This feature gives you complete control over the amount of detail displayed in your search results. It also allows you to specify the number of search results presented on each page, and whether your search terms should be presented in bold type so that they are easy to locate.

Language Options. You can change AltaVista's default language setting from any language to English, or whatever languages you can read and understand. Once you've identified your language preferences, AltaVista will automatically screen out Web pages written in other languages.

Family Filter. Allows you to set preferences for screening out objectionable material from your Web searches.

My AltaVista. Gives you the ability to pick and choose the "portal features" you find most useful, and decide how you want them arranged on the AltaVista home page.



Figure 5.44 You'll find the Customize Settings link on all the major AltaVista search forms: Basic, Advanced, and Power Search.

To set Results Options:

1. Click on Customize Settings (**Figure 5.44**) on the AltaVista Basic Search form, or on the Advanced or Power Search form. That will take you to the Results Options page.

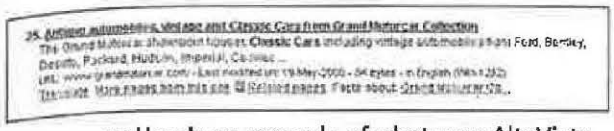


Figure 5.45 Here's an example of what your AltaVista results will look like if you choose to display all available options.

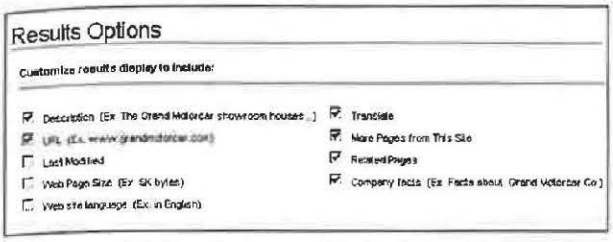


Figure 5.46 If there's any particular item you would prefer not to see in your results, click on the box next to it to remove the check mark.

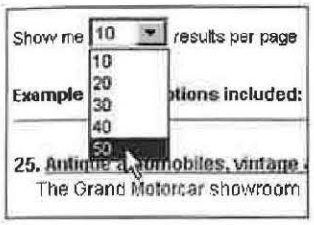


Figure 5.47 Increasing the number of results displayed on each page can be a time saver if you typically like to review 20 or more listings for your searches.

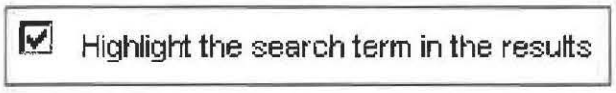


Figure 5.48 Remove this check mark to turn off highlighting of your search terms.

2. Scroll down the page and take a look at the example of how AltaVista displays search results (**Figure 5.45**). If there are any particular elements that you would like to remove, just click on the box next to the item to eliminate the check mark (**Figure 5.46**).

   If you're not sure whether or not to exclude an item, leave it checked until you've had more experience with AltaVista. At a minimum, you'll certainly want to keep `Description` and URL. We also find that `Last Modified` and `More Pages from This Site` come in handy from time to time.

   For an explanation of each item, scroll down the page and click on the `Help` link.

3. Once you've decided on the amount of detail to be included in your search results, use the drop-down menu (**Figure 5.47**) to choose the number of listings you want AltaVista to display on each page. The default setting is 10, but you can increase it to as many as 50.

   The advantage to doing so is that when you're reviewing your search results, you can simply scroll down the page, instead of having to load a new page for every 10 results.

4. Next, decide whether you want your search terms presented in bold in your search results. If so, make sure that there is a check mark in the `Highlight` box (**Figure 5.48**).

5. When you are finished setting your Results Options, click on the `Submit` button and AltaVista will begin using them immediately. To change them at any time, just repeat the steps presented here.

To set Language Options:

1. Click on Customize Settings on the AltaVista Basic Search form and then on Language Options (**Figure 5.49**).

2. Make your language selections by clicking on the box next to any language (**Figure 5.50**). Most Web pages are presented in English, but if you want to ensure that your results include *only* English-language pages, click on the box for English. Once you've done that (and made the setting permanent as described below), you won't have to bother with the drop-down languages menu on the AltaVista Basic, Advanced, and Power Search forms.

3. If you've selected languages that use non-Roman alphabets (Chinese, Czech, Greek, Hebrew, Japanese, Korean, Polish, or Russian), be sure to complete the sections on encoding options.

4. Finally, click on the Submit button. From this point forward, the default setting on the AltaVista search form's drop-down languages menu will be my languages(*) (**Figure 5.51**).

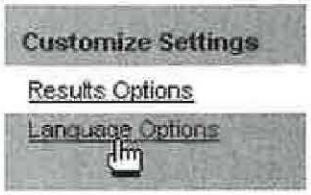


Figure 5.49 Click here to access the Language Options page.



Figure 5.50 The Language Options page allows you to limit your searches to English or more than two dozen other languages.

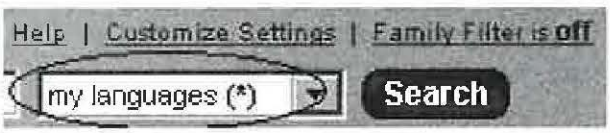


Figure 5.51 After setting Language Options, your searches will automatically default to your chosen language(s).

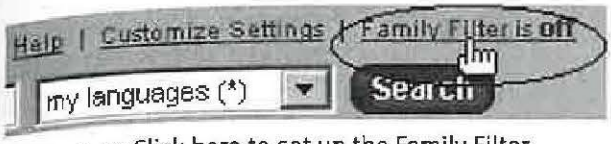


Figure 5.52 Click here to set up the Family Filter.

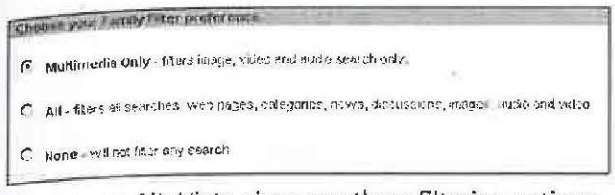


Figure 5.53 AltaVista gives you three filtering options, shown here.

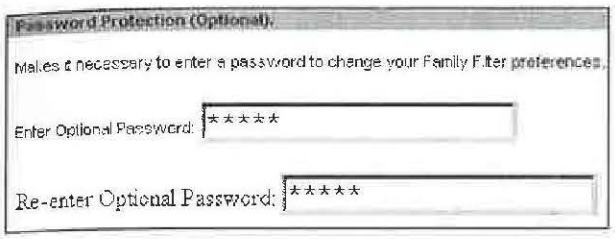


Figure 5.54 Password protection is an optional feature with the Family Filter.

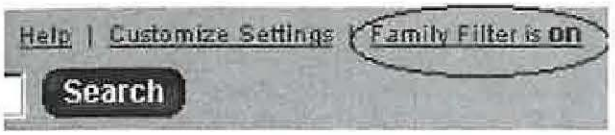


Figure 5.55 This link on the AltaVista search form tells you that the Family Filter is set to filter *all* Web content.

To set the Family Filter:

1. Click on `Family Filter is off` (**Figure 5.52**) on the AltaVista home page to access the Family Filter Setup Page.
2. Choose your filter preferences (**Figure 5.53**) by clicking the radio button to indicate whether you want to screen `Multimedia Only` (the default) or `All` Web content. To eliminate filtering completely, click on the radio button for None.
3. To activate password protection—optional, but probably a good idea if your objective is to prevent young family members from accessing inappropriate material—choose a password and enter it twice in the boxes provided (**Figure 5.54**).
4. Then click on the Done button to submit your Family Filter preferences.
5. If you have chosen to filter all Web content, when you return to the AltaVista search form, the Family Filter will be on (**Figure 5.55**). If you've chosen to filter multimedia content only, the filter will be activated only when you use one of the Multimedia search forms.

✔ Tips

- AltaVista's Family Filter works only for Web pages written in English.
- The Family Filter uses *cookie* technology to save your filter settings. If you have set your Web browser to reject cookies, you'll have to reset the Family Filter each time you visit AltaVista. (For more on this topic, including how to turn cookies back on, return to the Family Filter Setup Page and click on `Frequently Asked Questions`.)

To create your own AltaVista home page:

1. Click on My AltaVista (**Figure 5.56**) on the AltaVista home page.
2. Next click on Sign Up. You'll be asked to choose a member ID and password and to provide your name, address, and other information, most of which is optional.
3. Once you've completed the form, click on Next and you'll find yourself on the Customize My AltaVista page. Choose the features you want for each column of the AltaVista home page by clicking on the item and then on the Add button. Be sure to place AltaVista Search in one of the columns (**Figure 5.57**).
4. If you change your mind about a feature, just click on it again in the My Choices menu and then use the Remove button to delete it.
5. Once you've made all of your selections, click on the Done button.
6. Certain features can be further customized by clicking on the EDIT link (**Figure 5.58**) on your My AltaVista page.
7. You can, of course, return to the Customize My AltaVista page at any time to add or remove features.
8. To make My AltaVista your browser's default home page, click on Make this my start page and follow the instructions provided for Windows and Macintosh users.



Figure 5.56 Click on this link to create your own personalized version of the AltaVista home page.

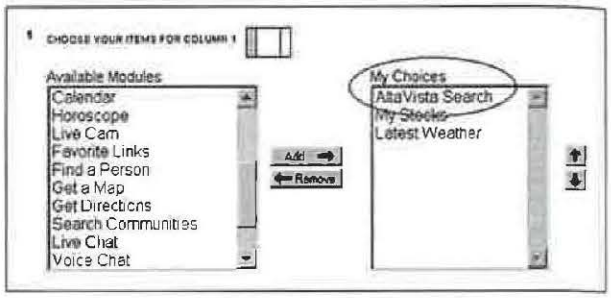


Figure 5.57 Three items have been selected for Column 1 of the My AltaVista page, including the AltaVista Basic Search form.

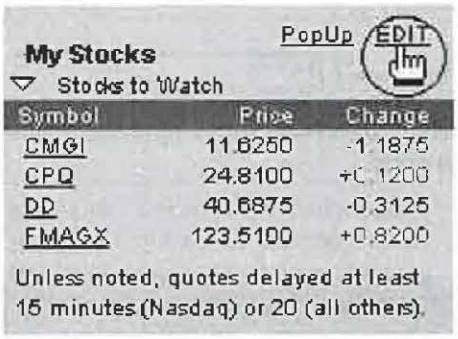


Figure 5.58 To further personalize a feature on your My AltaVista page, click on its EDIT link.

✔ Tips

- One disadvantage of the My AltaVista page is that it doesn't include links to Advanced Search, Power Search, Multimedia Search, and so forth. To use these search options, you first have to return to the AltaVista home page by clicking on Search Home or on the AltaVista logo at the top of the My AltaVista page.
- You can customize your My AltaVista page so that the Basic Search form is the only item that appears on it. But if using a stripped-down AltaVista without any of its "portal" features is what you're after, you might want to try Raging Search (www.ragingsearch.com) instead. It uses the same database of Web pages as AltaVista and gives you all the same great customization features.

6

GOOGLE



Good to Know

- Created at Stanford in 1998, Google is one of the top-rated search sites, earning high praise for its simple interface, spot-on search results, and ad-free home page.
- Google's database is the largest of all the leading search engines, with more than 1.2 billion Web pages.
- Search results include matches from Google's gigantic database as well as the human-compiled Web directory created by the Open Directory Project.
- Google is currently the only search engine that indexes the full text of PDF files and includes them in search results.
- Boolean operators and other "power search" capabilities are fairly limited, but Google does such a good job of producing relevant results that it doesn't matter for most searches.

Google is the newest of the Big Six search engines, but it's giving established competitors like AltaVista (`www.altavista.com`) and even Yahoo (`www.yahoo.com`) a run for their money. The company made headlines in June 2000 with the announcement that its Googlebot spider had indexed more than a billion Web pages—the first (and at this writing still the only) search engine to reach that milestone.

More importantly, Google gets high marks for the *relevancy* of its search results. Using a patent-pending technology, Google not only searches for keywords and phrases, it also employs sophisticated *link analysis* and *data mining* techniques to determine how search results are ranked and presented.

The underlying theory is that the best Web sites on a particular topic will have been discovered by other Web users, who will then include links to those pages in their own Web sites. So the more popular a site—based on the number of other sites that link to it—the higher it appears in your search results.

Google doesn't give you nearly as much control over your queries as AltaVista, HotBot, and Northern Light. You can't use Boolean operators other than OR, for example, nor can you do case-sensitive or wildcard searches. Phrase searches are problematic: If the phrase you're searching for includes a common word like *about* or *with*, Google will ignore it.

But frankly, these limitations aren't really a problem for most day-to-day search tasks. As long as you give Google two or three unique keywords to work with, there's a good chance you'll find what you're looking for within the first few listings in your search results.

✔ Tips

- Google was created in 1998 by Stanford doctoral students Larry Page and Sergey Brin, who dropped out of the computer science program, got venture-capital funding, and launched the service as a privately-held company in 1999.

  Sound familiar? That's probably because Yahoo was also created (four years earlier) by a pair of Stanford doctoral students.

- Both Yahoo and Netscape Search (`search.netscape.com`) have licensed the Google technology, so when you search at those sites, your results are coming (at least in part) from the Google database.

- Prepare to be amazed the first time you do a Google search. Because the work is spread out over a network of some 6,000 PCs, most searches take less than half a second.

- *Google* is a play on the word *googol* (**Figure 6.1**), coined in 1938 by an American mathematician's nine-year-old nephew. The boy had been asked by his uncle what word he would use for a truly gigantic number, represented by a one followed by 100 zeros.

  Google's founders chose the name to convey their commitment to organizing the immense body of information available on the Web. Today, a billion pages. Tomorrow, a googol!

- For more information about Google—press releases, corporate history, job openings, products from the company store—click `All About Google` on the search engine's home page.

goo·gol (go͞o'gôl')
n.

The number 10 raised to the power 100 (10^{100}), written out as the numeral 1 followed by 100 zeros.

[Coined at the age of nine by Milton Sirotta, nephew of Edward Kasner (1878 1955), American mathematician.]

Figure 6.1 The word *google* is based on a mathematical term coined in 1938 by a young boy named Milton Sirotta.

Contact Information

Google, Inc.
Mountain View, California
650-330-0100 (voice)
650-618-1499 (fax)
www.google.com

The Google Home Page

If you prefer to do your Web searches in a clutter-free environment, you'll *love* the Google home page (**Figure 6.2**). Talk about a search engine that's focused on searching! No Web directory to browse. No portal features like news headlines, weather reports, and stock quotes to procrastinate with. No banner ads (with their annoying animated graphics) to distract you.

The Google home page is devoted *exclusively* to its search form. The Web directory and advanced features like field searching and language options are all relegated to underlying pages.

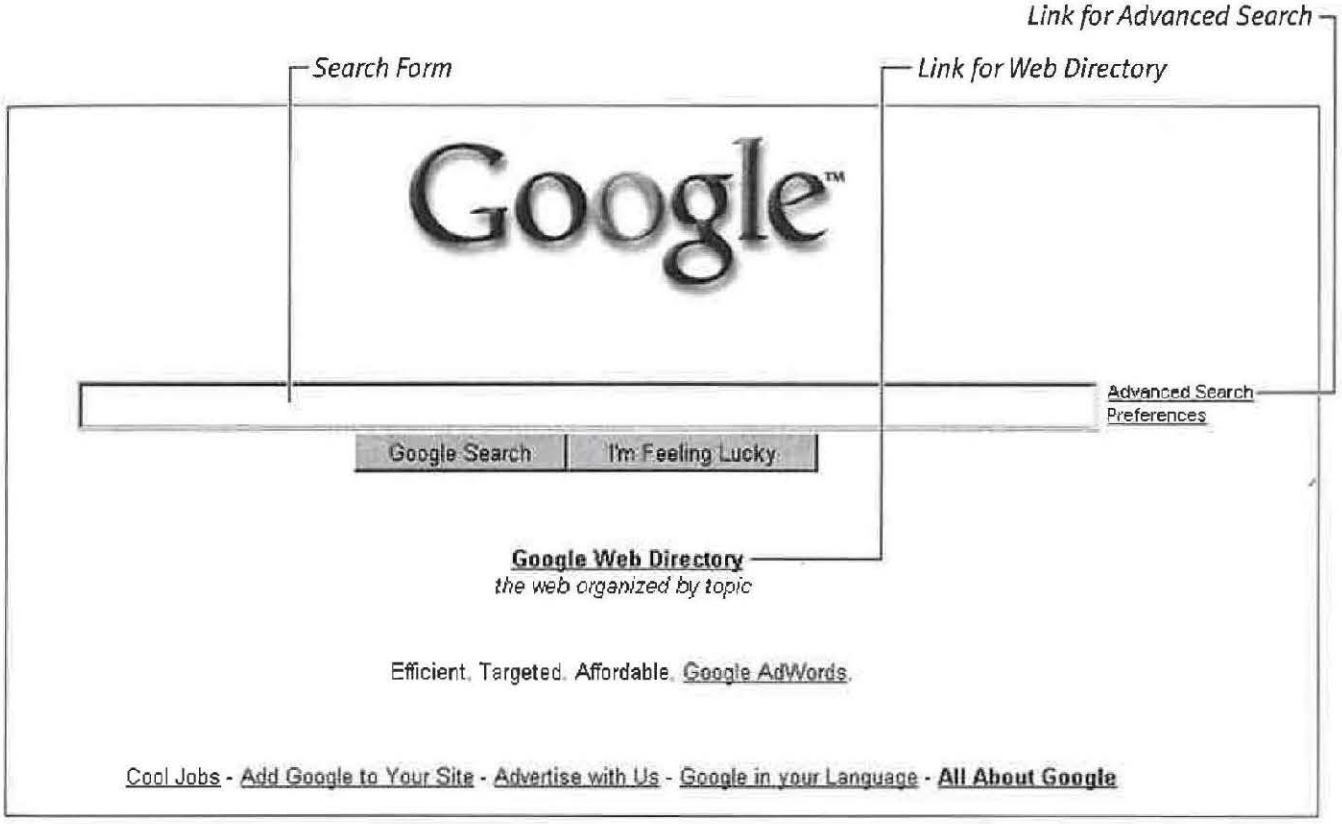


Figure 6.2 The Google home page is the least cluttered of all the leading search engines. The search form gets exclusive billing, and there are no banner ads.