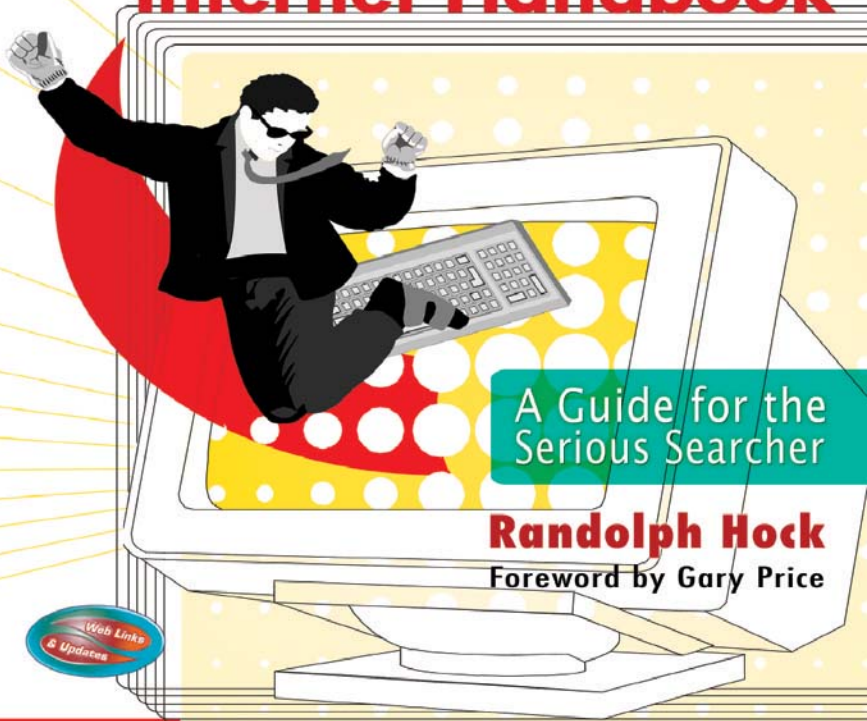
"Buying this book has to be the quickest way you can access years of hard-won Web searching experience. Invaluable."
—William Hann, FreePint

The EXTREME SEARCHER'S Internet Handbook



A Guide for the Serious Searcher

Randolph Hock
Foreword by Gary Price

Web Links & Updates

Extreme tools and techniques for Web information users

The EXTREME SEARCHER'S Internet Handbook

This page intentionally left blank

The EXTREME SEARCHER'S Internet Handbook

A Guide for the Serious Searcher

Randolph Hock

Foreword by Gary Price

CyberAge Books

Medford, New Jersey

The Extreme Searcher's Internet Handbook:
A Guide for the Serious Searcher

 Published by CyberAge Books, an imprint of Information Today, Inc., 143 Old Marlton Pike, Medford, New Jersey 08055.

Library of Congress Cataloging-in-Publication Data

Hock, Randolph, 1944-
The extreme searcher's Internet handbook : a guide for the serious searcher / Randolph Hock ; foreword by Gary Price.
p. cm.
Includes index.
ISBN 0-910965-68-4 (pbk.)
1. Internet searching--Handbooks, manuals, etc. 2. Web search engines--Handbooks, manuals, etc. 3. Computer network resources--Handbooks, manuals, etc. 4. Web sites--Directories. 5. Internet addresses--Directories. I. Title.
ZA4230.H63 2004
025.04--dc22

2003020596

Printed and bound in the United States of America.

Publisher: Thomas H. Hogan, Sr.
Editor-in-Chief: John B. Bryans
Managing Editor: Deborah R. Poulson
Copy Editor: Dorothy Pike
Graphics Department Director: M. Heide Dengler
Book Design: Erica Pannella
Cover Design: Jacqueline Walter
Indexer: Nancy Kopper

Dedication

To Pamela, Matthew, Stephen, and Elizabeth

This page intentionally left blank

CONTENTS

List of Illustrations and Tables

This page intentionally left blank

FOREWORD

Many people believe that searching the Web is as easy as typing a few terms into a box and clicking the search button. Like magic, in a matter of seconds, links to precise, accurate, and current answers will appear.

Unfortunately, this is not the case.

The term "search" is very broad and means different things to different people. For some people it means using an engine like AllTheWeb or Teoma. For others it includes the use of a Web directory focused on a specific topic. For some, search means utilizing not only Web engines but also specialized databases that may contain geographic data, full-text articles, or government information.

Another major issue for the searcher is where to begin. Questions revolve around what each resource does and does not offer. Which is most likely to hold the information I need? How often is the database updated? Can I limit my search to a particular format? Can I change the number of results I see on a results page? What advanced features are available? Knowing where to find this information and then how to apply it can help the Web searcher avoid coming face-to-face with massive amounts of aggravation and wasted time.

Complicating the situation is that as already large Web engines, directories, and databases get larger, it is becoming much more challenging to find what you're looking for. While the retrieval technology is getting better, to find information effectively your search skills must not only be up-to-date, they must be constantly improving.

The good news is that with just a little education and guidance, searching, retrieving, and accessing material on the Web can become easier. Having these skills will make you a better student. Knowing how to save search time will make you a more valuable employee.

These are a few of the reasons why the knowledge, experience, and opinions of Internet search expert Ran Hock are so valuable. This latest book of

Ran's, *The Extreme Searcher's Internet Handbook,* is a resource you'll find yourself referring to on a regular basis.

These days, people tend to rely on a single search tool for all of their Internet research needs. As Ran vividly illustrates, effective searching requires that you know how to use a number of tools. He does a great job of covering the wide range of resources available to the Web searcher. From news engines to quotation databases, specialized directories to online reference works, groups and mailing lists to image and audio finding tools, comparison shopping sites, portals, and more, Ran provides not only the addresses of these sources but the reasons you might want to use them. He also addresses copyright and citation issues, among other important topics for Web searchers.

Ran Hock has done more than write a book. He's created a key resource for both those who need a bit of education in the area of Web research and for experienced searchers who need to verify what a specific search tool offers.

I don't doubt that in a very short period of time your copy will be dog-eared, full of notes, draped with Post-Its, and nothing short of worn out.

Maybe you should buy two copies …

—Gary Price
November, 2003

Gary Price is a reference librarian and information consultant based in suburban Washington, DC. He is co-author of *The Invisible Web: Uncovering Information Sources Search Engines Can't See* and edits ResourceShelf (http://www.resourceshelf.com), a daily update on Web search and other online retrieval news.

ACKNOWLEDGMENTS

First, the great group of people at Information Today, Inc. are due my sincere thanks for their hard work, creativity, and enthusiasm in getting this book to press and into readers' hands. In particular, I am grateful to Tom Hogan, Sr. for the existence of Information Today, Inc., to John Bryans for his encouragement and support and for agreeing to do this book, to Deborah Poulson for shepherding it through the process, to Dorothy Pike for a great job of copyediting, to Heide Dengler for her role on the graphics side of things, and to Erica Panella, Kara Jalkowski, and Jacqueline Walter, the creative artists and designers who gave the book its unique look. Special thanks to Lisa Wrigley not just for her tireless efforts in promoting my books, but also for her unabated enthusiasm for them.

Once again, my appreciation to my friends in the New England Online Users Group for having suggested the phrase "Extreme Searcher" to me several years ago.

Thanks also to the readers of my earlier books for their support, encouragement, and comments. I also offer my gratitude to the many hundreds of students in the courses I teach, for their insights and comments on using the Internet effectively and on what excites them most about the wonders of the Internet.

This page intentionally left blank

INTRODUCTION

Several years ago, Thomas's English Muffins had an ad that proclaimed that the tastiness of their muffins was due to the presence of myriad "nooks and crannies." The same may be said of the Internet. It is in the Internet's nooks and crannies that the true "tastiness" often lies. Almost every Internet user has used Google and probably Yahoo!, and any group of experienced searchers could probably come up with a dozen or so sites that every one of them had used. But even for experienced searchers, time and task constraints have meant that some nooks and crannies have not been explored and exploited. These unexplored areas may be broad Internet resources such as newsgroups, specific types of resources such as multimedia, or the nooks and crannies of a specific site—even Google. This book is intended to be an aid in that exploration.

Back on the culinary scene, I am told that some people don't take the few extra seconds to split their English muffins with a fork, but, driven by their busy schedules, just grab a knife and slice them. This book is written for those seeking to savor the extra tastiness from the Internet. It will hopefully tempt you to discover what the nooks and crannies have to offer, and how to split the Internet muffin with a fork almost as quickly as you can slice it with a knife.

Less metaphorically, this book is written as a guide for researchers, writers, librarians, teachers, and others, covering what serious users need to know to fully take advantage of Internet tools and resources. It focuses on what the serious searcher "has to know" but, for flavor, a dash of the "nice-to-know" is occasionally thrown in. It assumes that you already know the basics, that you are signed up for and frequently use the Internet, and that you know how to use your browser. For those who are not experienced online searchers, my aim is to provide a lot that is new and useful. For those of you with more experience, I hope to reinforce what you know while introducing some new perspectives and new content.

If you are among those who find themselves not just using the Internet, but *teaching* it, the book should help you address an extensive range of questions. Much of what is included is based on my experience training thousands of Internet users from a wide range of professions, across a broad age range, and from more than 40 countries.

BRIEF OVERVIEW OF THE CHAPTERS

The choice of chapter topics reflects congruence between the types of things that experienced Internet users most frequently inquire about and a categorization of the kinds of resources available on the Internet. An argument could certainly be made that the content should have been divided differently. You will notice, for example, that there is a chapter on Finding Products, and you may wonder why there is not one specifically on "company information." This is because the latter topic pervades almost every chapter. Not every chapter will be of utmost interest to every reader, but give each chapter at least a quick glimpse. You may be surprised at what is in some of the nooks (and crannies, of course).

Although the nature of each chapter means that it has an organization of its own, they all contain some things in common. Typically, each chapter includes these aspects:

- Some useful background information, along with suggestions, tips, and strategies for finding and making the most effective use of sites in that area.
- Resource guides that will lead you to collections of links to major sites on the topic.
- Selected sites. I've selected these because (1) they are sites that many if not most readers should be aware of, and/or (2) they are *representative* of types of sites that are useful for the topic. Deciding which sites to include was often difficult. Many of the sites included in this book are considered to be "the best" in their area, but space limitation means that hundreds of great sites had to be excluded. These difficult decisions were made more palatable, however, because the resource guides included in the chapters will lead you quickly to those great sites—you're only one or two clicks away.

Following is a quick rundown of what each chapter covers.

Chapter 1. Basics for the Serious Searcher

This chapter covers background information that serious searchers need to know in order to be conversant with Internet content and issues. It includes some background for understanding more fully the characteristics, content, and searchability of the Internet. For those who find themselves teaching others how to use the Internet, it provides answers to some of the more frequently asked questions. Among the things included in Chapter 1 are a brief history of the Internet, a look at the kinds of "finding tools" available, issues such as retrospective coverage and copyright, resources regarding citing Internet sources, and others for keeping up-to-date.

Chapter 2. General Web Directories and Portals

Although they have quite a bit in common with Web search engines, general Web directories such as Yahoo!, Open Directory, and LookSmart also differ tremendously. This chapter addresses where these tools fit and when they may be most fruitfully used. Even though their databases may include less than 1 percent of what search engine databases cover, general Web directories still serve unique research purposes and in many cases may be the best starting point. This chapter looks at their strengths, their weaknesses, and their special characteristics. Since these general directories are positioned to varying degrees as "portals," this chapter also addresses the "portal" concept.

Chapter 3. Specialized Directories

For accessing immediate expertise in Web resources on a specific topic, there is no better starting point than the right "specialized directory." These sites bring together well-organized collections of Internet resources on specific topics and provide not just a good starting place, but also—importantly—confidence in knowing that no important tools in that area are being missed. Add some content such as news headlines, and you have not just a metasite but a "portal," making these tools even more important as starting points.

Chapter 4. Search Engines

This chapter attempts to provide the background and details about search engines that the serious searcher needs to know in order to get the best results. It examines the largest engines in detail, identifying their strengths and weaknesses and special features. It also presents the case for not getting too excited about metasearch engines.

Chapter 5. Groups and Mailing Lists

Newsgroups, mailing lists, and other interactive forums form a class of Internet resource that too few researchers take advantage of. Useful for a broad range of applications, from solving a software problem to competitive intelligence, these tools can be gold mines. This chapter outlines what they are, why they are useful, and how to locate the ones you need.

Chapter 6. An Internet Reference Shelf

All serious searchers have a collection of tools they use for quick answers—the Web equivalent of a personal reference shelf. This chapter emphasizes the variety of resources that are available for finding quick facts, offers some direction on how to find the right site for a specific need, and suggests several dozen sites that most serious searchers should be aware of.

Chapter 7. Sights and Sounds: Finding Images, Audio, and Video

Not only are there a half billion or so images, audio files, and video files available on the Web, but they are searchable (even better, findable). Whether you are looking for photos of world leaders or rare birds, a famous speech, or the sound of an elephant seal, this chapter provides a look at what resources and tools are available for finding the needed file and discusses techniques for doing so effectively.

Chapter 8. News Resources

This chapter covers the range of news resources that are available on the Internet—news services and newswires, newspapers, news consolidation services, and more—and explains how to most effectively and efficiently find what you are looking for. The chapter emphasizes, on one hand, the searchability of these

resources, and on the other, the limitations the researcher faces, particularly in regard to archival and exhaustivity issues.

Chapter 9. Finding Products Online

Whether for one's own or one's organization's purchase, or for competitive analysis purposes, some searchers find themselves tracking and comparing products online. This chapter shows where to look and how to do it efficiently and effectively.

Chapter 10. Becoming Part of the Internet: Publishing

Beyond using the Internet to gather information, many serious searchers need to have a Web site of their own. Reasons may range from communicating information about the services or products one may provide, to sharing resources with colleagues, to providing a syllabus and links for classes you may be teaching. Although this chapter does not provide the details of how to become a Webmaster, it does offer an overview of what is needed and the options that are available to those who want to move in that direction—including how to get started at no cost by taking advantage of free Web page sites.

Some Introductory Odds and Ends

Most of the sites I discuss in the book do not charge for access. Occasionally, reference is made to sites that require a paid subscription or offer information for a fee, in part as a reminder that (as the serious searcher is already aware) not all of the good stuff is available for free on the Internet. Commercial services such as Lexis/Nexis, Factiva, and Dialog contain proprietary information that is critical for many kinds of research and is not available on the free Web.

Sites are included here because they have useful content. Except for association, government, and academic sites, most of the sites mentioned are supported by ads. On the Internet, just as with television and radio, if the ratio of advertisements to useful content is too high, we can switch to another channel and another Web site. Some of us have come to appreciate the ads to some extent, aware as we are that advertising makes many valuable sites possible.

A Word on "Usage"

Although "Internet" and "Web" are not synonymous, most users do not distinguish between them. When it makes a difference, I use the appropriate term. Where I refer to resources that are generally on the Web part of the Internet, "Web" is used. Where the terms are interchangeable, either term may be used.

Some Final Basic Advice Before You Proceed

Most of us, as we have encountered the Internet over the last decade or so, have learned much of what we know about it in a rather piecemeal fashion, for instance, having been told about a great site, having bumped into it, or having read about it. Although this is, in many ways, an effective approach to exploring the Internet, it can leave gaps in our knowledge. Because each user has individual needs, no single book can fill all of the gaps, but this one attempts to help by providing a better understanding of what is out there as well as some starting points and suggestions for getting what you need—to help you find your way to the most useful nooks and crannies.

As you explore, keep in mind the following three guidelines to help you get the most value from the Internet:

One—"Click everywhere."

Two—"Click where you have never clicked before."

Three—"Split your muffins with a fork."

About The Extreme Searcher's Web Page

As a supplement to this book and to the author's previous book, *The Extreme Searcher's Guide to Web Search Engines*, the author maintains a Web site, The Extreme Searcher's Web Page, at http://www.extremesearcher.com. On the site, you will find links to all of the resources included in this book and updates regarding changes to the search engines discussed in Chapter 4. It is hoped that you will find the site useful enough to bookmark (add to your "favorites" list).

Most of the sites included in this book have been around for a while and will probably remain so for a long time because of their usefulness, quality, and established audience. A few will inevitably disappear or change their address. Every attempt will be made to keep the list of links up-to-date, but should you find a "dead link" there, the author will be most appreciative if you contact him.

Enjoy your visit there and please send any feedback to ran@extreme searcher.com.

Disclaimer

Neither the publisher nor the author makes any claim as to the results that may be obtained through the use of The Extreme Searcher's Web Page or of any of the Internet resources it references or links to. Neither publisher nor author will be held liable for any results, or lack thereof, obtained by the use of this page or any of its links; for any third-party changes; or for any hardware, software, or other problems that may occur as the result of using it. The Extreme Searcher's Web Page is subject to change or discontinuation without notice at the discretion of the publisher and author.

This page intentionally left blank

BASICS FOR THE SERIOUS SEARCHER

In writing this book, I have made the assumption that the reader knows the most basic of the Internet basics—what it is, how to get connected, and so forth. The "basics" covered in this chapter involve background information that serious searchers need to know in order to be fully conversant with Internet content and issues as well as general ways of approaching Internet resources in order to find just what you need. I go into some details already familiar to many readers, but I include this background material for two purposes: (1) so readers might understand more fully the characteristics, content, utility, and nuances of the Internet, in order to search more effectively, and (2) to help those who find themselves teaching others how to use the Internet by providing answers to some of the more frequently asked questions.

As for the general approaches to finding the right resources, this chapter provides an overview and comparison of the kinds of "finding tools" available and a set of strategies that can be applied. The strategies coverage goes into some detail on topics such as Boolean logic that will also be encountered elsewhere in the book. Integral to all of this are some aspects and issues regarding the content that is found on the Internet. These aspects include the questions of retrospective coverage, quality of content, and general accessibility of content, particularly the issue of the Invisible Web. Woven into this content fabric are issues, such as copyright, that affect how information found on the Internet can be used. Although only lightly touched upon, it is important that every serious user have an awareness of these issues. Lastly, the chapter provides some resources useful for keeping up on the latest Internet tools, content, and issues.

THE PIECES OF THE INTERNET

First, the "Internet" and the "Web" are not synonymous, although they are frequently used interchangeably. As late as the mid-1990s, the Internet had some clearly distinguishable parts, as defined by their functions. Much Internet

usage could be thought of as Internet *sans* content. It was simply a communications channel that allowed easy transfer of information. Typically, a user at one university could use it to send a file, or request a file, from someone at another university using FTP (File Transfer Protocol). Use of the Internet for sending e-mail was becoming tremendously popular at that time. A user of a commercial search service such as Dialog or LexisNexis could harness it as an alternative to proprietary telecommunications networks, basically sending and receiving proprietary information. "Content" parts of the Internet could be found, such as Usenet Newsgroups, where anyone with a connection could access a body of publicly available information. Gophers (menu-based directories allowing access to files, mainly at universities) were also beginning to provide access to content.

The world changed, and content was destined to become king, when, in 1991, Tim Berners-Lee at CERN (Conseil European pour la Recherché Nucleaire) in Geneva created the World Wide Web. The Web provided an easy-to-use interface for both potential content providers and users, with a GUI (Graphical User Interface) incorporating hypertext point-and-click navigation of text, graphics, and sounds, and creating what was at that time for most of us unimaginable potential for access to information.

Within less than a half-decade, the Web had overtaken e-mail and FTP in terms of Internet traffic. By 2000, usage of the other parts of the Internet was becoming fused into the Web. Usenet newsgroups were being accessed through a Web interface. Web-based e-mail was becoming the main—or only—form of e-mail for millions of people. FTP was typically being done through a Web interface. Gophers were replaced by Web directories and search engines, and any gophers you now find are likely to be in your backyard.

A VERY BRIEF HISTORY

The following selection of historical highlights provides a perspective for better understanding the nature of the Internet. It should be emphasized that the Internet is the result of many technologies (computing, time-sharing of computers, packet-switching, etc.) and many visionaries and great technical thinkers coming together over a period of a few decades. In addition, what they were able to accomplish was dependent upon minds and technologies of preceding decades. This selection of highlights, as merely a sampling, by necessity leaves out many essential technical achievements and notable contributors.

The points shown here are drawn primarily from the resources listed at the end of this time line.

1957 Sputnik launched by USSR.

1958 Largely as a result of the Sputnik launch, ARPA (Advanced Research Projects Agency) is created to put the U.S. ahead in science and technology. High among its interests was computer technology.

1962 J. C. R. Licklider writes of his vision of a globally interconnected group of computers providing widespread access to data and programs. RAND Corporation starts research on distributed communications networks for military purposes.

Early 1960s Packet-switching moves from theory to practice.

Mid- to Late 1960s ARPA develops ARPANET to promote the "cooperative networking of time-sharing computers" with four host computers connected by the end of 1969 (Stanford Research Institute, UCLA, UC Santa Barbara, and the University of Utah).

1965 The term "hypertext" is coined by Ted Nelson.

1968 The Tymnet nationwide time-sharing network is built.

1971 ARPANET has grown to twenty-three hosts, including universities and government research centers.

1972 The International Network Working Group (INWG) is formed to advance and set standards for networking technologies. The first chairman is Vinton (Vint) Cerf, who is later often referred to as the "Father of the Internet."

1972–1974 Commercial database services—Dialog, SDC Orbit, Lexis, New York Times DataBank, and others—begin making their subscription services widely available through dial-up networks.

1973 ARPANET makes its first international connections (University College of London (England) and Royal Radar Establishment (Norway).

1974 "A Protocol for Packet Network Interconnection" is published by Vint Cerf and Bob Kahn, which specified the details of TCP (Transmission Control Protocol).

Bolt, Beranek & Newman, contractor for ARPANET, opens a commercial version of the ARPANET, Telenet, the first public packet-data service.

1977 There are 111 hosts on the Internet.

1978 TCP is split into TCP and IP (Internet Protocol).

1979 The first Usenet discussion groups are created by Tom Truscott, Jim Ellis, and Steve Bellovin, graduate students at Duke University and the University of North Carolina. It quickly spreads worldwide.

The first emoticons (smileys) are suggested by Kevin McKenzie.

1980s The personal computer becomes a part of millions of people's lives.

There are 213 hosts on ARPANET.

BITNET (Because It's Time Network) is started, providing e-mail, electronic mailing lists, and FTP service.

CSNET (Computer Science Network) is created by computer scientists at Purdue University, the University of Washington, RAND Corporation, and BBN, with National Science Foundation (NSF) support. It provides e-mail and other networking services to researchers who did not have access to ARPANET.

1982 The term "Internet" is first used.

TCP/IP is adopted as the universal protocol for the Internet.

Name servers are developed, allowing a user to get to a computer without specifying the exact path.

There are 562 hosts on the Internet.

France Telecom begins distributing Minitel terminals to subscribers free of charge, providing videotext access to the Teletel system. Initially providing telephone directory lookups, then chat and other services, Teletel is the first widespread home implementation of these types of network services.

Orwell's vision, fortunately, is not fulfilled, but computers are soon to be in almost every home.

There are over 1,000 hosts on the Internet.

1985 The WELL (Whole Earth 'Lectronic Link) is started. Individual users, outside of universities, can now easily participate on the Internet.

There are over 5,000 hosts on the Internet.

1986 NSFNET (National Science Foundation Network) is created. The backbone speed is 56K. (Yes, as in the total transmission capability of a 56K dial-up modem.)

1987 There are over 10,000 hosts on the Internet.

1988 The NSFNET backbone is upgraded to a T1 at 1.544Mbps (megabits per second).

1989 There are over 100,000 hosts on the Internet.

ARPANET goes away.

There are over 300,000 hosts on the Internet.

1991 Tim Berners-Lee at CERN (Conseil European pour la Recherché Nucleaire) in Geneva, introduces the World Wide Web.

NSF removes the restriction on commercial use of the Internet.

The first gopher is released, at the University of Minnesota, which allows point-and-click access to files on remote computers.

The NSFNET backbone is upgraded to a T3 (44.736 Mbps).

1992 There are over 1,000,000 hosts on the Internet.

Jean Armour Polly coins the phrase "surfing the Internet."

1994 The first graphics-based browser, Mosaic, is released.

Internet talk radio begins.

WebCrawler, the first successful Web search engine is introduced.

A law firm introduces Internet "spam."

Netscape Navigator, the commercial version of Mosaic, is shipped.

1995 NSFNET reverts back to being a research network. Internet infrastructure is now primarily provided by commercial firms.

RealAudio is introduced, meaning that you no longer have to wait for sound files to download completely before you begin hearing them, and allowing for continued ("streaming") downloads.

Consumer services such as CompuServe, America Online, and Prodigy begin to provide access through the Internet instead of only through their private dial-up networks.

1996 There are over 10,000,000 hosts on the Internet.

1999 Microsoft's Internet Explorer overtakes Netscape as the most popular browser.

Testing of the registration of domain names in Chinese, Japanese, and Korean languages begins, reflective of the internationalization of Internet usage.

2001 Mysterious monolith does not emerge from the Earth and no evil computers take over any spaceships (as far as we know).

2002 Google is indexing more than 3 billion Web pages.

2003 There are more than 200,000,000 hosts on the Internet.

Internet History Resources

Anyone interested in information on the history of the Internet beyond this selective list is encouraged to consult the following resources.

A Brief History of the Internet, version 3.1

http://www.isoc.org/internet-history

By Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, Stephen Wolff. This site provides historical commentary from many of the actual people who were involved in the creation of the Internet.

Internet History and Growth

http://www.isoc.org/internet/history/2002_0918_Internet_History_and_Growth.ppt

By William F. Slater. This PowerPoint presentation provides a good look at the pioneers of the Internet and provides an excellent collection of statistics on Internet growth.

Hobbes' Internet Timeline

http://www.zakon.org/robert/internet/timeline

This detailed timeline emphasizes technical developments and who was behind them.

SEARCHING THE INTERNET: WEB "FINDING TOOLS"

Whether your hobby or profession is cooking, carpentry, chemistry, or anything in-between, you know that the right tool can make all the difference. The same is true for searching the Web. A variety of tools are available to help you find what you need, and each does things a little differently, sometimes with different purposes and different emphases, as well as different coverage and different search features.

To understand the variety of tools, it can be helpful to think of most finding tools as falling into one of three categories (although many tools will be hybrids). These three categories of tools are (1) general directories, (2) search engines, and (3) specialized directories. The third category could indeed be lumped in with the first because both are directories, but for a couple of reasons discussed later, it is worthwhile to separate them.

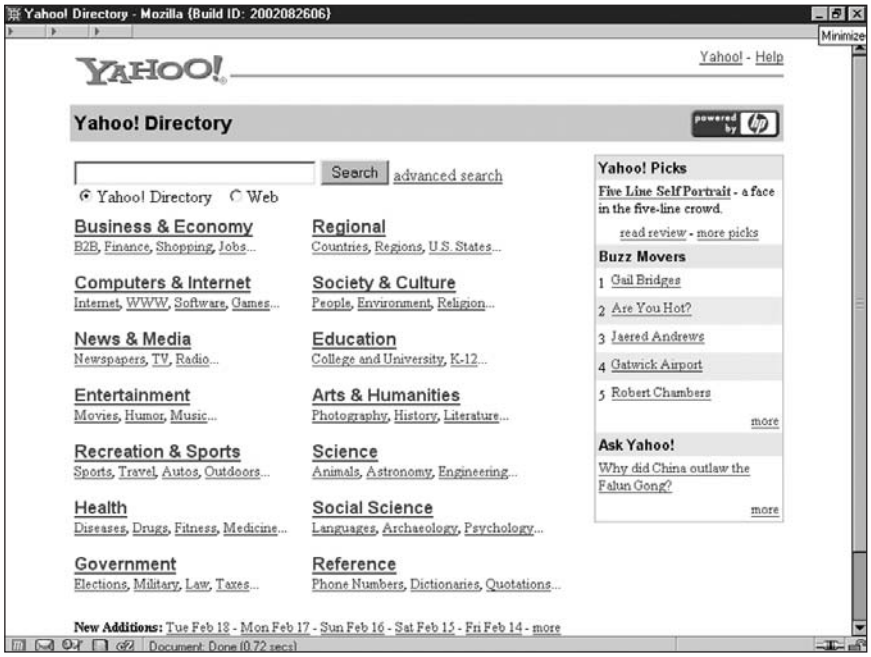
All three of these categories may incorporate another function, that of a portal, a Web site that provides a gateway not only to links, but to a number of other information resources going beyond just the searching or browsing function. These resources may include news headlines, weather, professional directories, stock market information, a glossary, alerts, and other kinds of handy information. A portal can be general, as in the case of Yahoo!'s My Yahoo!, or it can be specific for a particular discipline, region, or country.

Other finding tools serve other kinds of Internet content, such as newsgroups, mailing lists, images, and audio. These tools may exist either on sites of their own or they may be incorporated into the three main categories of tools. These specialized tools will be covered in later chapters.

General Web Directories

The general Web directories are Web sites that provide a large collection of links arranged in categories to enable browsing by subject area, such as Yahoo!, Open Directory, and LookSmart. Their content is (usually) hand picked by human beings who ask the question: "Is this site of enough interest to enough people that it should be included in the directory?" If the answer is yes (and in some cases, if the owner of the site has paid a fee), the site is added and placed in the directory's database (catalog) and is listed in one or more of the subject categories. As a result of this process, these tools have two major characteristics: They are *selective* (sites have had to meet the selection criteria), and they are *categorized* (all sites are arranged in categories—see Figure 1.1). Because of the selectivity, the user of these directories is working, theoretically, with higher quality sites—the wheat and not the chaff. Because the sites included are arranged in categories, the user has the option of starting at the top of the hierarchy of categories and browsing down until the appropriate level of specificity is reached. Also, usually only one entry is made for each site, instead of including, as in search engines, many pages from the same site. The size of the database of general Web directories is much smaller than that created and used by Web search engines, the former containing usually 2 to 3 million sites and the latter from 1 to 3 billion pages. Web directories are designed primarily for browsing and for general questions. Sites on very specific topics, such as "UV-enhanced dry stripping of silicon nitride films" or "social security retirement program reform in Croatia" are generally not included. As a result, directories are most successfully used for *general*,

Figure 1.1



Yahoo!'s Main Directory Page

rather than *specific* questions, for example, "Types of Chemical Reactions" or "social security." Although browsing through the categories is the major design idea behind general Web directories, they do provide a search box to allow you to bypass the browsing and go directly to the sites in the database.

When to Use a General Directory

TIP:

If your question contains one or two concepts, consider a directory. If it contains three or more, definitely start with a search engine.

General Web directories are a good starting place when you have a very general question (museums in Paris, dyslexia), or when you don't quite know where to go with a broad topic and would like to browse down through a category to get some guidance.

General Web directories are discussed in detail in Chapter 2.

Web Search Engines

Whereas a directory is a good start when you want to be directed to just a few selected items on a fairly general topic, search engines are the place to go when you want something on a fairly specific topic (ethics of human cloning, Italian paintings of William Stanley Haseltine). Instead of searching brief

descriptions of 2 to 3 million Web sites, these services allow you to search virtually every word from 2 to 3 billion Web pages. In addition, Web search engines allow you to use much more sophisticated techniques, allowing you to much more effectively focus in on your topic. The pages included in Web search engines are not placed in categories (hence, you cannot browse a hierarchy), and no prior human selectivity was involved in determining what is in the search engine's database. You, as the searcher, provide the selectivity by the search terms you choose and by the further narrowing techniques you may apply.

When to Use Search Engines

If your topic is very specific or you expect that very little is written on it, a search engine will be a much better starting place than a directory. If you need to be exhaustive, use a search engine. If your topic is a combination of three or more concepts (e.g., "Italian" "paintings" "Haseltine"), use a search engine. (See Chapter 4 for more details on search engines.)

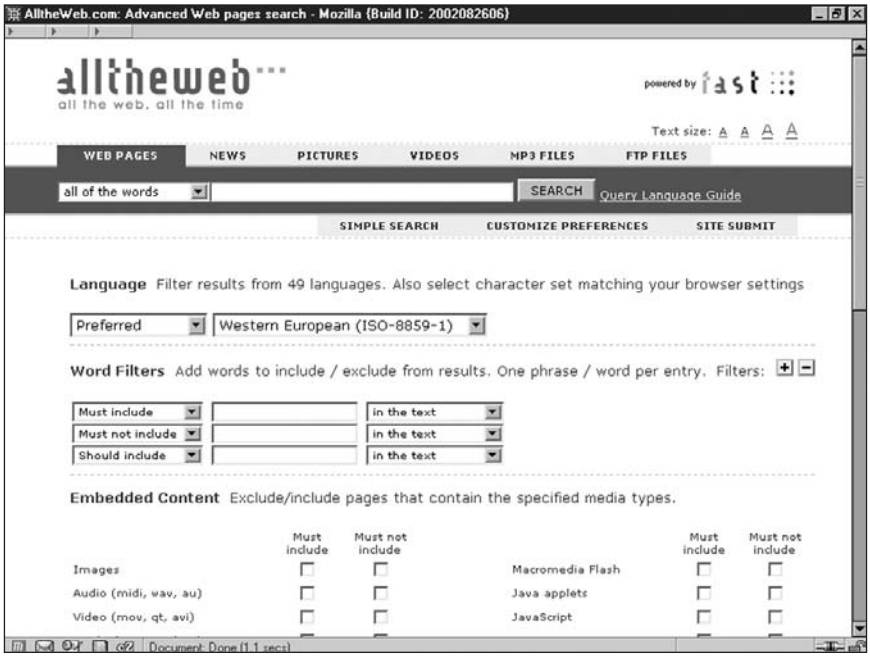


Figure 1.2

Web Search Engine—AllTheWeb's Advanced Search Page

Specialized Directories (Resource Guides, Research Guides, Metasites)

Specialized Web directories are collections of selected Internet resources (collections of links) on a particular topic. The topic could range from something as broad as medicine to something as specific as biomechanics. These sites go by a variety of names such as resource guides, research guides, metasites, cyberguides, and webliographies. Although their main function is to provide links to resources, they often also incorporate some additional portal features such as news headlines.

Indeed, this category could have been lumped in with the general Web directories, but it is kept separate for two main reasons. First, the large general directories, such as Yahoo! and Open Directory, all have a number of things in common besides being general. They all provide categories you can browse, they all also have a search feature, and when you get to know them, they all tend to have the same "look and feel" in other ways as well. The second main reason for keeping the specialized directories as a separate category is that they deserve greater attention than they often get. More searchers need to tap into their extensive utility.

When to Use Specialized Directories

Use specialized directories when you need to get to know the Web literature on a topic, in other words, when you need a general familiarity with the major resources for a particular discipline or a particular area of study. These sites can be thought of as providing *some* immediate expertise in using Web resources in the area of interest. Also, when you are not sure of how to narrow your topic and would like to browse, these sites can often be better starting places than a general directory because they may reflect a greater expertise in the choice of resources for a particular area than would a general directory, and they often include more sites on the specific topic than are found in the corresponding section of a general directory.

Specialized directories are discussed in detail in Chapter 3.

GENERAL STRATEGIES

First, there is no right or wrong way to search the Internet. If you find what you need and find it quickly, your strategy is good. Keep in mind, though, that

finding what you need involves issues such as Was it really the correct answer?, Was it the best answer?, and Was it the complete answer?

At the broadest level, assuming that your question is one for which the Internet is the best starting place, one approach to a finding what you need on the Internet is to first answer the following three questions.

1. Exactly what is my question? (Identification of what you really need and how exhaustive or precise you need to be.)
2. What is the most appropriate tool with which to start? (See the previous sections on the categories of finding tools.)
3. What search strategy should I start with?

These three steps often take place without much conscious effort and may take a matter of seconds. For instance, you want to find out who General Carl Schurz was, you go to your favorite search engine and throw in those three words. The quick-and-easy, keep-it-simple approach is often the best.

Even for a more complicated question, it is often worthwhile to start with a very simple approach in order to get a sense of what is out there, then develop a more sophisticated strategy based on an analysis of your topic into concepts.

Organizing Your Search by Concepts

Both a natural way of organizing the world around us and a way of organizing your thoughts about a search is to think in terms of concepts. Thinking in concepts is a central part of most searches. The concepts are the ideas that must be present in order for a resultant answer to be relevant, each concept corresponding to a required criterion. Sometimes a search is so specific that a single concept may be involved, but most searches involve a combination of two, three, or four concepts. For instance, if our search is for "hotels in Albuquerque," our two concepts are "hotels" and "Albuquerque." If we are trying to identify Web pages on this topic, any Web page that includes both concepts possibly contains what we are looking for and any page that is missing either of those concepts is not going to be relevant.

The experienced searcher knows that for any concept, more than one term present in a record (on a Web page) may indicate the presence of the concept, and these alternate terms also need to be considered. Alternate terms may include, among other things, (1) grammatical variations (e.g., electricity, electrical), (2) synonyms, near-synonyms, or closely related terms (e.g., culture, traditions), and (3) a term and its narrower terms. For an exhaustive search in which "Baltic states"

is a concept, you may want to also search for Latvia, Lithuania, and Estonia. In an exhaustive search for information on the production of electricity in the Baltic states, you would not want to miss that Web page that dealt specifically with "Production of Electricity in Latvia."

When the idea of thinking in concepts is expanded further, it naturally leads to a discussion of Boolean logic, which will be covered in Chapter 4. In the meantime, the major point here is that, in preparing your search strategy, think about what concepts are involved, and remember that, for most concepts, looking for alternate terms is important.

A BASIC COLLECTION OF STRATEGIES

Just as there is no one right or wrong way to search the Internet, there can be no list of definitive steps to follow, or one specific strategy to follow, in preparing and performing every search. Rather, it is useful to think in terms of a toolbox of strategies and to select whichever tool or combination of tools seems most appropriate for the search at hand. Among the more common strategies, or strategic tools, or approaches for searching the Internet are the following:

1. Identify your basic ideas (concepts) and *rely on the built-in relevance ranking* provided by search engines. In the major search engines and many other search sites, when you enter terms, only those records (Web pages)

Figure 1.3

Ranked Output

that contain all those terms will be retrieved, and the engine will automatically rank the order of output based on various criteria.

2. Use simple *narrowing techniques* if your results need narrowing:
   - Add another concept to narrow your search (instead of *hotels Albuquerque*, try *inexpensive hotels Albuquerque*)
   - Use quotation marks to indicate phrases when a phrase more exactly defines your concept(s) than if the words occur in different places on the page, for example, "foreign policy." Most Web sites that have a search function allow you to specify a phrase (a combination of two or more adjacent words, in the order written) by the use of quotation marks.
   - Use a more specific term for one or more of your concepts (instead of *intelligence*, perhaps use *military intelligence*).
   - Narrow your results to only those items that contain your most important terms in the title of the page. (These kinds of techniques will be discussed in Chapter 4.)
3. *Examine your first results* and look for, then use, terms you might not have thought of at first.
4. *If you do not seem to be getting enough relevant items, use the Boolean OR* operation to allow for alternate terms, for example, *electrical OR electricity* would find all items that have either the term *electrical* or the term *electricity*. How you express the OR operation varies with the finding tool.
5. Use *a combination of Boolean operations* (AND, OR, NOT, or their equivalents) to identify those pages that contain a specific combination of concepts and alternate terms for those concepts (for example, to get all pages that contain either the term *cloth* or the term *fabric* and also contain the words *flax* and *shrinkage*). As will be discussed later, Boolean is not necessarily complicated, is often implied without you doing anything, and can be as simple as choosing between "all of these words" or "any of these words" options.
6. *Look at what else the finding tools (particularly search engines) can do* to allow you to get as much as you need—and only what you need. Advanced search pages are probably the first place you should look.

Ask five different experienced searchers and you will get five different lists of strategies. The most important thing is to have an awareness of the kinds of

techniques that are available to you for getting everything you need and, at the same time, only what you need.

CONTENT ON THE INTERNET

Not only the amount of information but the kinds of information available and searchable on the Internet continue to increase rapidly. In understanding what you are getting—and not getting—as a result of a search of the Internet requires consideration of a number of factors, such as the time frames covered, quality of content, and a recognition that various kinds of material exist on the Internet that are not readily accessible by search engines. In using the content found on the Internet, other issues must also be considered, such as copyright.

Assessing Quality of Content

TIP:

For most sites, if you don't immediately see how to get back to the home page, try clicking on the site's logo. It usually works.

A favorite complaint by those who are still a bit shy of the Internet is that the quality of information found there is often low. The same could be said about information available from a lot of other resources. A newsstand may have both the Economist and The National Enquirer on its shelves. On television you will find both The History Channel and infomercials. Experience has taught us how, in most cases, to make a quick determination of the relative quality of the information we encounter in our daily lives. In using the Internet, many of the same criteria can be successfully applied, particularly those criteria we are accustomed to applying to traditional literature resources, both popular and academic.

These traditional literature evaluation techniques/criteria that can be applied in the Internet context include:

1. Consider the source.

From what organization does the content originate? Look for the organization identified both on the Web page itself and at the URL. Is the content identified as coming from known sources such as a news organization, a government, an academic journal, a professional association, or a major investment firm? Just because it does not come from such a source is certainly not cause enough to reject it outright. On the other hand, even if it does come from such a source, don't bet the farm on this criterion alone.

Look at the URL. Often you will immediately be able to identify the owner. Peel back the URL to the domain name. If that does not adequately identify it, you can check details of the domain ownership for U.S. sites on sites that

provide access to the Whois database, such as Network Solution's (VeriSign) http://www.networksolutions.com/cgi-bin/whois/whois. For other countries, similar sites are available.

Be aware that some look-alike domain names are intended to fool the reader as to the origin of the site. The top level domain (edu, com, etc.) may provide some clues about the source of the information, but do not make too many assumptions here. An edu or ac domain does not necessarily assure academic content, given that students as well as faculty can often easily get a space on the university server.

A cedilla " ~ " in a directory name is often an indication of a personal page. Again, don't reject something on such a criterion alone. There are some very valuable personal pages out there.

Is the actual author identified? Is there an indication of the author's credentials, the author's organization? Do a search for other things by the same author. Does she or he publish a lot on spontaneous human combustion and extraterrestrial origins of life on earth? If you recognize an author's name and the work does not seem consistent with other things from the same author, question it. It is easy to impersonate someone on the Internet.

2. Consider the motivation.

What seems to be the purpose of the site—academic, consumer protection, sales, entertainment (don't be taken in by a spoof), political? There is, of course, nothing inherently bad (or for that matter necessarily inherently good), in any of those purposes, but identifying the motivation can be helpful in assessing the degree of objectivity. Is any advertising on the page clearly identified, or is advertising disguised as something else?

3. Look at the quality of the writing.

If there are spelling and grammatical errors, assume that the same level of attention to detail probably went into the gathering and reporting of the "facts" given on the site.

4. Look at the quality of the documentation of sources cited.

First, remember that even in academic circles, the number of footnotes is not a true measure of the quality of a work. On the other hand, and more importantly, if facts are cited, does the page identify the origin of the facts. If a lot rests on the information you are gathering, check out some of the cited sources to see that they really do give the facts that were quoted.

5. Is the site and its contents as current as it should be?

If a site is reporting on current events, the need for currency and the answer to the question of currency will be apparent. If the content is something that should be up-to-date, look for indications of timeliness, such as a "last updated" date on the page or telling examples of outdated material. If, for example, it is a site that recommends which search engines to use, and if WebCrawler is still listed, don't trust the currency (or for that matter, accuracy) of other things on the page. What is the most recent material that is referred to? If a number of links are "dead links," assume that the author of the page is not giving it much attention.

6. For facts you are going to use, verify using multiple sources, or choose the most authoritative source.

Unfortunately, many facts given on Web pages are simply wrong, from carelessness, exaggeration, guessing, or for other reasons. Often they are wrong because the person creating that page's content did not check the facts. If you need a specific fact, such as the date of an historic event, look for more than one Web page that gives the date and see if they agree. Also remember that one Web site may be more authoritative than another. If you have a quotation in hand and want to find who said it, you might want to go to a source such as Bartleby.com (which includes very respected quotations sources), instead of taking the answer from Web pages of lesser-known origins.

For more details and other ideas on the topic of the evaluating quality of information found on the Internet, the following two resources will be useful.

The Virtual Chase:
Evaluating the Quality of Information on the Internet
http://www.virtualchase.com/quality

Created and maintained by Genie Tyburski, this site provides an excellent overview of the factors and issues to consider when evaluating the quality of information found on a Web site. She provides checklists and links to other checklists as well as examples of sites that demonstrate both good and bad qualities.

Evaluating the Quality of World Wide Web Resources
http://www.valpo.edu/library/evaluation.html

This site from Valparaiso University provides a detailed set of criteria and also several dozen links to other sites that address the topic of evaluating Web resources. It also has links to exercises and worksheets on the topic.

Retrospective Coverage of Content

It is tempting to say that a major weakness of Internet content is lack of retrospective coverage. This is certainly an issue for which the serious user should have a high level of awareness. It is also an issue that should be put in perspective. The importance and amount of relevant retrospective coverage available depends on the kind of information you are seeking at any particular moment, and on your particular question. It is safe to say that no Web pages on the Internet were created before 1991.

Books, Ancient Writings, and Historical Documents

The lack of pre-1991 Web pages does not mean that earlier content is not available. Indeed, if a work is moderately well-known and was written before 1920 or so, you are as likely to find it on the Internet as in a small local public library. Take a look at the list of works included in the Project Gutenberg site and The Online Books Page (see Chapter 6) where you will find works of Cicero, Balzac, Heine, Disraeli, Einstein, and thousands of other authors. Also look at some of the other Web sites discussed in Chapter 6 for sources of historical documents.

Scholarly and Technical Journals and Popular Magazines

If you are looking for the full text of journal or magazine articles written several years ago, you are not likely to find them free on the Internet (and, for most journal articles, you are not even likely to find the ones written this week, last month, or last year). This lack of content is more a function of copyright and requirements for paid subscriptions than a matter of the retrospective aspect. The distinction also needs to be made here between free material and "for fee" material on the Internet. On a number of sources on the Internet (such as ingenta) you can find references to scholarly and other material going back a several years. Most likely you will need to pay to see the full text, but fees tend to be very reasonable. Whatever source you use for serious research, Internet or other, examine the source to see how far back it goes.

Newspapers and Other News Sources

If, when you speak of news, you think of "new news," retrospective coverage is not an issue. If you are looking for newspaper or other articles that go back more than a few days, the time span of available content on any particular site is crucial. In 2000, many newspapers on the Internet contained only the current day's stories, with a few having up to a year or two of stories. Fortunately, more and more newspaper and other news sites are archiving their material, and you may find several years of content on the site. Look closely at the site to see exactly how far back the site goes.

Old Web Pages

A different aspect of the retrospective issue centers on the fact that many Web pages change frequently and many simply go away. Pages that existed in the early 1990s are likely to either be gone or have different content than they did then. This becomes a significant problem when trying to track down early content or citing early content. Fortunately, there are at least partial solutions to the problem. For very recent pages that may have disappeared or changed in the last few days or weeks, Google's "cache" option may help. For Web pages in Google's database, Google has stored a copy. If you find the reference to the page in Google, but when you try to go to it, the page is either completely gone, or the content that you expected to find on the page is no longer there, click on the "Cached" option and you will get to a copy of the page as it was when Google last indexed it. Even if you initially found the page elsewhere, search for it in Google, and if you find it there, try the cache.

For locating earlier pages and their content, try the Wayback Machine.

Wayback Machine—Internet Archive

http://www.archive.org

The Wayback Machine provides the Internet Archive, which has the purpose of "offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format." It allows you to search over 10 billion pages and see what a particular page looked like at various periods in Internet time. A search yields a list of what pages are available for what dates as far back as 1996. (See Figure 1.4.) As well as Web pages, it also archives moving images, texts, and audio. Its producers claim it is the largest database ever built.

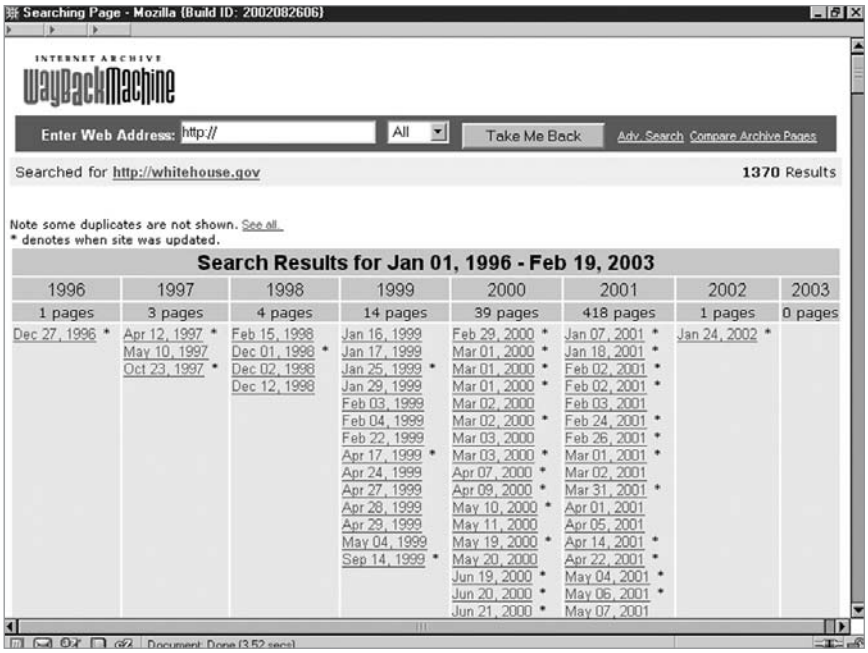


Figure 1.4

Wayback Machine Search Result Showing Pages Available in the Internet Archive for whitehouse.gov.

CONTENT—THE INVISIBLE WEB

No matter how good you are at using Web search engines and general directories, there are valuable resources on the Web that search engines will not find for you. You can get to most of them if you know the URL, but a search engine search will probably not find them for you. These resources, often referred to as the "Invisible Web," include a variety of content, including, most importantly, databases of articles, data, statistics, and government documents. The "invisible" refers to "invisible to search engines." There is nothing mysterious or mystical involved.

The Invisible Web is important to know about because it contains a lot of tremendously useful information—and it is large. Various estimates put the size of the Invisible Web at from two to five hundred times the content of the visible Web. Before that number sinks in and alarms you, keep in mind the following:

1. There is a lot of very important material contained in the Invisible Web.
2. For the information that is there that you are likely to have a need for, and the right to access, there are ways of finding out about it and getting to it.

3. In terms of volume, most of the material is material that is meaningless except to those who already know about it, or to the producer's immediate relatives. Much of the material that can't be found is probably not worth finding.

To adequately understand what this is all about, one must know why some content is invisible. Note the use of the word "content" instead of the word "sites." The main page of invisible Web sites is usually easy to find and is covered by search engines. It is the rest of the site (Web pages and other content) that may be invisible. Search engines do not index certain Web content mainly for the following reasons:

1. The search engine *does not know about the page*. No one has submitted the URL to the search engine and no pages currently covered by the search engine have linked to it. (This falls in the category, "Hardly anyone cares about this page, you probably don't need to either.")
2. The search engines have *decided not to index* the content because it is too deep in the site (and probably less useful), it is a page that changes so frequently that indexing the content would be somewhat meaningless (as, for example in the case of some news pages), or the page is generated dynamically and likewise is not amenable to indexing. (Think in terms of "Even if you searched and found the page, the content you searched for would probably be gone.")
3. The search engine is *asked not to index* the content, by the presence of a robots.txt file on the site that asks engines not to index the site, or specific pages, or particular parts of the site. (A lot of this content could be placed in the "It's nobody else's business" category.)
4. The search engine *does not have or does not utilize a technology that would be required* to index non-HTML content. This applies to files such as images and audio files. Until 2001, this category included file types such as PDF (Portable Document Format files), Excel files, Word files, and others, that began to be indexed by the major search engines in 2001 and 2002. Because of this increased coverage, the Invisible Web may be shrinking, proportionate to the size of the total Web.
5. The search engine cannot get to the pages to index them because *it encounters a request for a password or the site has a search box* that must be filled out in order to get to the content.

It is the last part of the last category that holds the most interest for the searcher—sites that contain their information in databases. Prime examples of such sites would be phone directories, literature databases such as Medline, newspaper sites, and patents databases. As you can see, if you can find out that the site exists, then you (without going through a search engine) can search the site contents. This leads to the obvious question of where one finds out about sites that contain unindexed (Invisible Web) content.

The three sites listed below are directories of Invisible Web sites. Keep in mind that they list and describe the overall site, they do not index the contents of the site. Therefore, these directories should be searched or browsed at a broad level. For example, look for "economics" not a particular economic indicator, or for sites on "safety" not "workplace safety." As you identify sites of interest, bookmark them.

You may also want to look at the excellent book on the Invisible Web by Chris Sherman and Gary Price (*The Invisible Web: Uncovering Information Sources Search Engines Can't See*. CyberAge Books. Medford, NJ USA. 2001).

Direct Search

http://www.freepint.com/gary/direct.htm

The "grandfather" of Invisible Web directories, this site was created and is maintained by Gary Price (co-author of *The Invisible Web*). The sites listed here are carefully selected for quality of content, and you can either search or browse.

invisible-web.net

http://www.invisible-web.net

By the authors of *The Invisible Web*, this is the most selective of the three Invisible Web directories listed here. It contains about 1,000 entries and you can either browse or search.

CompletePlanet

http://completeplanet.com

The site claims "103,000 searchable databases and specialty search engines," but a significant number of the sites seem to be individual pages (e.g., news articles) and many of the databases are company catalogs, Yahoo! categories, and the like, not necessarily "invisible." It lists a lot of useful resources, but the content also emphasizes how trivial much Invisible Web material can be.

COPYRIGHT

Because of the seriousness of the implications of this topic, this section could extend for thousands of words. Because this chapter is about basics, though, a few general points will be made and the reader is encouraged to go for more detail to the sources listed next, which are much more authoritative and extensive on the copyright issue. If you are in a large organization, particularly an educational institution, you may want to check your organization's site for local guidelines regarding copyright.

Copyright—Some Basic Points

Here are some basic points to keep in mind regarding copyright.

1. "Copyright is a form of protection provided by the laws of the United States (title 17, U.S. Code) to the authors of 'original works of authorship,' including literary, dramatic, musical, artistic, and certain other intellectual works." [http://www.copyright.gov/circs/circ1.html #wci]
2. Assume that what you find on a Web site is copyrighted, unless it states otherwise or you know otherwise, for example, based on the age of the item. See the U.S. Copyright Office site below for details as to the time frames for copyrights. (Of considerable use for Web page creators is the fact that "Works by the U. S. Government are not eligible for U.S. copyright protection" [http://www.copyright.gov/circs/circ1.html# wwp]. You should still identify the source when quoting something from the site.)
3. The same basic rules that apply to using other printed material apply to using material you get from the Internet, the most important being: For any work you write for someone else to read, cite the sources you use.

For more information on copyright and the Internet, see the following sources.

United States Copyright Office

http://lcweb.loc.gov/copyright

The official U.S. Copyright Offices site, for getting copyright information (for the U.S.) directly from the horse's mouth. (For other countries, do a search for analogous sites.)

Copyright Web Site

http://www.benedict.com

This site is particularly good for addressing in laypersons' language the issues involved in the copyright of digital materials. It also provides background and discussion on some well-known legal cases on the topic.

Copyright and the Internet

http://mason.gmu.edu/~montecin/copyright-internet.htm

For someone creating a Web page, this site from George Mason University is an excellent example of a site (written mainly for a particular institution) that provides an excellent, realistic, readable set of guidelines regarding copyright and the Internet.

CITING INTERNET RESOURCES

The biggest problem with citing a source you find on the Internet is identifying the author, the publication date, and so forth. In many cases, they just aren't there or you have to really dig to find them. Basically, in citing Internet sources, you will just give as much of the typical citation information as you would for a printed source (author, title, publication, date, etc.), add the URL, and include a comment saying something like "Retrieved from the World Wide Web, October 15, 2003" or "Internet, accessed October 15, 2003." If your reader isn't particularly picky, just give the information about who wrote it, the title (of the Web page), a date of publication if you can find it, the URL, and when you found it on the Internet. If you are submitting a paper to a journal for publication, to a professor, or including it in a book, be more careful and follow whatever style guide is recommended. Fortunately, many style guides are available online. The following two sites provide links to popular style guides online.

Karla's Guide to Citation Style Guides

http://bailiwick.lib.uiowa.edu/journalism/cite.html

Karla Tonella provides links to over a dozen online style guides.

Style Sheets for Citing Internet & Electronic Resources

http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Style.html

This site provides a compilation of guidelines based on the following well-known style guides: MLA, Chicago, APA, CBE, and Turabian.

TIP:

On virtually every site, look for a site index and a search box. They are often more useful for navigating a site than by means of the graphics and links on its home page.

KEEPING UP-TO-DATE ON INTERNET RESOURCES AND TOOLS

For someone who wants to be alerted to the more valuable resources that become available, the following sites will be useful. Also, numerous specialized sites cover specific areas or tools (such as science or search engines) that will be mentioned throughout the following chapters. All of the sites listed below provide free e-mail alerting services and also provide archives of past content.

The Resource Shelf

http://resourceshelf.blogspot.com

This site by Gary Price provides extensive updates on new resources. He also produces a Weblog ("blog") newsletter that is extremely useful for being alerted to new sites, particularly those in the Invisible Web.

FreePint

http://www.freepint.com

A U.K.-based site by Will Hann providing:

- A free e-mail newsletter with tips on Internet searching and reviews of Web sites.
- FreePint Bar: Subscribers' Internet-related questions and comment—and reviews.
- FreePint Portal—particularly good for business information.

ResearchBuzz

http://www.researchbuzz.com

A site by Tara Calishain covering news on a broad spectrum of Internet research tools and providing articles, archives, and a weekly newsletter.

Internet Resources Newsletter

http://www.hw.ac.uk/libwww/irn

Produced by the Heriot Watt University Library. "The free, monthly newsletter for academics, students, engineers, scientists and social scientists."

The Scout Report

http://scout.wisc.edu

The Scout Report, published since 1994, provides well-annotated reviews of new sites, with both a weekly general report and also specialized mailing lists in the areas of life sciences, physical sciences, mathematicals, engineering, and technology.

GENERAL WEB DIRECTORIES AND PORTALS

General Web directories are Web sites that selectively catalog and categorize the broad range of sites available on the Web, usually including only sites that are likely to be of interest to a large number of users. Although they have quite a bit in common with Web search engines, general Web directories, such as Yahoo!, also differ tremendously from search engines. This chapter addresses where these tools fit, when they should be used, and when they should not be used.

General Web directories serve unique research purposes and in some cases may be the best starting point, even though their databases may include less than 1 percent of what search engine databases cover. The chapter looks at their strengths, their weaknesses, and their special characteristics. Many of these types of sites put their directory in the context of a "portal," or "gateway," providing a selection of other tools and information on the same page as the directory, so the portal concept and function also are addressed here, but separately in the second half of the chapter.

STRENGTHS AND WEAKNESSES OF GENERAL WEB DIRECTORIES

Strengths

- ✔ Selective
- ✔ Classified (categorized)
- ✔ Easily browsed
- ✔ Good for general questions
- ✔ Most have some searchability

Weaknesses

- ✔ Relatively small database compared to Web search engines
- ✔ May not have sites addressing very specific topics
- ✔ Typically less search functionality than most search engines
- ✔ Paid inclusion may affect quality
- ✔ Tend to index only the main pages of sites

SELECTIVITY OF GENERAL WEB DIRECTORIES

The two most distinguishing characteristics of these tools (especially in contrast to Web search engines) are their *selectivity* and their *classification* (categorization) of sites. By selectivity, we mean that each site included in the directory is reviewed by a human being and included on the basis of some measure of quality. The underlying characteristic generally looked for is that the site must contain significant content and the content should be of interest to a fairly large number of people. Impinging on these decisions of the directory's editors is the issue of paid inclusion. For some directories, inclusion of a site is influenced by the payment of a fee. In the case of Yahoo!, a Web site owner technically does not have to pay to be included, but may pay to get a site considered for inclusion. Chances of being included in a timely fashion is greatly increased by the payment of a fee. Overture (not included in this chapter) is largely based on pay for placement. In contrast, Open Directory does not accept any fees for inclusion.

A third characteristic of these tools is that typically only the main page of a site is indexed (in contrast to Web search engines, which may index all pages of a site). Web sites rather than Web pages are what is included in general Web directories. One impact of this distinction is that if a term is not on the main page of a site, a directory will probably not identify that site as relevant for your search. Furthermore, directories are less likely to index every word even on the main page and may list and search only a brief description.

CLASSIFICATION OF SITES IN GENERAL WEB DIRECTORIES

General Web directories typically organize sites into a dozen or so broad categories, with each of those categories broken down into additional levels of hierarchy. This categorization can be the most important reason to go to a directory. It allows browsing down through the levels of the classification hierarchy and can provide valuable direction for a searcher who is not quite sure how to narrow down a broad topic.

Different directories use different classification schemes, which may influence a user to choose one over another. Yahoo! and Open Directory, for example both have a business category. LookSmart, which is much more consumer-oriented, does not have a business category. That doesn't mean that

you won't find business sites there, but it does mean you may have difficulty in finding them by browsing the categories.

Searchability of General Web Directories

All major directories have a search box on their main page, which causes confusion with Web search engines. (Technically, almost any Web site that has a search box does indeed have a search engine behind it, but that's not what is generally meant by Web search engine.) By entering a term in a directory's search box, you may be searching the directory's database. This brings us to two issues: (1) how big the database is and (2) how much search functionality is offered.

Size of Web Directory Databases

Whereas major Web search engines can contain as many as a few *billion* records (Web pages), directories have typically fewer than 4 *million* records (sites). This is a case of good news and bad news. Good because it is reflective of the high degree of selectivity, bad because you are missing out on the vast majority of Web content that is out there. Yahoo! is actually somewhat of a hybrid, because its search box leads to a search not of its selective directory database, but of a crawler-generated Web database. LookSmart does something similar using the Inktomi database as its backup.

Search Functionality in Web Directory Databases

Directories provide considerably less than search engines in the way of search functions. The major Web directories automatically AND all of the terms you enter. Some may allow you to use quotation marks to search for phrases, and allow you to use a minus sign to NOT a term. Yahoo! provides significant search features for its Web database, but not for its directory database. Remember that the main thrust of these tools is browsing, not searching.

When to Use a General Web Directory

When all of these factors are put together, they point to a couple of fairly obvious situations in which starting with a directory is your best bet:

1. For a general question, in other words, when you don't have something very specific in mind, a general Web Directory is the place to go. You're headed to Tblisi for the first time and you just want to look around on the Web to see what kinds of information are available about the city. What defines "general" vs. "specific"? As a rule of thumb you might think in terms of the number of concepts involved. One or two concepts such as "Tblisi" or "Tblisi museums" is fairly general, and you might want to head for a directory rather than a search engine. Three concepts is getting more specific than a general directory is able to support, for example, "Tblisi art museums." In addition, if a single term itself is very specific, such as "cyclopentanecarbaldehyde," don't count on a directory.
2. (This is basically a corollary of the previous point.) Start with a general Web directory when you know you need to get more specific than what you have in mind at the moment and you need to browse to help you narrow your search.

THE MAJOR GENERAL WEB DIRECTORIES

Three very large U.S.-based general Web directories, a number of large non-U.S. directories, and some U.S. directories that are smaller and more selective but not subject-specific make up the major general Web directories category. We'll look at the three largest, some additional representative sites, and point you toward sites that provide lists of others. Specialized directories that focus on particular subject areas will be dealt with in the Chapter 3.

Yahoo!

http://yahoo.com

Yahoo! is the best known of the general Web directories, although its own directory is probably smaller than either Open Directory or LookSmart. Its content is well organized, and in addition to the directory itself, it has an excellent collection of tools that may be more important for the serious searcher than the directory itself. These tools include a very personalizable portal aspect, country-specific versions of the directory (and portal), groups, free e-mail, a calendar, and channels on topics such as travel and health. (See the last part of the chapter for more information on the portal aspect.)

Browsing Yahoo!

Yahoo! has categorized the sites in its directory into 12 major categories found on the home page, each typically with from three to six sublevels, for example:

Home > Science > Mathematics > Geometry >
Computational Geometry > Trigonometry

A fairly full understanding of the capabilities Yahoo! provides when browsing can be gotten by a close examination of a directory page. In Figure 2.1, you will see a page that resulted from clicking on the Social Science category, and with that, on Anthropology and Archaeology.

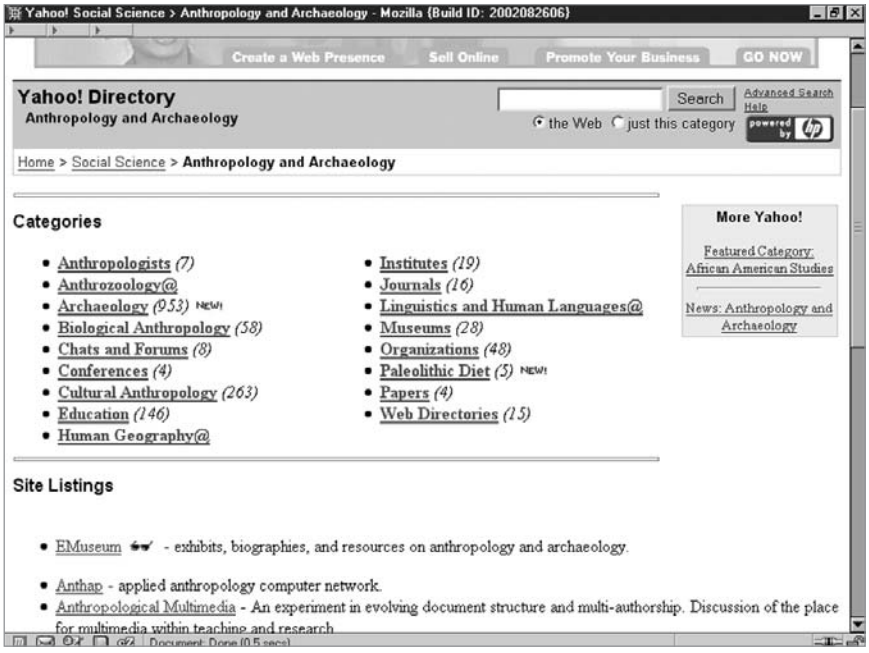


Figure 2.1

Yahoo! Directory Page

Note the following points:

1. There is a search box that allows you to search the Web or just within the current category. More on searching "all of Yahoo!" in a few paragraphs, but here notice that the "just this category" choice provides an extremely powerful tool. If you are looking for "graphics" sites, from the "graphic arts" side rather than from the computer and Web side, you might start by browsing from "Arts and Humanities" to "Design Arts" to "Graphic

Design." At that point, because over 1,000 Yahoo! listings still remain, you might use the search box to search just in the current category. In that way, you can avoid bumping into many sites that may be irrelevant.

2. Toward the top of the pages, Yahoo! reminds you where you are in the directory (Home > Social Science > Anthropology and Archaeology). The preceding levels are clickable here, allowing you to go back up one or more steps.
3. The Categories section of the pages shows what additional subcategories are available and how many listings are in each. The @ sign indicates that this is a cross-reference for a category primarily found elsewhere in the hierarchy. In this example, if you click on Anthrozoology, you will be taken to a page from the Biology category.
4. "Site Listings" lists the sites classified at this current level of specificity. Clicking on them will take you to the actual site. In some cases, they are broken down by "Most Popular" and an alphabetical listing. "Sponsor Listings" found here are ads.
5. (Not shown in this figure.) "Inside Yahoo!" listings takes you to potentially relevant Yahoo! resources such as "News" and various channels such as "Finance and Health."

Searching Yahoo!

Using the Yahoo! search box provides the user with a very different situation than is encountered when browsing through the categories. Until 2001, when you entered a term in the search box, the search was performed on the Yahoo! directory database of around 2 million sites. In 2001 Yahoo! began supplementing that with a search of a crawler-created Web database (Google's, with over 2 billion items), and then moved to having search results provided primarily by such a database. In addition, Yahoo! enhances search results with links to matching categories from the directory database, and "Inside Yahoo" links to supplementary resources such as the World Factbook and the Concise Britannica. In the context of this discussion of directories, be aware that most of the results shown on Yahoo! search results pages do not reflect the main advantage provided by directories, that of "selectivity." Most importantly, when using Yahoo!, remember that browsing the categories provides access to the smaller, but editorially selected, collection of sites. Yahoo's search box provides access to a much larger, but nonselective, Web database. Look for Yahoo!

to continue to shift its focus from the selective directory function to more general and extensive Web searching.

As well as having its search box provide access to more of a "search engine" size database, Yahoo! has also enhanced the searchability provided from within the search box. As has been true for some time, all terms you enter are automatically ANDed. You can also use quotation marks to specify a phrase, a minus sign in front of a term to eliminate pages containing that term, and now an OR (in capital letters) between terms to "OR" them. (See page 66 for an overview of Boolean AND, OR, and NOT.)

Yahoo!'s Advanced Search Page

Yahoo!'s Advanced Search page also moves Yahoo! more in the direction of "search engine" than "directory." To get to Yahoo!'s Advanced Search page, click on the Advanced Search link to the right of the search box. On the resulting page you will find options that allow simple Boolean ("all the words," "any of the words," "none of the words"), and options for limiting retrieval to title words, URL, date, domain, country, and language. Using the advanced search page you can also apply an adult content filter.

Yahoo! Search Results Pages

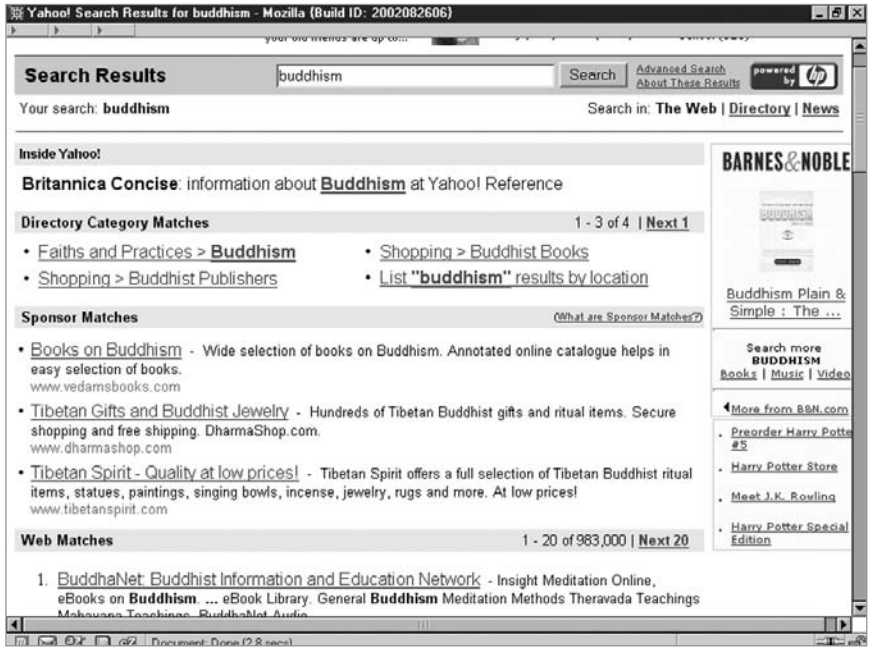
As with the browsing function, a good idea of the potential of Yahoo!'s search function can be gotten by looking at a search results page. When you do a search using Yahoo!, you will get results of the type shown in Figure 2.2

Note the following:

1. "Inside Yahoo!" gives links to other resources that Yahoo! automatically searches when you do any search. Depending upon the search, you will find links to reference resources such as The Britannica Concise and the World Factbook. Sometimes a list of matching headlines from Yahoo! News will be shown here.
2. "Category matches" are headings from Yahoo!'s directory that contain your term. Clicking on these will take you to that directory page. If "More" appears here, click on it to get a complete list of matching directory categories. This little section is where you can take advantage of selectivity and easily focus in on a specific aspect of your topic.
3. The Directory, News, Yellow Pages, and Images links shown near the top of the search results page is a menu that will lead you, respectively, to a list of matching categories and sites from the directory, matches from

Figure 2.2

Figure 2.2



Yahoo! Search Results Page

Yahoo!'s news collection, Yahoo!'s Yellow Pages, and images from an image collection.

4. You will also often see a list of "Related topics" near the top of the page. Clicking on one of these will take you to a list of sites from that category of Yahoo!'s directory.

Other Yahoo!s

Country-Specific

At the bottom of Yahoo!'s home page there are links to special versions of Yahoo! for over 20 countries (plus the U.S. site in Chinese and Spanish). These versions provide more localized content and should be considered if you are searching either from one of those countries or in detail about any of those countries.

City Yahoo!s

Likewise at the bottom of the main page, there are links to Yahoo! sites with extended coverage for over 200 U.S. cities. They all follow a similar pattern containing local phone directories, real estate listings, maps, and so on.

Yahooligans

Yahooligans is the very popular version of Yahoo! built for kids ages 7 to 12. The directory portion of the site contains age- and content-appropriate sites

and there are a number of other references and other features of use at home and in the classroom.

Open Directory

http://dmoz.org

Open Directory is the largest of the general Web directories (4 million sites) and differs from Yahoo! in several significant ways: (1) Instead of paid editors, Open Directory uses volunteers (over 50,000 of them); (2) it is pure "directory" and does not position itself as a portal and has no portal features; (3) it is used by other sites, most notably by Google. Google's implementation of Open Directory is different enough that it is treated separately later.

Browsing Open Directory

Open Directory divides its sites into 16 top-level categories, and each of these is further categorized by several additional levels, such as:

Top: Society: Government: Finance: Central Banks: Supranational

A look at an example of a directory page, (as with Yahoo!), can identify some of Open Directory's most important aspects.

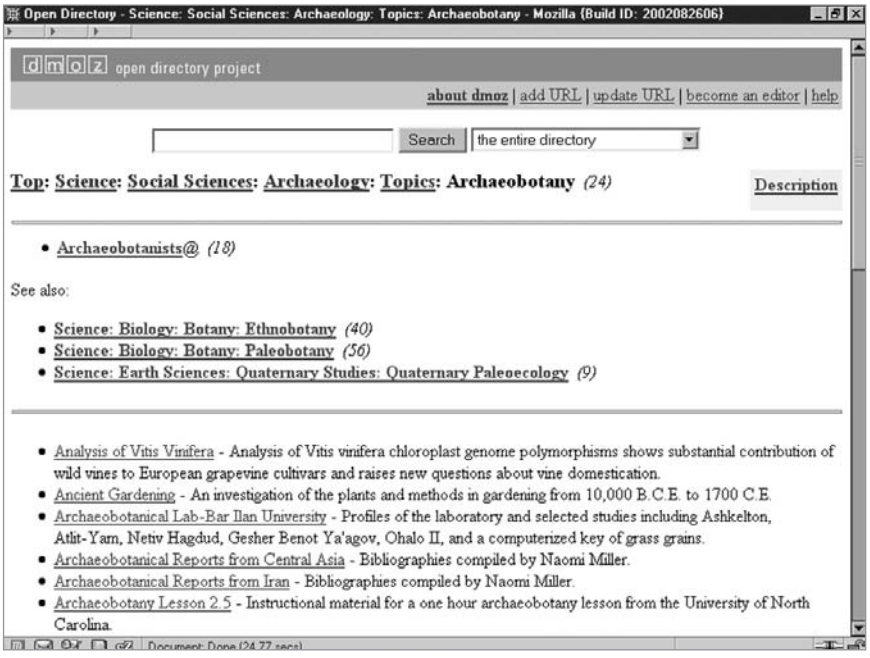


Figure 2.3

Open Directory Directory Page

The most significant features here are:

1. A search box that gives the option of searching the entire directory or just the current category.
2. A reminder, under the search box, of where you are in the subject hierarchy, each section being clickable, allowing you to move back up the hierarchy easily.
3. The subject hierarchy is followed by a list of the subcategories and usually a "See also" list of categories. The latter points to other sections in the Open Directory, as does the @ sign that occurs after some of the subcategories.
4. If the directory database contains articles on this topic in languages other than English, you will see a listing for "This category in other languages."
5. Following that will be the listings of the sites themselves, with brief annotations.
6. Unique to Open Directory is the "Descriptions" link in the upper right-hand corner of the page. Clicking on this will take you to a "scope note" defining what kinds of things are placed in this category.
7. (Not shown in the figure.) At the bottom of the pages are links to search engines and even to Yahoo!. Clicking the links will cause the name of the current category to be searched in these tools.

Searching Open Directory

The Open Directory database can be searched using either the search box found on the main page, at the top of directory pages, and at the bottom of search results pages. Search syntax is a bit more sophisticated than that offered by Yahoo!:

- Multiple terms are automatically ANDed. "Eastern Europe" will get only those items containing both terms (capitalization is ignored).
- The automatic AND can be overridden by use of an OR (capitalization not required). For example: cycling OR bicycling.
- You can specify a phrase using quotation marks, e.g., "Native American."
- A minus sign or "andnot" will exclude a term, e.g., "vienna -virginia" will eliminate records containing the term "virginia" from the listing of Web sites (but not from categories).
- Prefixes can be used to limit results to records that have a particular term in the title, URL, or descriptions. For example: t:austria, u:cam, or u:cam.ac.uk.

- You can use right-hand truncation. german* will retrieve german, germany, germanic.
- Various combinations of these functions can be used in combination. However, if you are looking for that degree of specificity, consider using a search engine instead of a directory.

Primarily because of the lack of related portal features, Open Directory search results pages are much simpler than Yahoo!'s (see Figure 2.4).

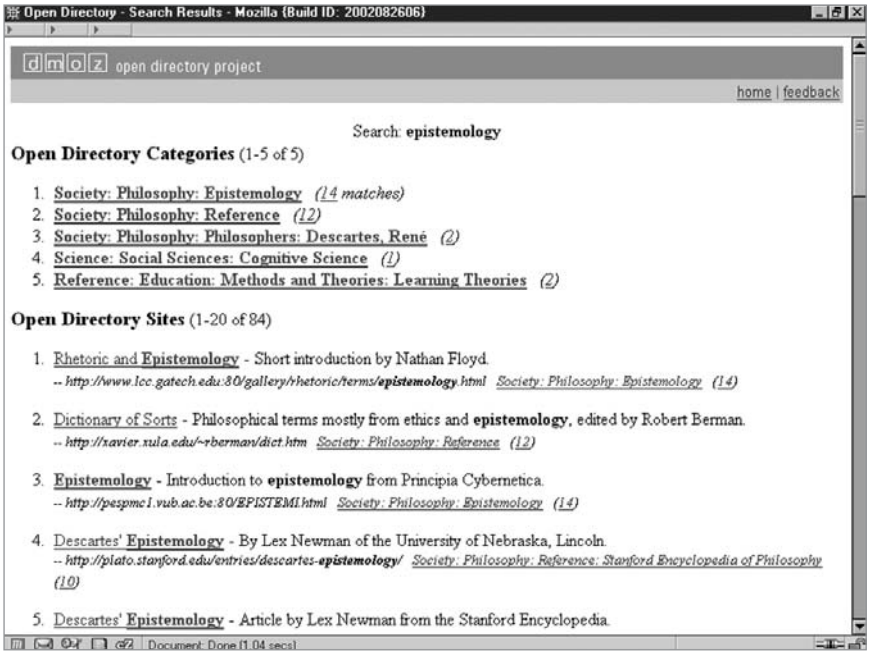


Figure 2.4

Open Directory Search Results Page

Open Directory search results pages contain the following details:

- Category headings containing the term you searched for or that were identified through the Web sites identified by the search. The number of sites in the category is also shown.
- Sites where the title of the site or the annotation contained your term(s). The category in which the term occurred is also shown and is clickable to take you to that category.
- As when browsing through categories, links to search engines are given at the bottom of search results pages. Clicking on any of these links will

cause you to be switched to that engine, and your search will be executed there. Another Open Directory search box will also be found at the bottom of search results pages.

Open Directory's Advanced Search Page

The link to the Advanced Search page, found on Open Directory's main page beside the search box, takes you to a page where you can limit your search to a particular category, to "categories only" or "sites only," or to sites that fall in the categories of Kids and Teens, Kids, Teens, or Mature Teens.

Google's Implementation of Open Directory

For its Web directory (click the Directory tab on Google's home page), Google uses the Open Directory database. You will find that the layout of directory and results pages there are almost identical to the pages you see when using Open Directory at http://dmoz.org, with a couple of important exceptions.

1. Whereas the dmoz.org site ranks retrieved records by relevance ranking, Google's results are ranked by the same popularity-based approach as is the Google Web search.
2. Searching is done using the same syntax as for Google's Web search:
   - OR to "OR" terms
   - Quotation marks for phrases
   - -term to exclude a term

One very important aspect of the way Google uses Open Directory is that, at the same time a regular Web search is done in Google, a search on the Open Directory database is also done. Any matching Open Directory categories found are shown at the top of the regular Google Web page results and any matching Open Directory sites are integrated into the regular Google results.

LookSmart

http://looksmart.com

Although its database is not as large as that of Open Directory, LookSmart's database is still significantly larger than Yahoo!'s. As can be seen by a look at the main categories used, LookSmart has more of a consumer orientation (see Figure 2.5). Its categories have, however, come to look more and more like those of its two main directory competitors. LookSmart positions itself as a supplier of directories for other (portal) sites and LookSmart.com is largely a

demo site for potential customers. You will actually find the LookSmart directory to be the directory used by sites such as Microsoft's MSN, AltaVista, Netscape Netcenter, CNN, AskJeeves, and many other high-profile sites. Paid inclusion is central to LookSmart's business plan, but LookSmart also has a program of volunteer editors.

Browsing LookSmart

LookSmart arranges its content under 12 main categories. For each of those, several major subcategories are also shown on the home page, making it a bit easier to find your way to what you need. Each typically has from three to five sublevels of categories.

As you browse down through these categories, you will typically see the following on the directory pages:

1. A search box, with a pull-down window enabling you to search all of LookSmart or just within the current category
2. "Directory Categories"—Subcategories, including a line showing where you are in the hierarchy (with each previous level clickable)
3. "Directory Listings"—the actual sites from the LookSmart directory database

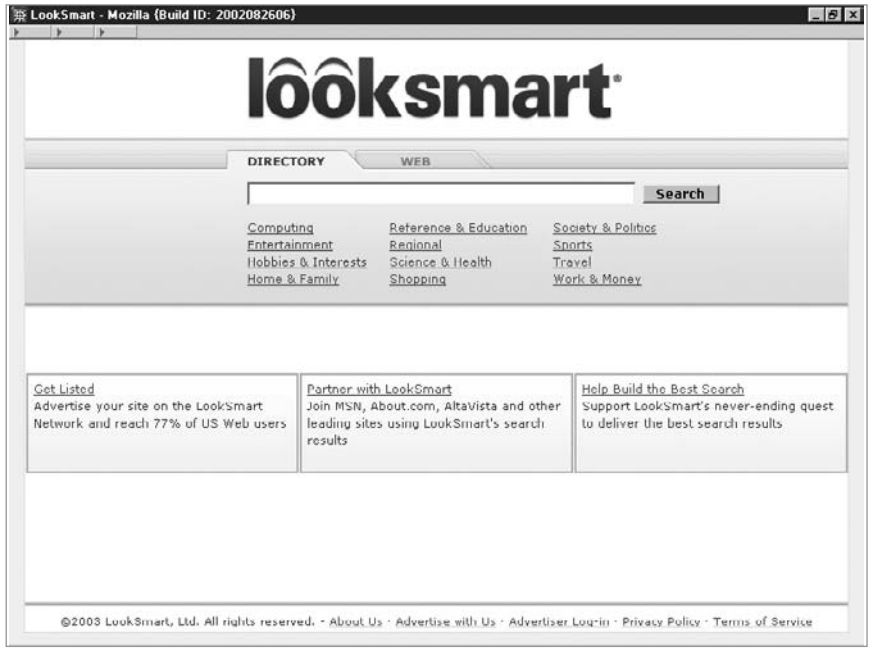
Searching LookSmart

LookSmart's home page (see Figure 2.5) has tabs for "Directory" and "Web," each providing a search box. The Directory search box allows a search of the selective ("reviewed") sites in LookSmart's own directory collection, while, like Yahoo!, the "Web" search searches a nonselective machine-created (crawler) database (in this case, WiseNut).

In either case, you will find that the first category of results listed is "Results from our sponsors," i.e., "paid listings." (See Figure 2.6.) If you searched from the directory tab, you will then find a listing of sites from LookSmart's directory collection. If you searched from the "Web" tab, you will find up to 300 listings from the WiseNut database.

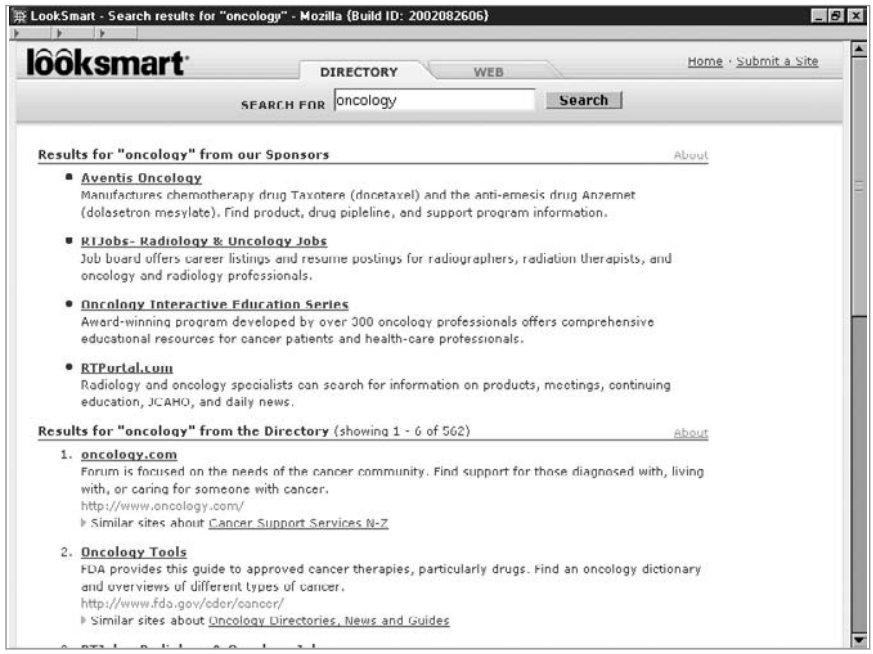
Search Features: LookSmart is the least searchable of the major directories. Terms are automatically ANDed, and you can use "-term" to exclude a term, but you cannot use quotation marks for phrases.

Figure 2.5



LookSmart Home Page

Figure 2.6



LookSmart Search Results Page

Other General Directories

Numerous other general Web directories are available, although none as large as the three just discussed. Most of the others specialize in some way, and the dividing line between general and specialized is a bit hazy. Many directories are general in regard to subjects covered, but specialized with regard to geographic coverage, such as the numerous country-specific directories. How to find them is covered later in this chapter. Those directories that are specialized by subject are covered in the next chapter. Here, though, we will look at one more directory that is general with regard to subject, but much more selective and, hence, much smaller: Librarians' Index to the Internet. Many others fall in this category, but this one is certainly among the best and is fairly representative of the genre.

Librarians' Index to the Internet

The highly respected Librarians' Index to the Internet (http://lii.org) is a collection of over 11,000 carefully chosen resources selected on the basis of their usefulness to public library users. Provided by the Library of California, it is well annotated, easily browsable, and also searchable.

Browsing Librarians' Index to the Internet

The contents of the site are broken down into 14 top-level categories, each usually has from one to three additional sublevels. The moderately lengthy annotations also provide links to the category in which they were placed, the date the annotation was created, and a link for users to comment on the site.

Searching Librarians' Index to the Internet

A search box appears on most pages. The search automatically ANDs your terms, but you can use an OR between terms and you can truncate using an asterisk (e.g., transport*). A spell-checker kicks in for terms that appear to be misspelled. An Advanced Search page allows you to search by the following fields: description, title, subject, author, publisher, URL, indexer initials, and category. Advanced Search also allows a Boolean AND, OR, and NOT, by use of pull-down windows, and here stemming (truncation) is automatic unless you check the "No Stemming" box.

Librarians' Index to the Internet also provides a free subscription to weekly e-mail updates on new sites added.

Where to Find Other General Directories

Unfortunately, most lists of searching tools do not adequately distinguish between search engines and directories and lump the two species together. Keeping that in mind, one place to go for a list of regional (continent or country-specific tools) is Search Engine Colossus at http://www.searchenginecolossus.com.

Most Important Things to Remember About Directories

1. Web Directories are most useful when you have a general rather than a specific question.
2. The content of directories is selected by humans, who evaluate the usefulness and appropriateness of sites considered for inclusion.
3. Directories *tend* to have one listing per Web site, rather than indexing individual pages.

GENERAL WEB PORTALS

Portals, or gateway sites, are sites that are designed to serve as *starting places* for getting to the most relevant material on the Web. They typically have a *variety of tools* (such as a search engine, directory, news, etc.) all on a single page designed so that a user can use that page as the "start page" for his or her browser. Portals are often personalizable regarding content and layout. Many serious searchers choose a portal, make it their start page, and personalize it. Thereafter, when they open their browser, they have in front of them such things as news headlines in their areas of interest, the weather for where they are or where they are headed, stock performance, and so on.

The portal concept goes considerably beyond the idea of general Web directories as we have been discussing them. However, this chapter seemed the appropriate place to discuss them for two reasons: (1) General Web directories (such as Yahoo! and the numerous sites that make use of Open Directory) are often presented in the context of a portal; (2) general portals embody the concept of getting the user quickly and easily to the most relevant Web resources.

In addition, when specialized directories are discussed in Chapter 3, we will see that their directory and portal natures meld so tightly that it is not feasible to try to separate them in that discussion. Hence, this chapter seemed the place to discuss general portals.

In addition to Yahoo!, well-known general portals include AOL, MSN (http://msn.com), Netscape (http://netscape.com), Lycos (http://lycos.com), Excite.com, and many others. For most countries there are popular general portals, for example, the French portal Voila! (http://www.voila.fr).

General portals usually exhibit three main characteristics: *a variety of generally useful tools*, positioning as a *start page*, and *personalizability*.

General Web Portals as Collections of Useful Tools

In line with the "gateway to Internet resources" idea, general portals provide a collection of tools and information that allows users to easily put their hands on information they frequently need.

Instead of having to go to different sites to get the news headlines and weather or to find a phone directory, general Web directory, search engine, and so forth, a portal puts this information—or a link to this information—right on your start page. General portals usually include some variety of the following on their main page:

- A general Web directory
- A Web search engine
- News
- Weather
- Stock information
- White pages
- Yellow pages
- Sports scores
- Free e-mail
- Maps/directions
- Shopping
- Horoscope
- Calendar
- Address book
- Chat, message boards, newsgroups

General Web Portals as Start Pages

Most general portals are designed to induce you to choose their site as your browser's start page. Because at least part of their support comes from ads, you will find a lot of those on the page, but the portal producer knows that the useful information must not be overpowered by ads or no one will come to the page. The overall thrust is to provide a collection of information so useful that it makes it worthwhile to go to that page first.

TIP:

To make a chosen page your browser's start page:

Internet Explorer: From the main menu bar: *Tools* > *Internet Options* > then, under the "General" tab, put the URL in the "Address" box.

Netscape: From the main menu bar: *Edit* > *Preferences* > then, under the "Navigator" section, put the URL in the "Home Page" box.

General Web Portals—Their Personalizability

Most successful general portals make their pages personalizable, allowing the user to choose which city's weather appears on the page, which stocks are shown, what categories of headlines are displayed, and so on. If you look around on the main pages of these sites, you will usually see either a "personalize" link or a link to a "My" option such as My Yahoo!, My Netscape, or My MSN that will allow you to sign up and personalize the page or take you to your personalized page if you have already done so. A sign-in link will do likewise.

Yahoo!'s Portal Features

A look at Yahoo! offers a good idea of the types of things most general portals can do. Yahoo! is undoubtedly one of the best of the general portals, particularly with regard to the personalization features. As a matter of fact, a case could be made that, for the serious searcher, Yahoo!'s personalized portal (My Yahoo!) is more important than the Yahoo! directory (and Yahoo!'s designers have now actually moved the directory categories rather far down on the home page).

Yahoo! has a number of portal features on its main, nonpersonalized page. Some of them, such as news headlines, are displayed directly on the page and links are provided to over 30 other portal features. Some of these links lead to a channel such as Autos, Real Estate, and Classifieds. "Channels," a term that has been used at various times by most portals, really refers to a more specialized portal page provided by the site with, again, a collection of tools and links specific to the topic of the channel. Other links on Yahoo!'s main page take you to a phone directory, maps, groups, and more. The best way to understand a portal such as Yahoo! is to lock yourself in your office and not leave until you have clicked on every link on the page. (Skip the ads, though.)

My Yahoo!

An example of a personalized general portal page (My Yahoo!) is shown in Figure 2.7. Yahoo! provides one of the most personalizable general portals, with possibly the widest variety of choices. It also provides personalized versions for most of its 24 country or language-specific versions.

My Yahoo! Personalized Portal Page

Figure 2.7

A Few of the Popular General Portals

The sites listed below all exhibit the three characteristics of general portals, to varying degrees and with varying content. Which of them is the best for any individual is probably dependent upon what content is available on the portal and how it is presented. Try more than one before deciding. Lycos has over 40 options you can place on the page, Yahoo! has over 60. Such items as "Word of the Day" and "Pregnancy Watch" may not necessarily be of interest to you. Your personal stock portfolio is handled very differently by various portals and what data the portal displays and how it displays them may make the difference in your choice. A portal may allow very detailed specification of what categories of headlines are displayed, or only very general categories, and so on. The ones listed below are among the best known in the U.S. For non-U.S. portals, take a look at the "World" section of Open Directory (dmoz.org/World), choose your country, and search for the term "portal" in the relevant language.

Selected Examples of Leading General Portals

Excite (http://excite.com)—Once the best, and might be on the way back.

Lycos (http://lycos.com)—Very good content and personalization, but ads take up too much of the space.

AOL—Mentioned here because it was the first popular general portal. Available only to AOL subscribers.

MSN (http://msn.com)—Widely used because it came pre-installed on so many computers. For those of you who can't get enough of Bill Gates, here's one more opportunity to have him around.

Netscape (http://netscape.com)—Very good content, very cleanly laid out, and very personalizable. (Netscape was acquired by AOL in 1999.)

Other Resources Relating to General Directories and Portals

Traffick: The Guide to Portals and Search Engines. Frequently Asked Questions about Portals.

http://www.traffick.com/article.asp?aID=9#what

This site provides an overview and history of the concept of Web portals.

SUMMARY

Remember that general Web directories provide sites that are evaluated and selected by human beings. This, along with the fact that all sites are placed in categories to allow browsing, makes these tools a good starting place when you want selected sites, when you want only a few sites, and when your question has a general rather than a specific nature.

This page intentionally left blank

SPECIALIZED DIRECTORIES

For some immediate expertise in Web resources on a specific topic, there is no better starting point than the right specialized directory, or portal. Also known as resource guides, metasites, cyberguides, Webliographies, or just plain collections of links, these sites bring together selected Internet resources on specific topics. They provide not just a good starting place for effectively utilizing Internet resources in a particular area, but also, very importantly, a confidence in knowing that no really important tools in that area are being missed. The variety of theses sites is endless. They can be discipline-oriented or industry-oriented; they may focus on a specific kind of document (e.g., newspapers or historical documents) or take virtually any other slant toward identifying a useful category of resources.

If the producer of the site adds to the collection of links some valuable content such as news headlines or lists of events, you have not just a specialized directory, but a specialized portal or gateway, making it even more useful as a starting point.

STRENGTHS AND WEAKNESSES VS. OTHER KINDS OF FINDING TOOLS

Strengths

- ✔ Specialized
- ✔ Very selective
- ✔ Provide some immediate "Web expertise"

Weaknesses

- ✔ Relatively small
- ✔ Variable quality and consistency
- ✔ Most are browsable but not searchable

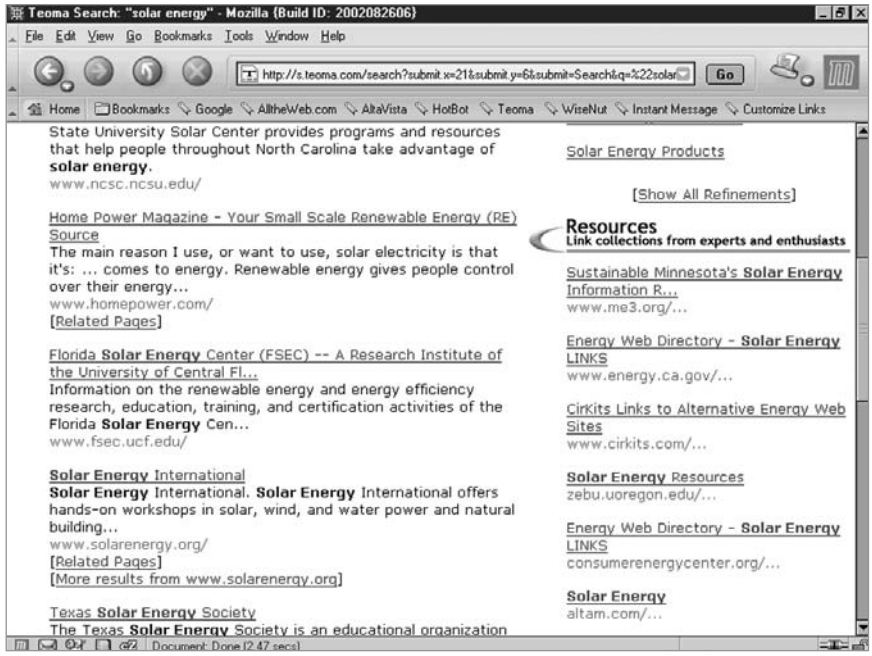
HOW TO FIND SPECIALIZED DIRECTORIES

There are several ways of systematically identifying a specialized directory for a particular area of interest. These include: Teoma's "Resources" results; Yahoo!'s "Web Directories" subcategory; searching for them in search engines, professional journal articles, and books; and directories of directories.

Finding Specialized Directories Using Teoma

The Teoma search engine (http://teoma.com) provides a unique section on its results pages that specifically identifies resource guides. Do a search on the topic for which you would like to find a specialized directory and look for the "Resources" section on the first page of Teoma's results. Among the sites that it finds, Teoma lists separately the sites that have a large number of links. Not all of the sites listed will truly be a specialized directory, but Teoma usually identifies several. You will notice that Teoma, wisely, identifies these as "Link collections from experts and enthusiasts," not guaranteeing the level of expertise involved (see Figure 3.1).

Figure 3.1



Resources Section of a Teoma Results Page (a Search on "Solar Energy")

Finding Specialized Directories by Using Yahoo!

Yahoo! lists thousands of specialized directories. As a matter of fact, it has lists of one or more specialized directories for almost 1,000 categories, from parasitology to sumo. The trick to finding them in Yahoo! is simple: Look for the Web Directories subcategory either by browsing down through the Yahoo! categories list or by putting your subject and the phrase "Web directories" in Yahoo!'s

Directories search box. (A similar thing can be done in Open Directory using the subcategory "Directories," but you will find significantly fewer results. Only around 200 categories in Open Directory list this subcategory.)

Finding Specialized Directories in Professional Publications

Keep an eye out for articles that discuss Internet resources for particular areas in professional publications (printed and online): journals such as *Online* and *Searcher* and Web sites for searchers such as FreePint (http://freepint.com). A book by Nora Paul and Margot Williams, *Great Scouts*: *CyberGuides to Subject Searching on the Web* (CyberAge Books, Medford, NJ, 1999) is specifically about specialized directories and lists over 500 such sites.

Finding Specialized Directories Using Directories of Directories

Directories of directories are valuable sources for locating topic-specific information. The following two sites contain collections of specialized directories (and may contain other content as well).

The WWW Virtual Library

http://www.vlib.org

Perhaps the best known catalog of Web directories, The WWW Virtual Library, started by none other than Tim Berners-Lee, founder of the Web, contains an excellent selection of specialized directories arranged by category. In one sense, it is one large directory with individual sections done by a large number of volunteers, but because the format of each section is also very independently done, WWW Virtual Library is indeed a collection of individual directories. The quality of the individual directories tends to be quite high.

Search Engine Guide

http://www.searchengineguide.com

Although this site does not adequately distinguish between search engines and directories, if you use the search box under the Search Engines link or browse the categories listed there, you will find a useful collection of specialized directories.

Finding Specialized Directories Using Search Engines

In addition to using Teoma's special "Resources" section, you may be successful in finding a specialized directory in your area by searching a term for your area AND the word "resources," for example, "biotechnology resources." Also try using "metasite" in addition to, or instead of, "resources." For industry portals, search for the industry plus the word "portal," for example, "nuclear industry portal." If you would like to get a site that provides a list of printed resources for a subject, as well as Internet resources, use the word "pathfinder." Many libraries provide pathfinders that are guides to the literature and to Internet resources in their library. Even if you don't have access to the library that produced it, the guide can provide reminders of printed tools you might want to track down.

WHAT TO LOOK FOR IN SPECIALIZED DIRECTORIES AND HOW THEY DIFFER

For many areas there are numerous directories to choose from. If you want to find the best, several factors must be considered. An excellent specialized directory does not have to be strong in all of these facets, but, depending on your need, you might want to focus on a few particular aspects. They tend to differ mainly in these terms:

- Size—Sometimes large is good, sometimes a smaller number of sites to focus on is good.
- Categorization/Classification—Especially if the number of sites included is large, it is helpful if they are divided by a useful categorization.
- Annotations—A large portion of specialized directories (including many very good ones) do not have annotations describing the sites they list. Annotations, though, can be very useful by providing a quick overview of what the sites cover and any special characteristics of the sites.
- Searchability—A fairly small portion of specialized directories provide a search box, to save having to browse. If the directory is large, this can be quite useful.
- Origin—Who (or what organization) produced the site is sometimes, but not always, a good indicator of the quality you might expect from the site. Unfortunately, many sites do not give a clear indication of who produced them, and you have to rely on the URL for a clue.

- Portal features—If, in addition to the collection of links, other features are included, the site can be especially powerful. Look for such things as news headlines, lists of events (conferences, etc.), professional directories (e.g., a list of members if it is a site produced by an association), directories of companies in that area, and so on.

SOME PROMINENT EXAMPLES OF SPECIALIZED DIRECTORIES

The following are chosen for a variety of reasons. Some are chosen because they are simply sites that most serious searchers should be aware of. Some are listed here because they demonstrate particularly good or unique characteristics of a specialized directory. Some are given because they are very wide-ranging (as well as having other values as a specialized directory). In some categories, such as Government, more than one is listed in order to provide contrasts between sites. (Sometimes multiple directories are listed for an area because I just could not make up my mind which one to choose.)

Don't forget that effective use of a directory approach to identifying relevant sites can mean using a combination of the general Web directories covered in the previous chapter and the specialized directories covered here. In one sense, each section of a general directory such as Yahoo! or Open Directory is itself a specialized directory.

General and Reference Tools

The first three sites listed here provide an extensive collection of links to reference tools such as encyclopedias, dictionaries, and so forth. These directories vary in terms of exhaustiveness and method of arrangement. Each is worth getting to know. The last two included in this section, Project Gutenberg and the Library of Congress Z39.50 Gateway, provide links to books and library catalogs available online.

Internet Public Library Reference Ready Reference

http://www.ipl.org/ref/RR

From the School of Information, University of Michigan, and created by librarians and students, this is a great collection of ready reference links, including almanacs, biographies, census data, dictionaries, encyclopedias, and other reference resources.

refdesk.com

http://refdesk.com

A fairly extensive collection, actually arranged more as a portal with news headlines, and other features, as well as links to valuable reference resources. (It was achieving a deserved status on its own, but got a boost when Colin Powell said something to the effect that it should be on the screen of every State Department employee.)

InfoMine

http://infomine.ucr.edu

A well organized, categorized, and searchable collection of over 40,000 links, this directory is specialized in that it focuses on "Academically Valuable Resources." Look here for sources that will be useful in an academic environment (all levels). For a specialized directory, the Advanced Search page has quite extensive searching capabilities. It comes from the University of California, with contributions from librarians at a number of other universities.

BUBL LINK

http://bubl.ac.uk/link

This site, from the University of Strathclyde, includes over 12,000 resources, covering all academic areas. Part of its uniqueness is that the categories used are based on the Dewey Decimal Classification, and it has a particularly strong focus on library and information science. It is very easily browsable and also has good search capabilities on its Search page.

Project Gutenberg

http://www.promo.net/pg

Want to read a good book? Come here. This is the site for a project that dates back to early years of the Internet and has the objective of making available to the world all books that are out of copyright and in full-text online. It leads to around 6,000 books, from Cicero to the Bobbsey twins. All are books that are no longer under copyright (therefore, almost all are from before 1923). For many of the books, the entire text is available in a single file, allowing a researcher to quickly find all mentions of a word in a text (by using the "Edit > Find in page" function of your browser). Using this approach (not just here but elsewhere) you can, for example, go to the text

of the Odyssey and quickly, one-by-one, find every mention of Telemachus, if you are inclined to do such things.

Library of Congress Gateway to Library Catalogs

http://lcweb.loc.gov/z3950/gateway.html

Going beyond just the collection of links level, this site brings together, using a consistent interface, the capability of searching (one at a time) the contents of the online catalogs of almost 500 libraries, both in the U.S. and elsewhere. All of these are catalogs that use the Z39.50 standard for online library catalogs.

Social Sciences and Humanities

Social Science Information Gateway

http://sosig.esrc.bris.ac.uk

This collection, aimed at students and researchers in the social sciences, actually consists of two collections: the SOSIG Internet Catalogue of thousands of carefully selected Internet resources and the Social Science Search Engine with a database of over 50,000 resources identified by crawlers (hence, less selective than the catalog). The catalog itself is searchable as well as browsable. The overall site is much more of a portal than just a specialized directory. The Grapevine section contains extensive listings of conferences, courses, events, departments, and CVs. If you register (it is free), you have added capabilities, such as free e-mail alerts of new sites, conferences, and more.

Tennessee Tech History Web Site

http://www2.tntech.edu/history

At first it looks like it's simply about history and about Tennessee Tech, but there's much more, with excellent large collections of resources for both history and historiography. Although anyone interested in history will find it valuable to browse most sections of this site, for many, the most profitable part may be under the heading Internet Resources in History, and under that, the sections History Sites by Subject and History Sites by Time Period.

Virtual Religion Index

http://religion.rutgers.edu/vri

With a focus on scholarly sites, this directory site contains extensive links on the world's major (and minor) religions, and on the academic study of religion and religious issues.

Physical and Life Sciences

At present, there does not seem to be a single broad-reaching directory for the sciences in general. Your best bet for focusing on a specific science may be to try the techniques mentioned earlier for finding specialized directories, or try the appropriate section on sites such as InfoMine. The following are some notable examples of science sites in some specific areas.

ChemDex

http://www.chemdex.org

This site contains over 7,000 chemistry-related links. The links are arranged by 13 top-level categories and include both scholarly sites and links to chemical companies and suppliers. Go to "WebElements" for an outstanding online periodic table. Even if you have no connection with chemistry, you will find it interesting and even fun, with contents ranging from the usual periodic table data for each element, to bond enthalpies, to cartoons about the element.

HealthFinder

http://www.healthfinder.gov

As its subtitle says, "your guide to reliable health information," this consumer-oriented site comes from the U.S. Department of Health and Human Services. The links it includes range from medical dictionaries to background on diseases to directories of physicians, hospitals, nursing homes, and a variety of other easily understandable resources.

MEDLINE Plus Health Topics

http://www.nlm.nih.gov/medlineplus/healthtopics.html

A combination of information provided directly on the site and extensive collections of links, a good sense of what the site provides can be gotten by looking at the categories into which it is arranged: Health Topics (over 570 topics on conditions, diseases, and wellness), Drug Information, Medical

Encyclopedia, Dictionaries, News, (health news from the past 30 days), Directories (doctors, dentists, and hospitals), and Other Resources.

Engineering

EEVL: The Internet Guide to Engineering, Mathematics, and Computing
http://www.eevl.ac.uk

The EEVL site, based at the Heriot Watt University in Edinburgh, U.K., is undoubtedly one of the best specialized directories on the Internet. It contains over 9,000 links on the topics defined in its title and the well-annotated links are easily browsed using the detailed categories provided. The "Search All," "Key Sites," "EEVL Catalogue," and "Web Sites" tabs shown on the main page provide easy and quite extensive searchability. Sites are well-annotated and the main page also provides links to news and events in the areas covered, plus a variety of other resources. ("EEVL" is now the acronym for Enhanced and Evaluated Virtual Library.)

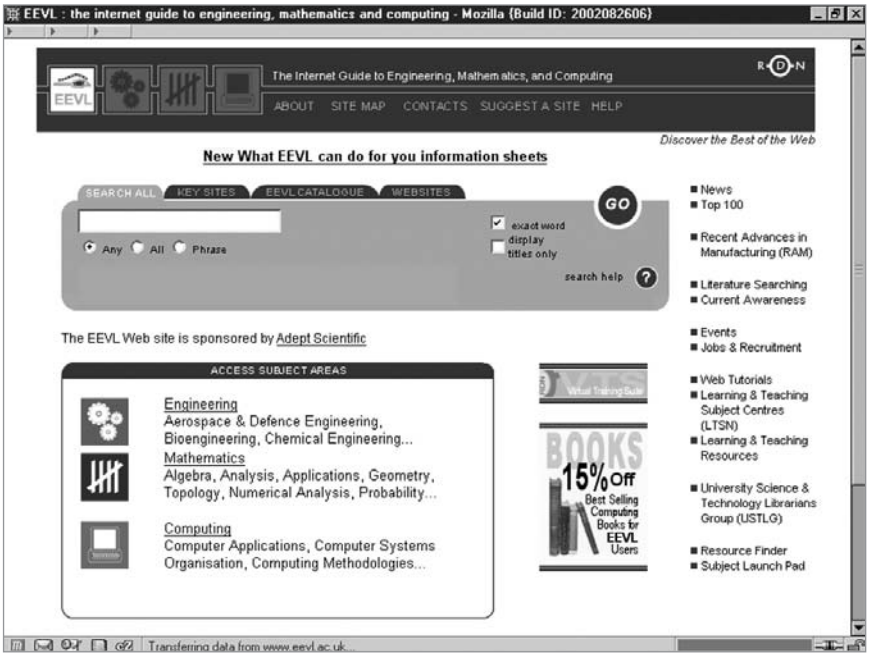


Figure 3.2

EEVL: The Internet Guide to Engineering, Mathematics, and Computing

Business and Economics

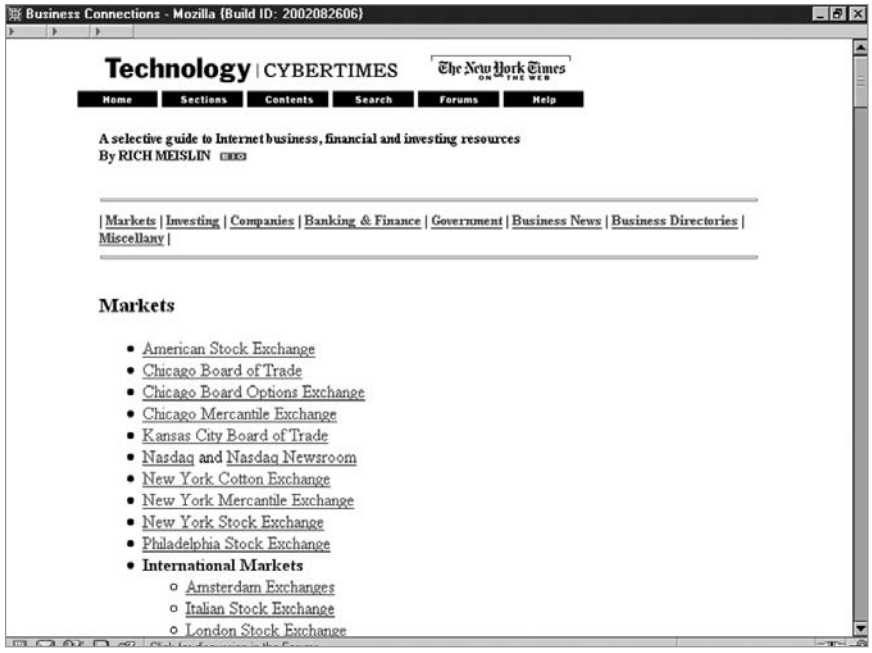
In addition to the specialized directories listed here, for business-related information, be sure to look at the sites listed in Chapter 6 (Reference Shelf) for company information. Some of the sites listed there, such as Corporate-information.com, can also be considered specialized directories.

New York Times Cybertimes—A Selective Guide to Internet Business, Financial, and Investing Resources

http://www.nytimes.com/library/cyber/reference/busconn.html

This bare-bones collection of business-related links provides categories for Markets (Stock, Options, etc.), Investing, Companies (directories, news, etc.), Banking & Finance, Government (Federal Reserve, IRS, BLS, etc.), Business News, Business Directories, and so on. Only about half of the 200 or so sites it includes are annotated (and just briefly), but the clarity, selectivity, and categories into which they are divided make it an easy and quick guide to critical business resources.

Figure 3.3



New York Times Cybertimes—Business, Financial, and Investing Resources

CEOExpress

http://ceoexpress.com

CEOExpress is a cluttered looking but rich site with a strong emphasis on business news sites. To get a good understanding of what it can provide, spend three or four minutes browsing the somewhat unique categories into which the links are arranged.

Virtual International Business and Economic Sources

http://libweb.uncc.edu/ref-bus/vibehome.htm

Divided into "Comprehensive," "Regional," and "National" sites, the 1,600 links on this site emphasize "full-text files of recent articles and research reports," "statistical tables and graphs," and other business-related directories.

Resources for Economists on the Internet

http://rfe.wustl.edu

Edited by Bill Goff and sponsored by the American Economic Association, this site lists over 1,300 categorized into 93 sections. These sections range from the obvious things of interest to economists, such as data, to less obvious but very useful categories, such as software and mailing lists.

WebEc

http://www.helsinki.fi/WebEc

A member, so to speak, of the World Wide Web Virtual Library, WebEc is edited by Lauri Saarinen and sponsored by two Finnish organizations, the Center for Innovative Education and the Helsinki School of Economics. WebEc covers an extensive range of economics sites, provides good annotations, is categorized for easy browsing, and also provides a search capability for searching descriptions and keywords for the sites covered.

I^3—Internet Intelligence Index

http://www.fuld.com/i3

Produced by Fuld and Company, a leader in the competitive intelligence field, this directory provides well organized and annotated links to over 600 sites that competitive intelligence researchers should be aware of.

Government and Governments

Although some countries have single sites that provide links to sites for individual departments or ministries, many do not, and it is not always easy to identify the particular agency site you need. The following directories make this much easier by bringing together large collections of sites by country or other category. For the main site for any U.S. state, use the following "recipe": http://www.state.pc.us, where pc is the two-letter postal code for the state, e.g., http://www.state.md.us.

Governments on the WWW

http://www.gksoft.com/govt

Although a bit slow in updating, this site contains links to over 17,000 Web sites from governments (and multinational organizations) around the world, including sites for parliaments, law courts, embassies, cities, public broadcasting corporations, central banks, political parties, and the like. There are no annotations, but the names of the sites are translated into English.

Foreign Government Resources on the Web

http://www.lib.umich.edu/govdocs/foreign.html

Whereas the preceding site provides access by country, this site provides both a country index and a subject index, the latter with over 30 headings, such as anthems, decolonization, economics, human rights, and so on. There are fewer sites for each country, but annotations are provided for the sites that are included.

FirstGov

http://firstgov.gov

This site is the official portal to U.S. government sites and also contains links to state sites. The Agencies link will take you to links arranged by branch of government, and the main divisions (Citizens, Business, Federal Employees, and Government-to-Government) allow browsing by type of information sought.

UK Online

http://www.open.gov.uk

This is the official U.K. government portal site and provides links to U.K. public sector information. The Quick Find links are particularly useful and provide an alphabetic index by subject for central and local government resources.

Political Resources on the Net

http://www.politicalresources.net

This is an excellent resource for quickly identifying the sites for political parties of any country. On the map on the home page, click on a continent, then the country. Links for international parties and other related resources are also provided.

Legal

FindLaw

http://www.findlaw.com

This very rich portal contains links to a broad range of legal subjects from lawyers and law firms to cases and codes. Don't expect it to turn you into an expert legal researcher, but if you are one, you are probably already making good use of this site. If you aren't one, it will point you in the right direction for the best legal resources on the Internet.

Education

Kathy Schrock's Guide for Educators

http://school.discovery.com/schrockguide

This well-known directory for K-12 teachers and parents contains hundreds of sites, each with a brief annotation. You can browse by subject or you can search (either the whole site or the parents or teachers areas). Among other things, it is a good source for links to lesson plans.

Education World

http://education-world.com

Education World contains a searchable database of over 500,000 sites related to education. The site itself is more portal than merely a directory and contains much original content by the producers of the site (such as articles and lesson plans) as well as the links to other sites.

Education Index

http://www.educationindex.com

Education Index contains over 3,000 sites, with annotations, arranged in 66 categories. You can browse either by subject area or by "Lifestage" The "Coffee Shop" section is a collection of online discussion groups.

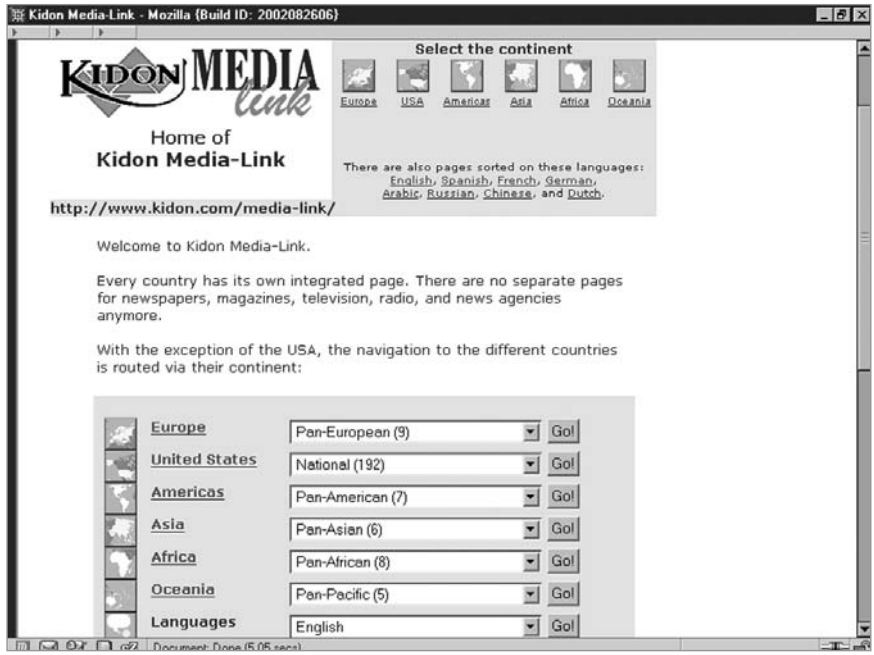
News

Kidon Media-Link

http://www.kidon.com/media-link

Although a number of sites serve as directories of newspapers and other news sources on the Internet, Kidon Media-Link is one of the most extensive and seems to have relatively few dead links, a problem with some of the other news directories. The site is arranged by continent, then country, and provides links to newspapers, news agencies, magazines, radio, and TV sites.

Figure 3.4



Kidon Media Link

Genealogy

Cyndi's List of Genealogy Sites on the Internet

http://www.cyndislist.com

This is perhaps the best known of the numerous genealogy directories with links to over 205,000 sites. You can browse through the 150 categories or take advantage of the search box. Either the beginner or the experienced genealogist should find it useful.

SEARCH ENGINES

General Web search engines, such as AltaVista, AllTheWeb, and Google, stand in contrast to Web directories in three primary ways: (1) They are much larger, containing over a billion, instead of a few million (or fewer) records, (2) there is virtually no human selectivity involved in determining what Web pages get included in the search engine's database, and (3) they are designed for searching (responding to a user's specific query) rather than for browsing, and, therefore, provide much more substantial searching capabilities than do directories.

For someone using Internet resources, a workable definition of a Web search engine is that it is a service on the Web that allows searching of a large database of Web pages by word, phrase, and other criteria. There is actually some ambiguity involved when one speaks of "search engines." From a slightly more technical perspective, when we use a site such as AltaVista, we are utilizing a "service" that facilitates searching of a database. In the narrower sense, the "search engine" is the program utilized by the service to query the database. Almost any site that provides a search box could be considered to have a search "engine." Here, when we speak of "search engines," we will really be referring to a service, such as the three just mentioned, that provides searching of a very large database of Web pages and may provide other services as well, such as translations, shopping, and others.

HOW SEARCH ENGINES ARE PUT TOGETHER

To fully take advantage of search engines, it is useful to understand the basics of how they are put together. Four major steps are involved in making Web pages available for searching by a search engine service. These steps also correspond to the "parts" of a search engine and are: the spiders, the indexing program and index, the search engine program, and the HTML user interface.

1. Spiders (a.k.a. crawlers). These are programs used by the search engine services to scan the Internet to identify new sites or sites that have

changed, gather information from those sites, and feed that information to the search engine's indexing mechanism. For some engines, more popular sites (such as those that have lots of links to them) are crawled more thoroughly and more frequently than less popular sites.

Tied into this crawling function is a second way for Web pages to get identified—by the process of submitted URLs. By means of a link on most search engines' home pages, anyone can submit a URL, and with the exception of those pages that are identifiable as "spam" (pages that are designed to mislead the search engine and search engine users and/or illegitimately lead to high rankings) or pages that are unacceptable for other reasons, the pages will get indexed and added to the database.

2. The indexing program and the index. Once a new page is identified by the search engine's crawler, the page will typically be indexed under virtually every word on the page. Other parts of the page may also be indexed, parts such as the URL, metatags, the URLs of links on the page, and image filenames.
3. The search "engine" itself. This is the program that identifies (retrieves) those pages in the database that match the criteria indicated by a user's query. Another important and more challenging process is also involved, that of determining the order in which the retrieved records should be displayed. The relevance-ranking algorithm may take a number of factors into account, such as the popularity of the page (as measured by how many other pages link to it), the number of times the search terms occur in the page, the relative proximity of search terms in the page, the location of search terms (for example, pages where the search terms occur in the title of the page may get a higher ranking), and other factors.
4. The HTML-based (HyperText Markup Language) interface that gathers query data from the user (the "search page"). The home page of the search service and advanced search pages are the parts we usually envision when we think of a particular search engine. These pages contain the search box(es), other search qualifiers, links to the various databases that are searchable (images, news, etc.), and perhaps a number of other features.

HOW SEARCH OPTIONS ARE PRESENTED

Exactly what search options are available varies from search engine to search engine. In any particular search engine, some available options use the

features shown on the home page, but on the advanced search page, usually several more options are clearly displayed. Options are typically made available in one of two ways: (1) by means of a menu or (2) by the searcher directly qualifying the term when it is entered in the main search box.

An example of the menu approach is shown in Figure 4.1, where (in AllTheWeb) a pull-down menu allows the term entered in the box to be qualified. In this example, the search is requesting that only those pages be retrieved that have the term "antioxidants" in the title of the page.

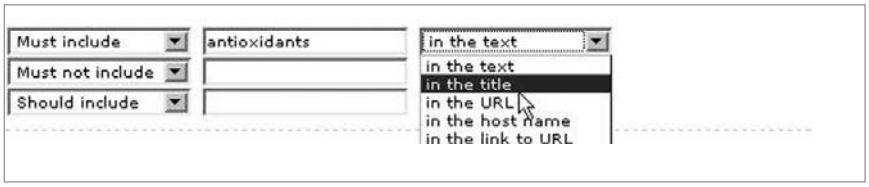


Figure 4.1

Example of the Menu Approach to Qualifying a Search Term

Figure 4.2 shows an example of qualifying a term directly. Here (in AltaVista) the "title:" prefix is inserted to accomplish the same thing as in the menu example in the previous figure.

Figure 4.2

Example of Using a Prefix to Qualify a Term

Usually you have a choice as to which approach to use. The menu approach is easier in that you do not need to know the somewhat cryptic prefixes. If you do know the prefixes, you can sometimes more quickly and easily accomplish your search.

TYPICAL SEARCH OPTIONS

A number of search options are fairly typical. These include phrase searching, language specification, and specifying that you retrieve only pages where your term appears in a particular part (field) of the record such as the title, URL, or links. Date searching is also common. Now that major engines include more than just HTML pages, for those engines, you can also specify file type

(Web pages, PDF files, Excel files, etc.). Every engine also offers some form of Boolean operations.

The following paragraphs give a quick look at why you might want to use (or not use) those options. The chart at the end of this chapter (Table 4.2 beginning on page 112) identifies which options are available in which engines, and the profiles that follow provide some details for using the search options in each engine. Expect some changes in exactly which options are offered by which engines.

Phrase Searching

Phrase searching is an option that is available in every search engine, and perhaps surprisingly, can be done the same way in all of them. To search for a phrase, put the phrase in quotation marks. For example, searching on "Red River" (with the quotation marks) will assure that you get only those pages that contain the word "red" immediately in front of the term "river." You will avoid records such as one about the red wolves of Alligator River. When your concept is best expressed as a phrase, be sure to use the quotation marks. You are not limited to two words, but can use several. For example, to find out who said "When I'm good I'm very good, but when I'm bad I'm better," search for a few of the words together, such as "when I'm bad I'm better." (Search engines have limits on the number of words you can enter.)

Some engines automatically identify common phrases and most engines give a higher ranking to pages that have your terms next to each other. To be sure, though, that you are only getting records with your terms adjacent to each other and in the order you wish, be sure to use quotation marks.

Title Searching

This is often the most powerful technique for quickly getting to some highly relevant pages. It may also cause you to miss some good ones, but what you do get has an excellent chance of being relevant. Almost all of the major engines have this option and most of them allow you to search titles by either menu options or prefixes (see Figures 4.1 and 4.2).

URL and Domain Searching

Doing a search in which you limit your results to a specific URL allows you, in effect, to perform a search of that site. Even for sites that have a "site search" box on their home page, you may find that you get better results by doing a URL

search in a large search engine. If you want to find where on the FBI site the term "internship" is mentioned, use a search engine and specify the term "internship" in the search box and "fbi.gov" in the box that allows you to specify URL. Most engines will allow you to accomplish the same thing using a prefix. For example, in Google, you could search for:

internship inurl:fbi.gov

Most engines allow you to be more specific and search a portion of a site, for example (again in Google):

internship inurl:baltimore.fbi.gov

Domain searching is, in many search engines, identical to URL searching. The use of the term, though, points out that you can use this approach to limit your retrieval to sites having a particular top-level domain, such as: gov, edu, uk, ca, or fr. This could be used to identify only Canadian sites that mention tariffs, or to only get educational sites that mention biodiversity.

Link Searching

There are two varieties of "link" searching. In one variety, you can search for all pages that have a hypertext link to a particular URL, and in the other variety, you can search for words contained in the linked text on the page. In the former, you can check, for example, which Web pages have linked to your organization's URL. In the second variety, you can see which Web pages have the name of your organization as linked text. This can be very informative in terms of who is interested in either your organization or your Web site. It can be very useful for marketing purposes, and can also be used by nonprofits for development and fundraising leads. Also, if you are looking for information on an organization, it can sometimes be useful to know who is linking to that organization's site.

This searching option is available in most major search engines on their advanced page and/or on the main page with the use of prefixes. Most engines allow you to find links to an overall site, or to a specific page within a site. If you want to search exhaustively for who is linking to a particular site, definitely use more than one search engine. In link searching, the difference in retrieval is even more pronounced than in keyword searching.

Language Searching

Although all of the major engines allow you to limit your retrieval to pages written in a given language, they differ in terms of which languages can be

specified. The 20 or 30 most common languages are specifiable in all of those engines, but if you want to find a page written in Galician, not all engines will give you that option. If you find yourself searching by language, be sure to look at the various language options and preferences provided by the different engines, particularly if a non-Western character set is involved.

Date

"Date" is one of the most obviously desirable options, and all major engines provide you with such an option. Unfortunately, it may not have much meaning. Due to no fault of the search engines, it is often impossible to determine a "date created" or the "date of publication" of the content of the page. As a "workaround," most engines take the date when the page was last modified and, if that cannot be determined, may assign the date on which the page was last crawled by the engine. For searching Web pages, keep this approximation in mind and do not expect much precision. (On the other databases an engine may provide, such as news or groups, the date searching may be very precise.)

Searching by File Type

Now that search engines are indexing non-HTML pages, including Adobe Acrobat (PDF) files, Word documents, Excel files, and so on, there are times when you may want to limit your retrieval to one of those types. For example, if you wanted to print out a tutorial on using Dreamweaver, you might prefer the more attractive PDF (Personal Document Format) over the format of an HTML page. Specifying file type may not be required very often, but at times it will be useful.

Boolean Search Options

In the context of online searching, "Boolean searching" basically means the following: the process of identifying those items (such as Web pages) that contain a particular combination of search terms. It is used to indicate that a particular group of terms must all be present (the Boolean "AND"), that any of a particular group of terms is acceptable (the Boolean "OR"), or that if a particular term is present, the item is rejected (the Boolean "NOT").

This can be represented by the dark areas in the Venn diagrams shown in Figure 4.3.

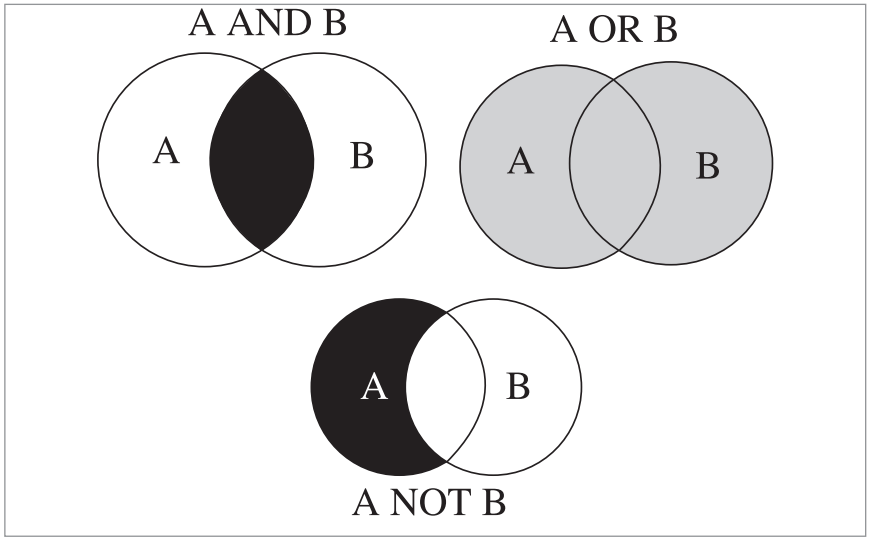


Figure 4.3

Boolean Operators (Connectors)

Very precise search requirements can be expressed using combinations of these operators along with parentheses to indicate the order of operations. For example:

(grain OR corn OR wheat) AND (production OR harvest) AND oklahoma

The use of the actual words AND, OR, and NOT to represent Boolean operations has been downplayed in Web search engines and has been replaced in many cases by the use of menus or other syntax. Even if you have never typed the AND, OR, or NOT, you have probably still used Boolean. (One point here being that Boolean is "painless.") If, from a pull-down menu, you choose the "all the words" option, you are requesting the Boolean AND. If you choose the "any of the words" option from such a menu, you are specifying an OR. Because all major search engines automatically AND your query terms (if you do not specify otherwise), any time you just enter two or more terms in a search box, you are implicitly requesting an AND (even if you do not realize it).

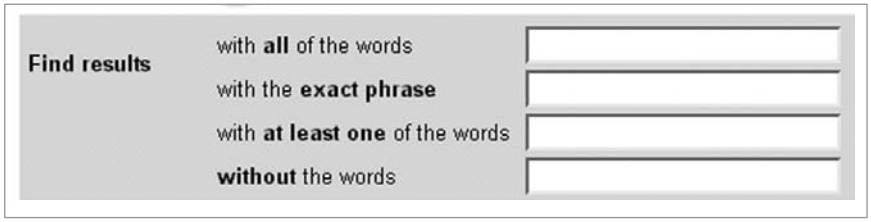
Varieties of Boolean Formats

Just as with title, URL, and other search qualifications, with Boolean you usually have two options for indicating what you want: (1) a menu option or (2) the

option of applying a syntax directly to what you enter in the search box. Using the menus can be thought of as "simplified Boolean" or "simple Boolean."

An example of a Boolean menu option is shown in Figure 4.4.

Figure 4.4



Menu Form of Boolean Choices

The syntax approach varies with the search engine. All major engines currently automatically AND your terms, so when you enter:

prague economics tourism

what you are really going to get is what more traditionally would have been expressed as:

prague AND economics AND tourism

How Boolean operators are expressed varies among engines, and even between the home and advanced pages of the same engine. Figure 4.5 shows an example of Boolean syntax (from AltaVista's Advanced page).

Figure 4.5

Example of Boolean syntax

Full Boolean

Even though most engines provide a syntax that allows you at least to get close to maximum Boolean capabilities, unfortunately each engine has decided to do Boolean syntax in its own way. For example, Google uses an OR but does not use parentheses and AllTheWeb in its home page mode uses parentheses as a substitute for an OR.

Table 4.1 shows how a typical Boolean-oriented search would be structured in the major engines.

Search Engines' Boolean Syntax

Table 4.1

Search Engine/	Boolean Patttern	Full Boolean?	Expression
AllTheWeb (Advanced)	**A and (B or C) andnot D**	yes	"endangered species" and (maryland or virginia) andnot rockfish
AltaVista (Advanced)	**A AND (B OR C) AND NOT D**	yes	"endangered species" AND (maryland OR virginia) AND NOT rockfish
Google	**A B OR C -D**	almost	"endangered species" maryland OR virginia -rockfish
HotBot (Inktomi)	**A B -D**	no	"endangered species" maryland -rockfish
Teoma	**A B -D**	no	"endangered species" maryland -rockfish
WiseNut	**A B -D**	no	"endangered species" maryland -rockfish

SEARCH ENGINE OVERLAP

It is important to recognize that no single search engine covers everything. Due to differences in crawling, indexing, and other factors, each engine includes Web pages that the others do not. In a typical search, if you search a second engine, it will often increase the number of unique records you find by 20–30 percent. Searching a third and fourth engine will also often yield records not found by the first engines. Therefore, if you need to be exhaustive—if it is crucial that you find everything on the topic—do your search in a second and third engine. (Near the end of this chapter, you will see why metasearch engines are NOT the solution to this problem.)

RESULTS PAGES

One of the most useful things a searcher can do is to take a few extra seconds and look not just at the titles of the retrieved Web pages listed there, but look for other things included on results pages and also at the details provided in each record. Most engines provide some potentially useful additional information besides just the Web page results. At the same time they search their Web database, they may search the other databases they have, such as news, images, and directories. You may find some news headlines that match your topic; a link to images, audio, or video on your topic; a directory category; and more.

Also look closely at the individual Web results records. In most search engines, results are "clustered," that is, only the first one or two records from any site will be shown, and there will be a link in the record leading you to "more results from …" or more hits from … ." If you are not aware of these links, you may miss relevant records from that site.

PROFILES OF SEARCH ENGINES

The following detailed profiles provide a look at each of the top five search engines in terms of size and popularity. The descriptions give an overview of the engine, a look at the features provided on the home page and advanced page, and a list of particularly notable additional features provided. For some features, such as news and image databases, just a brief mention is given in the profile, because the subject is covered in detail in the relevant chapter elsewhere in the book. Features that are common to all engines, such as phrase searching, and have already been covered, will not be repeated in the profiles. As you use these engines, expect to occasionally find new features, new arrangements of home pages, and other changes. For updates on such changes, take a look at http://extremesearcher.com, the companion Web site for this book.

ALLTHEWEB

http://alltheweb.com

Overview

AllTheWeb (formerly FastSearch) has been maintaining a position as one of the three largest Web databases, with over 2 billion pages indexed, and it also provides searching of image, news, video, MP3, and FTP databases. The News database covers over 3,000 sources with continual updates. AllTheWeb has a very simple home page, but the advanced search mode provides substantial menu-accessed search functionality with good field-searching capability. Full Boolean capabilities are also available on the home page. More than any other major engine, AllTheWeb allows customization of what appears on search and results pages, and how results and queries are handled.

Figure 4.6

AllTheWeb Home Page

On AllTheWeb's Home Page

You will find the following main features on AllTheWeb's home page:

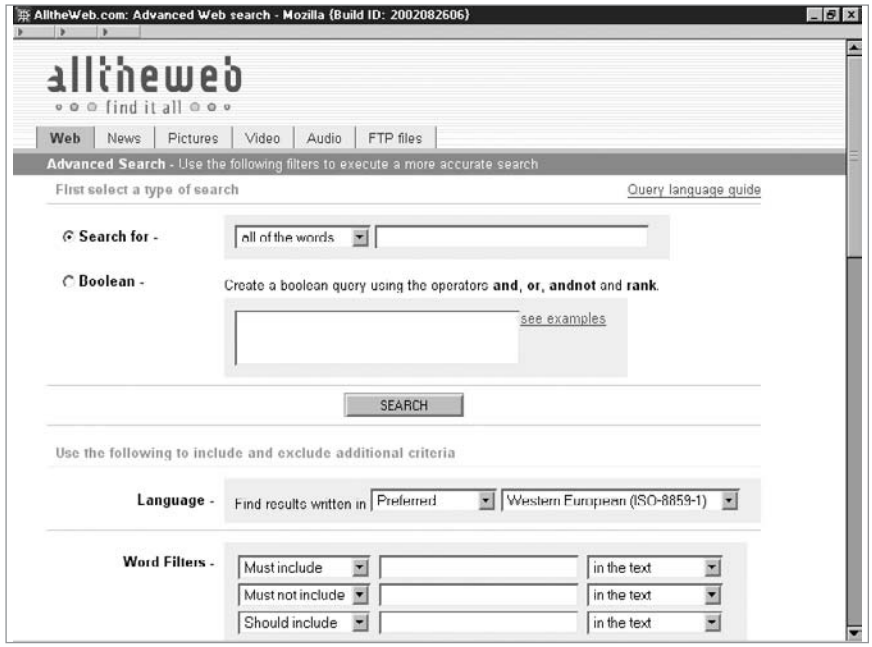
- Search Box. You can enter single words or phrases. Terms are automatically ANDed, but you can also OR terms by putting them in parentheses and you can use a minus sign in front of a term to "NOT" it.
- Links (Tabs). Types of resources offered include News, Pictures, Videos, Audio Search, and FTP searches.
- Customize Preferences Link. This allows you to choose the following options:
  - Offensive Content Reduction
  - Language Settings (Preferred language and encoding)
  - "Site Collapsing"—Clustering or unclustering of results by site
  - Mark Search Terms in Results (highlighting)
- Link to Advanced Search
- Language Option—To view Web pages in any language, or just English. (Note that the default is for English, so you may miss important items in other languages if you do not change this.)

AllTheWeb Advanced Search

AllTheWeb's Advanced Search provides considerably more options than its Simple search. These options include search filters, options for appearance and content of the advanced search page itself, and options for content of the results pages:

- Tabs to other AllTheWeb databases (News, Pictures, Videos, MP3 files, FTP files).
- Search Options. Choose whether you want the terms you enter to be searched as: "all of the words," "any of the words," as "the exact phrase," or as a full Boolean expression. (See discussion of AllTheWeb's Boolean features later.)
- Search Box. Enter terms, prefixed terms (such as "title:term"), or a full Boolean expression.
- Query Language Guide. Leads to a help screen that covers features that can be used in the search box, such as Boolean operators.

Figure 4.7



AllTheWeb Advanced Search Page

- "Site Submit" link to submit a Web site to AllTheWeb.
- Language and Character Setwindows. Offers the choice of searching only those pages in any one of 49 languages.
- Pull-down "Word Filters" windows to specify simple Boolean and fields to be searched:

  Should include (equivalent of Boolean OR)

  Must include (equivalent of Boolean AND)

  Must not include (equivalent of Boolean NOT)

  Field Qualifiers: Text, Title, Link name, URL, Link to URL
- Check boxes to retrieve only pages with the specified embedded content (images, audio, video, RealAudio, RealVideo, Flash, Java, JavaScript, VBScript).
- Domain Filters. To limit to or exclude a specific domain (for example, mit.edu, fr, com). You can also limit to pages from a specific region of the world (based on country codes present in the URLs).
- IP Address Filters. You can limit to, or exclude specific IP addresses. Very esoteric and not really of use to many searchers.
- Result Restrictions:

  File Format. Restrict to PDF, Flash, or Word documents

  Dates pages were updated

  Document size
- Result Presentation

  Number of Results per page. Choices include 10, 25, 50, 75, 100.

  Adult content filter.
- Advanced Search Page Settings

  Save Settings. Saves your selections so that the next time you go to the Advanced Search page, those settings will already be chosen.

  Load Saved. Loads your saved settings.

  Clear Settings. Clears your own settings and goes back to the standard AllTheWeb defaults.

  At the bottom of the page are "Help" and other links.

Search Features Provided by AllTheWeb

AllTheWeb provides all of the more common search capabilities, such as title, URL, and Boolean searching, plus some unique filters, such as for personal homepages. The main options are shown below, but AllTheWeb also provides some additional options for field-searching using prefixes. Take a look at AllTheWeb's help screens for the additional prefix options.

Title Searching

To search for only those pages with your search terms in the title of the page, you can either use the pull-down window on the advanced page (in the "Word Filters" section) or you can use the "title:" prefix in front of your term in the main search box on either the home page or the advanced search page. For example:

title:peugeot

URL Searching

You can limit your search to only those pages from a particular URL or containing a particular term in the URL by either using the pull-down window in the "Word Filters" section of the advanced search page or by using the "url:" prefix in the main search box on either the home page or advanced page. For example

url:fujifilm.com

url:edu

url:uk

The Domain Filters window can likewise be used to limit or exclude a particular domain.

Link Searching

To locate pages that link to a particular site, use the "in the link to URL" option from the pull-down window on the advanced page (Word Filters section), or use the "link:" prefix in the main search boxes.

Language Searching

You can use the Language window on the advanced search page to select only those pages written in any one of 49 languages. On the Customize Preferences page (Language Preferences link), you can select up to eight

"preferred" languages. When you do so, your results will contain only pages in those languages.

Other Fields and Special Search Features

AllTheWeb's advanced search page also allows you to specify special page content such as audio and video, to limit retrieval to personal home pages, and to specify date, file type (Adobe Acrobat, PDF, Flash, Word), document size, and document depth.

Boolean

AllTheWeb's Home Page:

AllTheWeb automatically ANDs all terms unless you specify otherwise.

You can use a minus immediately in front of a term to NOT that term

Example: muskrat -recipes

You can put words in parentheses to do an OR

Example: muskrats (recipe recipes)

AllTheWeb's Advanced Search Page:

On the advanced search page, you can use the pull-down window next to the main search box for simple Boolean by your choice of the "any of the words" or "all of the words" options.

Plus, in the "Word Filter" boxes, you can do simple Boolean and at the same time apply it to a specific field (title, URL, link) by using the two sets of boxes (see Figure 4.1).

"should include"

"must include"

"must not include"

You can also use full Boolean in the main search box by choosing the "boolean expression" radio button and using the following operators: "**and**," "**or**," and "**andnot**." For example:

coffee and decaffeination and (process or method) andnot cancer

Results Pages

Depending upon your search, you may find the following on AllTheWeb results pages:

- Sponsored Results (ads)
- Latest news. Recent headlines that contain your search

Figure 4.8

AllTheWeb Results Page

- Clusters. Retrieved records grouped by category, to enable you to easily narrow your search.
- Multimedia Results. At the same time it does the regular Web search, AllTheWeb also checks its photos and videos databases and, if there are matches, provides a link to those matching items.
- FTP Results. If anything is found in AlltheWeb's FTP collection, a link is provided.
- A link to a dictionary definition of your search terms

When using the advanced search page, you can specify 10, 25, 50, 75, or 100 results per page.

Other Searchable Databases

News Search

The News Search option on AllTheWeb's home page gives access to current news from over 3,000 sources. For details on this feature, see Chapter 8.

Pictures, Audio, and Video

AllTheWeb has an extensive collection of searchable photos, audio files, and videos. Each of these collections is reached by use of the corresponding tab above the search box on either the home page or the advanced page. You will find these discussed in Chapter 7.

FTP Search

AllTheWeb provides an extensive collection of downloadable files. Click on the FTP tab on the main or advanced page. The advanced FTP search page features extensive search options, but the only description of content in results is a brief title, so unless you know exactly what you are looking for, you may find this less easy to use than similar functions on download sites such as CNET Shareware.com (shareware.cnet.com).

Other Special Features

Customize Preferences Page

This page allows you to do the following:

- Change your default database (catalog) to news, pictures, videos, MP3 files, or FTP files.
- Turn Offensive Content Reduction on or off.
- Specify 10, 25, 50, 75, or 100 results per page.
- Turn off highlighting of search terms in results listings.
- Have results you click on automatically open in a new window.
- There are also links for Advanced, Language Preferences, and "Look and Feel" preferences search pages and results.

Advanced Settings

The Advanced Settings page allows you to change some aspects of what appears on the search pages and results pages. Theses choices include turning off automatic rewriting of queries (such as automatically adding quotation marks to common phrases), adding an "any, all, phrase" window to the search box on the main page, turning off site collapsing, and turning on or off some of the features that appear on the results pages.

Language Preferences

To get to this, click the Language link on the Customize Preferences page. That page allows you to set your preference for having results returned only

for languages you choose, or for all languages. You can choose up to eight "preferred languages."

NOTE: AllTheWeb's default is to return only those records in your default language. If you want ALL results, go to the Languages Preferences page and under Select Language, choose Any Language. This can make a big difference in your results!

"Look and Feel" Preferences

Searchers who are bored can change the "skins" and alter the appearance of the AllTheWeb pages.

AllTheWeb Special Features

AllTheWeb also provides a number of interesting and useful special features, including the following:

- URL Investigator—Enter a URL in the search box and AllTheWeb will return information about the URL, including links to information on who owns the site, etc.
- Conversion Calculator. In the search box, enter the word "convert," followed immediately by a colon and a number and unit of measure and AllTheWeb will do metric to Imperial (or vice-versa) conversions. For example, enter convert:27miles
- Spell-Check. If as part of your search, you enter a word of questionable spelling, you will see "Did you mean" and the suggested spelling.
- Calculator. Enter 27*(12+48) in the search box and AllTheWeb will provide the answer. You can use +, -, *, /, and, for an exponent, ^.

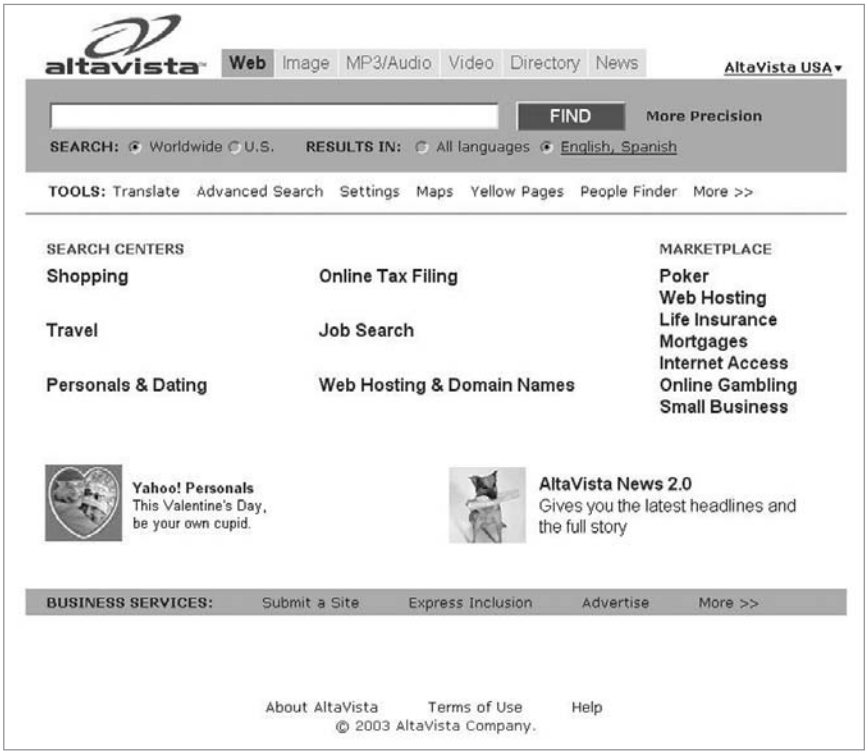
ALTAVISTA

http://www.altavista.com *or* http://av.com

Overview

AltaVista provides a large database and a very broad range of traditional search functionality, with some powerful features, particularly truncation and case sensitivity—that are now rare among Web search engines. As well as the Web database, it also provides databases for searching images, MP3's/audio, video, a Web directory (Open Directory), and News. The latter is updated continually and

Figure 4.9



AltaVista Home Page

includes over 3,000 sources. In its main Web database, AltaVista indexes PDF files as well as HTML files and contains about 1.1 billion pages.

On AltaVista's Home Page

Throughout its history, AltaVista has vacillated between a home page interface that is pure search engine and a portal interface with lots of added features on the home page. It seems to have found a middle road, with visual emphasis on the search features, but retaining a number of links to added portal services and features. The most significant features you will find on the page are these:

- Tabs leading to the different databases: Images, MP3/Audio, Video, Web directory (Open Directory), and News.
- A link to country-specific versions. "AltaVista USA" is the default for U.S. searchers.

- Search Box. Terms are automatically ANDed, but you can qualify a term with a minus sign for a NOT, or apply various prefixes to search a specific field and also use a full Boolean statement.
- "More Precision." This links to a page with boxes for simple Boolean ("all these words," "any of these words," etc.).
- Search Worldwide or U.S. Radio Buttons. The default is "Worldwide."
- Results in All languages or English and Spanish only. Note well that in the U.S. the default is for only English and Spanish! Click on the "English, Spanish" link to get more languages (26 total).
- Tools. Translate (see later discussion), Advanced search link, Settings (country, language, family filter, display options), maps, yellow pages, People Finder (phone numbers).
- Search Centers. Mostly personal services, shopping, and ads.
- Business Services. For information on submitting sites and advertising on AltaVista.

AltaVista's Advanced Search

AltaVista's Advanced Search provides the following functions:

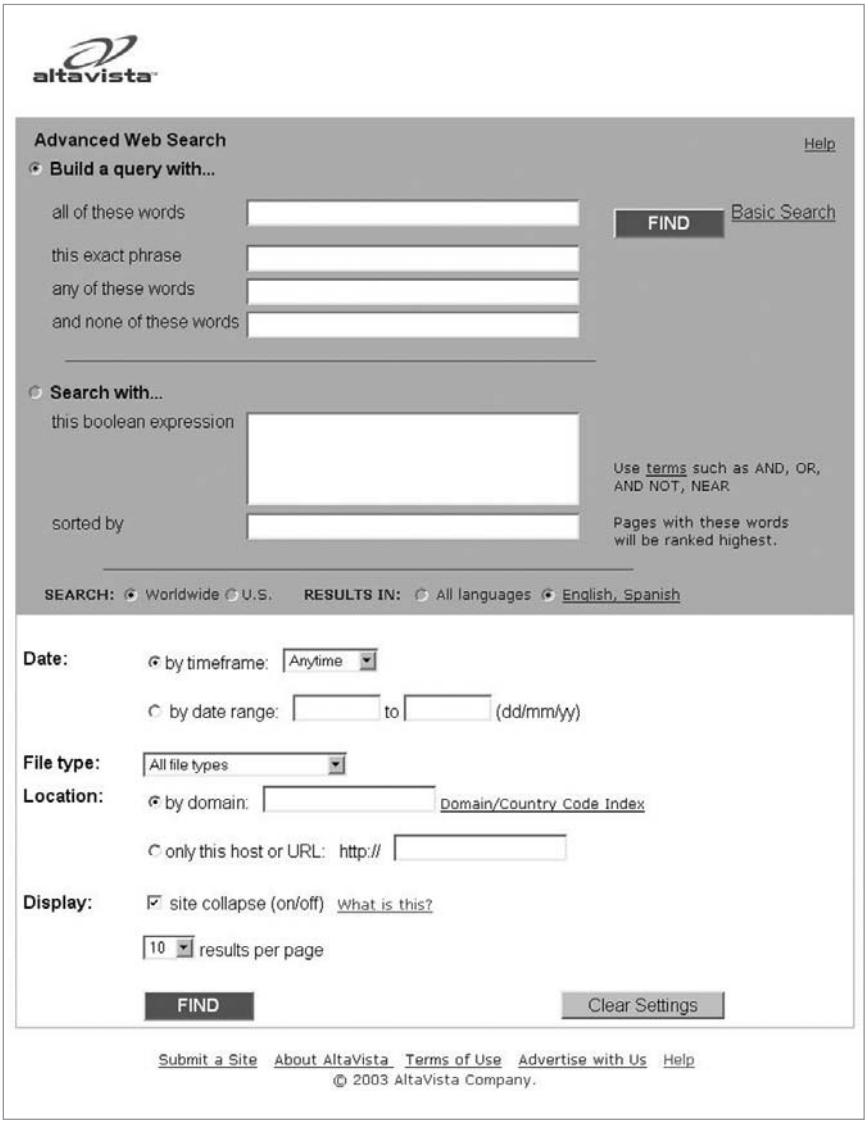
- "Build a query with." Simple Boolean using the "all of these words," "any of these words," and "none of these words" boxes, and also boxes for "exact phrase."

  Full Boolean using the "Search with this boolean expression" box: You can use the operators AND, OR, AND NOT, and NEAR. Be sure to put one or more of your terms also in the "sorted by" box to make the ranking work.
- Search Worldwide or U.S. Radio Buttons. The default is "Worldwide."
- Radio buttons for results in All languages or English and Spanish only. Note that the default is for English and Spanish only. Click on the "English, Spanish" link to get a choice of more languages (26 total).
- Date searching using either a pull-down window or a date range.
- File type. Allows you to select all file types, only HTML, or only PDF.
- Location. You can limit by domain or URL. The Domain/Country Code Index link provides a list of all country codes and U.S. top-level domains.
- Option of turning off the "site collapse" (clustering) option.
- Choice of number of results per page—10, 20, 30, 40, or 50.

Search Features Provided by AltaVista

AltaVista provides all of the most common field search capabilities and three features that are currently unique for Web search engines, although they are common in proprietary search services (NEAR, truncation, and case sensitivity). It also provides full Boolean capabilities.

Figure 4.10



AltaVista's Advanced Search Page

Title Searching

To search for pages that have your term(s) in the title, you must use the "title:" prefix.

Examples: title:palamino title:"new caledonia"

URL Searching

On AltaVista's home page, you can search for pages from a specific URL by using the "url:" prefix. You can use the URL specification by itself, to find all pages from the site, or you can combine it with another term, e.g., "flat panel monitors" url:dell.com.

On the advanced search page, you can search for pages that are from a particular URL by using the "only this host or URL" box in the Location section of the advanced search page. When you search using this approach, however, you must have other terms in the search boxes in order for it to work.

The "by domain" box should be used when you want to limit to a top-level domain such as gov or fr. You can also limit such searches on the main page by using the "domain:" prefix.

Link Searching

To find pages that link to a specific page, use the "link:" prefix on AltaVista's home page, for example: link:extremesearcher.com.

Language Searching

On both the home page and the advanced page, there are radio buttons to specify that you retrieve results in "All languages" or "English, Spanish" only. The default is for only English and Spanish, so if you don't want to miss anything, click on "All languages." If you click on the "English, Spanish" link, a table will appear, allowing you to choose from 26 languages.

Date

In Advanced Search mode, AltaVista also allows for specifying a period (the last week, month, year, etc.) or a date range using the date range boxes.

The date should be entered in the dd/mm/yy format:

31/10/99

Remember that generally, date searching is only "approximate."

File Type

Because AltaVista now indexes PDF files as well as HTML files, the File Type window allows you to retrieve either or both of the file types.

Other Fields

The following fields are also searchable by the use of the prefix shown:

anchor: Searches for clickable text terms.

applet: Finds particular Java applets used on a page.

object: Finds programming objects such as Flash objects.

host: Acts the same as using the url: prefix.

image: Searches for a term in an image file name.

like: Finds similar pages.

text: Finds text anywhere on the page other than an image tag, link, or URL.

Boolean

From the home page, if you click on the More Precision link, you are presented with a page that allows you to use simple Boolean by means of the "all these words," "any of these words," and "none of these words" boxes. The same boxes are available on the advanced search page.

You can use full Boolean (AND, OR, AND NOT) in either the search box on the home page or in the "boolean expression" box on the advanced search page. For example:

haseltine AND (painter OR painting) AND (italy OR italian)

Other Search Features

NEAR

One of the unique and powerful features of AltaVista is the NEAR operator. When used between two terms, it specifies that the two words must be within 10 words of each other. This is especially useful for names, since it allows the words in either regular or inverted order and also allows one or more middle names. It should be used whenever you need two words near each other but want to allow for intervening words and for the words to occur in either order. It can also be used along with the Boolean operators.

Examples: john NEAR kennedy
speeches AND john NEAR kennedy

Truncation

Sometimes referred to as "wildcard" searching, this feature allows you to end a string of characters with an asterisk and automatically retrieve all terms that begin with that string. For example, "metal*" will retrieve "metal," "metals," metallic," and so on.

One asterisk will retrieve any number of additional characters. You can also use the asterisk in the middle of a word.

Truncation can be used with prefixes:

Example: title:russia*

Automatic Phrases

AltaVista automatically identifies thousands of common and not-so-common phrases and automatically treats those as if they had quotation marks around them. To be safe, put in the quotation marks yourself when you need them. Also be aware that you may be getting some unwanted narrowing done if you do not remember that the automatic phrasing may be taking place. "Military history" and military history both yield the same result. "Military intelligence" and military intelligence do not.

Case Sensitivity

AltaVista is the only major Web search engine that allows you to specify case sensitivity. To indicate that you want an exact case match, enter your term with the appropriate case and within quotation marks in the home page search box. Otherwise, case is ignored and all case variations are retrieved.

"SALT" will retrieve SALT, but not Salt or salt (unless those words happen to also appear on the same page). Without the quotation marks, all case variations are retrieved. Taking advantage of this can be especially useful when searching for acronyms.

On the advanced search page, whenever any term containing one or more uppercase letters is entered in the Boolean expression box, case is also recognized, even if you do not put your term within quotation marks.

Translate

AltaVista, utilizing the SYSTRAN company's Babel Fish translation software, offers an immediate machine translation of a Web page by clicking on

the Translate link at the end of a results record. It will translate either way between English and French, German, Italian, Portuguese, Spanish, Japanese, Korean, and Chinese; from Russian to English; and also some non-English combinations. You can also take advantage of the translation feature by clicking the Translate link under the Tools section of the home page. By doing so, you can enter either a URL to have a page translated, or enter up to 150 words in the text box.

Don't expect a good translation, but it may be an adequate translation for a basic understanding of the content of a Web page or a block of text.

Although it may take a while to load, the "World Keyboard" link on the translation page will pull up an on-screen keyboard that allows you to type in any one of seven languages (French, German, Italian, Portuguese, Russian, Spanish, English), with all of their unique accent marks and characters.

Settings Page

The Settings Page (found under Tools on the home page), allows you to specify these items:

- Languages to search in
- What you want to see in Web results records (description, URL, page size, language, translate link, related pages link)
- Highlighting of search terns
- Number of results per page

Results Pages

On AltaVista results pages, in addition to the Web results and "Sponsored Matches," you will find a list of phrases under "Refine your search with AltaVista Prisma." These phrases are the most common terms found in the records retrieved in the current search and can provide useful ways of refining your topic.

Other Searchable Databases

Images, Audio, and Video

AltaVista has one of the largest image databases and also has significant and easily searchable MP3/Audio and Video databases. These databases are accessed by clicking the appropriate tab on AltaVista's home page. For details on using them, see Chapter 7.

Directory

Clicking on the Directory tab on the home page takes you to AltaVista's implementation of Open Directory. You can either browse through its 10 top-level categories or search it using the search box on the directory page.

News

Clicking the News tab on AltaVista's home page takes you to a page that provides headline stories in several categories as well as a box that allows searching of the 3,000 news sources included. For details, see Chapter 8.

GOOGLE

http://www.google.com

Overview

In a period of only about four years, Google went from being a brand new introduction to becoming the favorite search engine for the majority of search engine users. Its own popularity has been based on its use of the popularity of a Web site as the major ranking factor, its simplicity for the casual user, and its vigorous efforts to increase both the size of its database and the provision of additional features and types of content. It ranks records mostly on the popularity of the page as measured by how many pages link to that page and how popular those linking pages are. (Web pages are known by the friends they keep.) Google's output is unique in that it allows you to go to the page as it is currently on the Web, or to go to a cached copy that Google stored when it retrieved the page. Google is at present also the best source for newsgroup searching (with a Usenet collection going back over 20 years), for images, and for PDF and other non-HTML files**.** Google's Web database contains about 3 billion records.

On Google's Home Page

One of the reasons for Google's immense popularity is its insistence on a simple, uncluttered home page. Even though the home page has been kept simple, a single click uncovers a number of features. The home page includes the following items:

Figure 4.11

Google's Home Page

- Links to Google's databases:
  - Web (the default) database. Images. Leads to one of the largest image search databases on the Web.
  - Groups. Allows searching of 800 million Usenet postings back to 1981!
  - Directory. Link to Google's implementation of Open Directory.
  - News. Covers 4,500 news sources going back 30 days.
- Link to Advanced Search
- Language and Display Preferences
  - Language search and interface preferences
  - Number of results per page
  - Option to have results opened in new window
  - Safe Search option (adult content filter)
- Language Tools providing these capabilities:
  - Limiting retrieval to a specific language or country of origin

  - Translating a specific Web page between English and five languages (French, German, Italian, Spanish, Portuguese, or Russian) or between French and German
  - Choice of having the Google interface in any one of over 60 languages
  - Links to the Google country-specific versions for 77 countries
- Search box. Enter one or more words. The minus sign in front of a term (for NOT) and ORs can be used.

  Google will ignore small, very common words unless you insert a plus sign in front of them. Google will ignore quotation marks.
- "I'm Feeling Lucky." This selection automatically takes you to the page that Google would have listed first in your results (mostly a gimmick).
- Various special options. Links for information on advertising, the company, and Google Services and Tools, which provides links to a number of special Google offerings and tools such as the Froogle shopping search engine, the Google toolbar for your browser, the Google Answers service, Google catalog search, and other features.

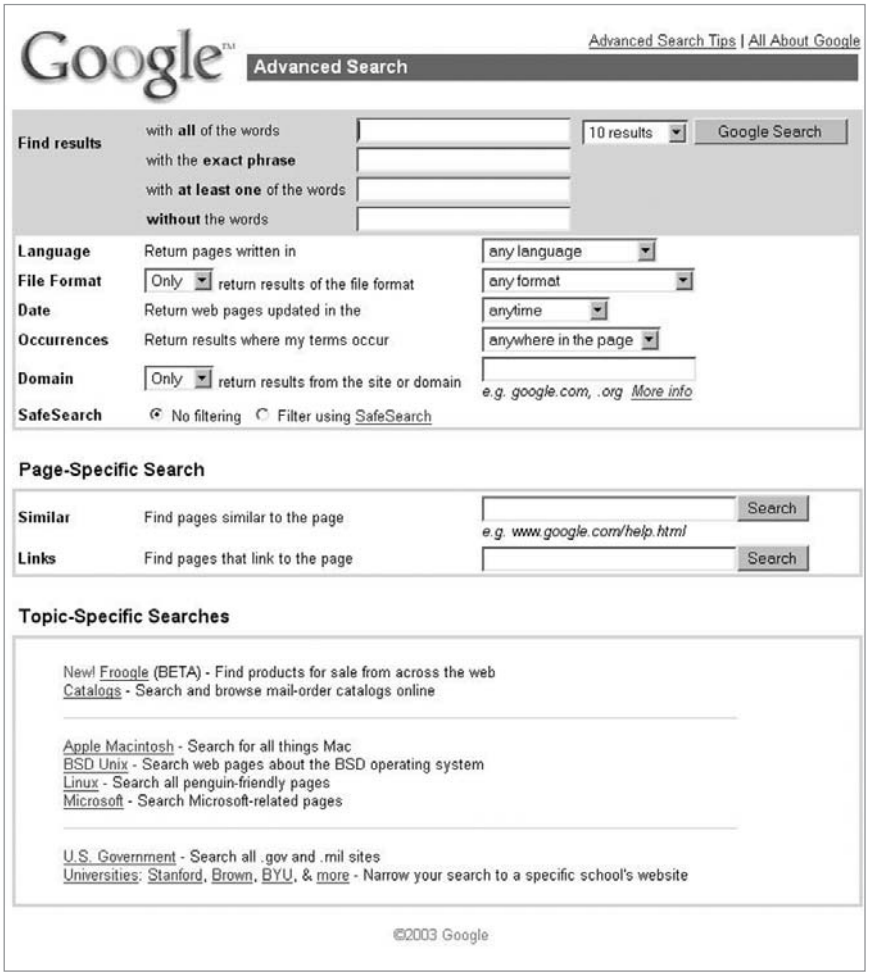
Google's Advanced Search

Although, as with other engines, many searches can be effectively accomplished by putting one or two terms in the home page search box, if you need enhanced capabilities, Google's advanced search page provides them. It has all of the common field search options (title, URL, link, language, date) and less common options as well.

In roughly this order, you will find the following on Google's advanced search page:

- Boxes to perform simple Boolean combinations ("all the words," etc.).
- Choice of 10, 20, 30, 50 or 100 results per page.
- Choice of searching for documents in all languages or any one of 35 languages.
- Option to retrieve only a specific file format (PDF, xls, doc, ps, Ppt, rf).
- Date restriction (anytime, last 3 months, last 6 months, last year).
- Window to limit retrieval to title or URL fields.
- Box for limiting to (or excluding) a particular domain or URL.
- Adult content filter option.

Figure 4.12



Google's Advanced Search Page

- "Page Specific Search" for pages that are *similar* to a particular page whose URL you enter in the box.
- "Page Specific Search" for pages that link *to* a particular page (enter the URL of the page of interest).
- Links to "Topic-Specific Searches."
  - Froogle product search and Catalog Search.
  - Apple Macintosh. Searches for Mac-related pages.
  - BSD Unix. Searches Web pages about the BSD operating system.
  - Linux. Searches Linux-related pages.

- U.S. Government. Searches all .gov and .mil sites.
- Universities. Searches pages from selected universities.

Search Features Provided by Google

Both by the menus on the advanced page and prefixes on the main page, Google provides field searching for all of the commonly searchable Web page fields (title, URL, link, language, date), plus searching by file format and for "similar" pages.

Title Searching

Searches can be limited to words appearing in the page title by either of two ways. On the advanced search page, you can enter your terms in the search boxes, then, under the Occurrences section of the page, choose "in the title of the page" from the pull-down menu.

You can also, on the home page, use the "intitle:" or "allintitle:" The "intitle:" prefix specifies that a single word or phrase be in the title.

Examples: intitle:online intitle:"online strategies"

"Allintitle:" is used to specify that all words that follow the colon be in the title, and not necessarily in that order. For example, the following would retrieve titles with both words somewhere in the title, not necessarily in that order:

allintitle:nato preparedness.

These prefixes can be combined with a search for a word anywhere on the page.

Example: summit intitle:nato.

You cannot do a combination like the one just mentioned using the menus on the advanced page because your single menu choice there will apply to all terms you enter in the search boxes.

URL Searching

Limiting retrieval to pages from a particular URL is done in a way that is parallel to title searching. You can do it either on the advanced search page with menus or on the home page by using prefixes. On the advanced search page, enter a URL or part of a URL in the search boxes, then choose "in the url of the page" from the Occurrences section of the advanced page.

On the home page, you can use the "inurl:" or "allinurl:" prefixes.

Examples: inurl:bbc inurl:bbc inurl: "bbc.co.uk"
allinurl:bbc co uk

On the advanced search page, you can also use the "Domain" box to search for a part or all of a URL (uk, edu, ford.com). To do a "site search" for a particular topic, enter terms for your topic in the search boxes and the URL in the Domain box.

A site search can be done on the home page as follows:

hybrid inurl:ford.com

The prefix "site:" is almost identical to the "allinurl:" prefix. With "site:", however, you have to use a search term as well.

Example: hybrid site:gm.com.

Link Searching

To find pages that link to a particular site, you can use the "Links" box in the "Page-Specific" section of the advanced search page by entering the URL in the box, or you can perform the search on Google's home page by using the "link:" prefix. For example, to find pages that link to the Modern Language Association site, search for:

Example: link:mla.org.

Language Searching

To limit retrieval to a particular language, use the Language menu on the advanced search page. The default is "all languages," but you can choose any one of over 30 languages. If you wish to make a particular language your default choice, you can do so on Google's Preferences page. On that same page, you can also choose to have the Google search pages appear in any of those languages.

Date Searching

The Date window on the advanced search page allows you to limit results to pages that are new in the last three months, six months, or year. Keep in mind that date searching is only an approximation, because the origination date or last updated date is often not clearly identified on most Web pages.

File Type

Google indexes more file types than any other Web search engine and includes the following in its index in addition to HTML pages: Adobe Acrobat files (.pdf), Adobe Postscript files (.ps), Microsoft Word files (.doc), Microsoft Excel files (.xls), Microsoft PowerPoint (.pdf), and rich text format (.rtf.). You can limit your retrieval to any one of these by using the File Format window on the advanced search page.

On Google's home page, you can accomplish the same thing by using the "filetype:" prefix. For example, if you want a 1099 IRS tax form that you can print out, you could search for: 1099 IRS form filetype:pdf.

Related (Similar) Pages

You can search for pages that are similar to a particular page by using the Similar box in the Page-Specific section of Google's advanced search page. Enter the URL of the page in the box. This can also be done by using the "related:" prefix on Google's home page.

Example: related:searchenginewatch.com

Other Fields

The following prefixes can be used on Google's home page to search for the information indicated.

cache: Enter a URL after the colon and you will get Google's cached version of the page.

Example: cache:www.aps.org

info: Enter a URL after the colon to get basically the same information that is shown on a results page when the record for that site is retrieved.

Example: info:cyndislist.com

stocks: Enter one or more stock symbols after the colon to get links to stock quotes.

Boolean

On the home page, Google automatically "ANDs" all of your words. You can also use a minus sign in front of a term to NOT a term, and you can use one or more ORs. (The OR must be capitalized.)

Example: warfare chemical OR biological -anthrax

This search expression would get all records that contain the word "warfare" and also contain either "chemical" or "biological" but would eliminate all records containing the word "anthrax."

On Google's advanced search page, simple Boolean is done by use of the "with all the words," "with at least one of the words," and "without the words" boxes.

Other Search Features

"Wildcard" Words

Google allows the use of one or more asterisks for "wildcard" *words* (not to be confused with "truncation," which is for wildcard *characters* within or

at the end of a word). You can use the asterisk for unknown words in a phrase search. The use of *each asterisk insists on the presence of one word.*

Example: "erasmus * rotterdam"

will retrieve "Erasmus Universiteit Rotterdam" and "Erasmus von Rotterdam." It will not necessarily retrieve any "Erasmus Rotterdam" records.

Example: "erasmus * * rotterdam"

will retrieve "Erasmus University of Rotterdam," but not necessarily the "Erasmus Universiteit Rotterdam" records.

If you want "Franklin Roosevelt" and also "Franklin D Roosevelt" and also "Franklin Delano Roosevelt," you would search for:

"Franklin Roosevelt" OR "Franklin * Roosevelt"

Results Pages

On Google results pages, it pays to look closely at the entire page and also at the content of the individual records.

On the line where Google reports your results, for example,

Searched the web for belgium. Results 1 - 10 of about 7,800,000

look for underlined terms. Clicking on them will lead to dictionary definitions from Dictionary.com.

When you do a Web search, Google also searches its directory (Open Directory). If you see a Category link near the top of the page, it means that Google found a category heading that matches your search term. Clicking on the category link will take you to that category in the directory. Matches from individual sites from Open Directory will appear among your results (and are identifiable by the presence of a Category in the record).

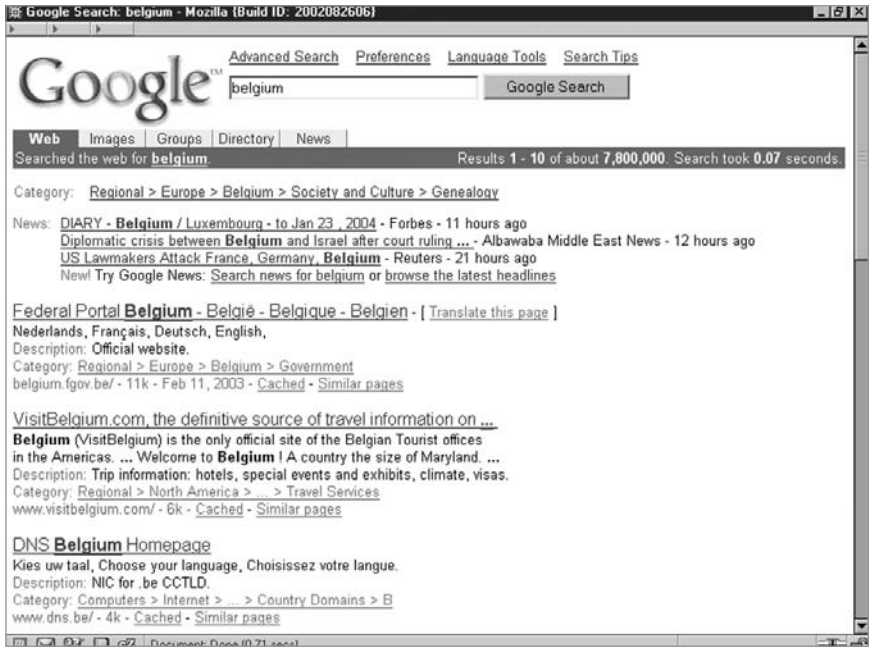
Google also searches its News database whenever you do a Web search. If your topic has been in the news recently, you may see up to three headlines. Click on them to go to the news stories.

There are several parts of individual records that are worthwhile to examine.

Example of Output:

VisitBelgium.com, the definitive source of travel information on ...
Belgium (VisitBelgium) is the only official site of the Belgian Tourist offices in the Americas. ... Welcome to Belgium ! A country the size of Maryland. ...
Description: Trip information: hotels, special events and exhibits, climate, visas.
Category: Regional > North America > ... > Travel Services
www.visitbelgium.com/ - 6k - Cached - Similar pages

Figure 4.13



Google Results Page

Clicking on the Cached link in the record will take you to a cached copy that Google stored when it retrieved the page. This feature is especially useful if you click on a search result and the page is not found, or it is found, but the terms you searched for do not seem to be present. If this happens, go back to the Google results page and click on the Cached link.

Clicking on "Similar pages" will take you to pages with similar content ("More like this"). Take advantage of this capability to find related pages that may be difficult to find otherwise.

Other Searchable Databases

In addition to the Web database of over 3 billion pages, Google also provides searching of Images, Groups, Directory, and News databases. Each of these is accessible by clicking the appropriate tab above the search box on

Google's main page (and on many other Google pages). Because each of these Google databases is discussed in some detail in either Chapter 7 (…Images, Audio and Video), Chapter 5 (Groups …), Chapter 2 (General Web Directories …) or Chapter 8 (News…), they are mentioned just briefly here.

Google Image Search

Google's Image Search is possibly the largest searchable image collection on the Web, containing over 400 million images. Details on this type of searching are covered in Chapter 7.

Directory

Google uses Open Directory for its browsable and searchable directory database. A search of the directory categories is integrated, automatically, into all searches, with matching categories appearing near the top of the results page and hits from Open Directory incorporated into the results list. For details on Open Directory itself, please see Chapter 2. Although Open Directory category pages and results pages look slightly different whether you are searching its own site (http://dmoz.org) or through Google, the content, arrangement, searchability, and browsability are virtually the same. The biggest difference is that when you search the directory through Google, results are ranked by Google's ranking algorithm.

Google Groups (Newsgroups)

Google provides access to the Usenet collection of newsgroups, covering over 20 years and containing over 800 million messages. For details on Google Groups, please see Chapter 5.

Google News

Google's News Search is reachable by the tab on Google's home page, or directly at http://news.google.com. It covers about 4,500 news sources and is updated continually. Records are retained for 30 days. For details, see Chapter 8.

Other Google Features and Content

The folks at Googleplex, Google's headquarters, let no grass grow beneath their thousands of computers. They are constantly adding new things. Interestingly, many of the new things receive relatively little press. Informal polling shows that many Google users have not even clicked on the tabs on Google's home page to see what is there, and even many very experienced searchers

have not had time to fully explore everything Google offers. The Google offerings described below are some of the more significant of these features and content. For a look at the other offerings, use the links at the bottom of Google's home page, particularly Services & Tools and Jobs, Press, & Help. The names of these links change occasionally, so also look around for All About Google and Cool Things links.

PDF Files and Other File Formats Retrieved by Google

PDF (Adobe's Portable Document Format) files were formerly a part of the Invisible Web, and not identifiable or retrievable by general Web search engines. Google started indexing documents in this file format in 2001 and fairly quickly began adding other files types, including Word (.doc), Excel (.xls), PowerPoint (.ppt), and rich text format (.rtf) files. Now if a Web page contains a link to any of these types of files, the file not only gets indexed, but gets indexed in depth. In the case of Excel files for example, when Google finds one and indexes it, not just column and row headings get indexed, but every cell. This level of access can be quite a boon for researchers in areas such as demographics and trade. For those who do not have the corresponding software (Word, PowerPoint, etc.), Google also provides a link in each record to view the file in HTML format. Specific file types can be selected by using the Format window on the Advanced Search page, or, on the home page, by using the "filetype:" prefix.

Example: filetype:doc

Phone Book and Address Lookup

A phone book lookup for U.S. phone numbers and addresses can now be done on Google, directly from the home page search box. For a business, type a business name and either city and state or ZIP code. For individuals, give the first name or initial, the last name, and either state, area code, or ZIP code. It will also work without either the first name or initial if the last name is not very common. As with all phone directory sites on the Web, do not expect perfect results all the time.

You can also do a reverse lookup just by entering the phone number in the search box, with or without punctuation. Include the area code.

Stock Search

Enter a ticker symbol in the search box to get a link to stock quotes (from Yahoo! Finance). You can actually enter several at the same time.

Preferences Page

Click on the Preferences link on the home page to get to this. Once there, you will find that you can change the default interface language (for tips and messages), specify which languages you want to see in your results, turn off the adult content filter, specify the number of results per page, and have results opened in new windows.

Language Tools Page

This page, that you get to from the Language Tools link on the home page, provides another place where you can specify a language to which you want your results limited. This page also allows you to limit results to only those from a particular country. Because the Language Tools page sets up defaults that will control your results until you go back to the page again, for most people it will probably be wiser to use the Domain box on the advanced search page to specify country only when needed.

On this page you will also find a translation program (from SYSTRAN, the translation program also used by AltaVista) that allows you to translate blocks of text or a Web page between various combinations of English, German, French, Italian, Portuguese, and Spanish.

Froogle

Google's shopping engine, Froogle.com, was introduced in 2002 and contains product pages Google has identified by crawling the Web to identify product sites as well as pages derived from catalogs submitted by merchants. For more details on Froogle, see Chapter 9, Finding Products Online.

Catalog Search

Google's Catalog Search is a database of published merchant catalogs and contains catalogs of over 5,000 merchants. It is accessible either by links on various Google pages or by going directly to http://catalogs.google.com. The main page contains a subject directory that allows you to browse by category, a search box, and also a link to an advanced catalog search. Using the advanced search, you can search the entire collection, a category, or an individual catalog. You can view an actual image of every catalog page, or just the portion for a particular product.

Google Toolbar

The Google Toolbar is a free downloadable feature that allows you to have the Google search box and additional features as a toolbar on Internet Explorer. Go to the "Services and Tools" link on the home page to find out about what the Google Toolbar provides:

- Google **Search**: The search box can always appear on your browser screen.
- **Search Site**: To search only the pages of the site currently displayed.
- **PageRank**: See Google's ranking of the current page.
- **Page Info**: Get more information about a page, similar pages, and pages that link to a page. You also get a cached snapshot.
- **Highlight**: Will highlight your search terms (each word in a different color).
- **Word Find**: To find search terms wherever they appear on the page.

The Google Toolbar can be customized to include most of the features on the regular Google home page (and in several languages).

Calculator

For a quick arithmetic calculation, as with AllTheWeb, you can use the Google search box. Enter 46*(98-3+32), and Google provides the answer. You can use +, -, *, /, and, for an exponent, ^.

Google Answers

This is a service whereby users can ask questions that are then answered by other users who have signed up as researchers. You submit a question, and pay a 50¢ fee plus an amount that you are willing to pay for the answer (from $2 to $200). Researchers then bid to answer your question. See the Google Answers FAQs at: http://answers.google.com/answers/faq.html. Be aware that no particular qualifications are required for a person to become a researcher for this service.

Figure 4.14

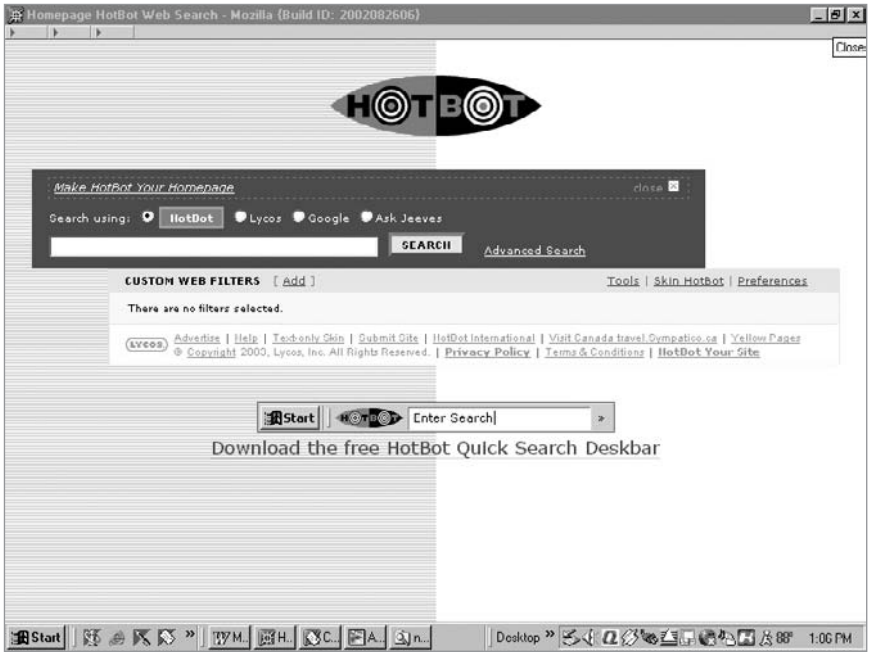
Google Toolbar

HOTBOT

http://www.hotbot.com

Overview

HotBot is one of the oldest Web search engines. It remained quite unchanged and unenhanced from 1998 until 2003, when it reengineered its site, leaving virtually nothing intact and adding some good new—and unique—features. The new interface has a single search box, but with radio buttons allowing your search to be done in either the Lycos (AllTheWeb's) database; Google's database; HotBot's original, main database (Inktomi); or Ask Jeeves (Teoma's) database. For its advanced version, HotBot provides a somewhat standardized interface for each of the four databases, allowing you to take advantage of most of the advanced features of those databases without having to reorient yourself in very differently arranged advanced search pages. The home page is customizable to the extent that it can contain all of the features provided on the advanced page for searching the Inktomi database. For a quick comparison of the top results from some of the top search engines, or to move quickly from the advanced search features of one engine to another, HotBot may be a good starting place. HotBot's Inktomi database contains about 1.5 billion records.



HotBot Home Page

Figure 4.15

On HotBot's Home Page

On HotBot's home page you will find the following elements:

- Radio buttons allowing you to choose the database to be searched: Lycos, Google, the main HotBot database (Inktomi) or Ask Jeeves
- Search box
- Link to Advanced Search
- Customize Web Filters/Preferences

You can add any or all of the following search features to the home page:

- Language
- Domain/Site
- Region (continent)
- Word filters menu (any, all, none of the words, and phrase), and field specifications for title, URL, and contained URLs (link-to's).
- Date
- Page content (audio, image, etc.)
- Block Offensive Content option

You can specify that the following appear on results pages:

- Number of results
- Description shown in records
- URL shown in records
- Date shown in records
- Page size shown in records
- Related searches shown
- Related categories shown
- Whether you want results opened in the same or a new window.

On the definitely trivial side, you can also choose "skins" that have varying degrees of the old HotBot green and blue.

HotBot's Advanced Version

To understand both the nature and the power of HotBot, keep in mind that it has its own database (Inktomi) and also provides, in a consistent-as-possible format, interfaces for three other Web databases. When using the advanced page for Inktomi, you have the following options:

- Choice of database (engine). Use the radio buttons to switch to HotBot's interface for Lycos, Google, or Ask Jeeves

- Search box
- Link to Advanced search to get to filter options for the other databases
- Filters:
  - Language. For limiting your retrieval to any one of 35 languages
  - Domain/Site. To limit to, or exclude a specific domain
  - Region. To limit retrieval to a specific continent, and within North America (to limit to com, edu, gov, mil, net, org)
  - Word Filter (Simple Boolean). All, Any, None of the words, phrase
  - Fields. Limiting retrieval to pages with your terms in the body, title, URL, or referring URL.
  - Date. Limiting to anytime; the last week or month; or before, after, or on a specific date
  - Page Content. Limiting retrieval to pages containing audio, video, Java, or other file format

HotBot Advanced Search Interface to Lycos, Google, and Ask Jeeves

For the advanced interfaces for the other three databases, HotBot provides the following options:

- Lycos. Language, Domain/Site, Region, Word Filter, Date, Page Content, Adult Filter
- Google. Language, Domain/Site, Word Filter, Date, Adult Filter
- Ask Jeeves. Language, Region, Date, Adult Filter

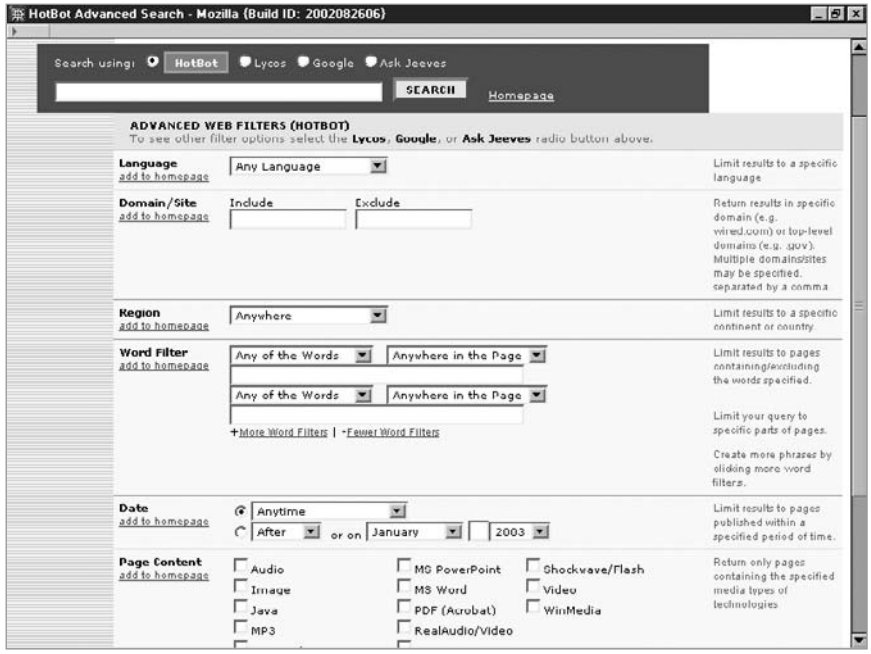
Search Features Provided by HotBot

HotBot's interface for Google, Lycos, and Ask Jeeves provides searchablilty of many but not all of the fields that are searchable in those engines directly. HotBot's version of Inktomi offers a very good collection of searchable fields by using the appropriate windows on the advanced search page.

Title Searching

To perform a title search on HotBot, enter your term(s) in the search box and choose "title" in the Word Filters menu.

Figure 4.16



HotBot's Advanced Page

URL Searching

To perform a search for all pages from a specific URL, enter the URL in the search box and choose "In Contained URLs" in the Word Filters menu.

Link Searching

To use HotBot to identify those pages that link to a particular site, enter the URL in the search box and choose "referring link" in the Word Filters menu.

Language Searching

To perform a search by language, enter your term(s) in the search box and choose the language from the language menu.

Date Searching

To limit retrieval by date, you can either choose a time frame such as last week, or last month or you can specify before, after, or on the date you select in the date boxes.

Page Content

You can use the checkboxes on HotBot's advanced page to limit retrieval to those pages that contain one or more of the following content types: audio, image, Java, MP3, MS Excel, MS PowerPoint, MS Word, PDF, Real Audio/Video, Script, Shockwave, Flash, video, or WinMedia. You can also specify a specific extension such as .gif or .jpg.

Boolean

If no qualifiers are inserted between terms, HotBot (for any of the four databases) will AND the terms.

You can use Google's, AllTheWeb's, or Teoma's Boolean syntax, but it will probably only work correctly in that engine, so you will probably be better off going to the engine itself if you want to use Boolean syntax.

You can do simple (all the words, any of the words, none of the words) Boolean by using the Word Filters menu on the advanced pages.

OR will work, but it is not currently documented on the HotBot site.

Example: turkey dressing OR stuffing

You can use a minus to NOT a term

Example: turkey dressing OR stuffing -oyster

Output

HotBot's results pages show the first 10 records from the selected database (with the usual links at the bottom to get to the rest of the results) and a few sponsored links (ads) at the top. The records are all in a HotBot format, with the page title, a line or two of description, the URL, and the page size. Content of results records is also customizable. The downside to the results pages is that you do *not* get much of the significant additional output content and features that you will find if you search Google, AllTheWeb, or Teoma directly.

Also, you may get fewer matches in HotBot's interface for the other engines than in the engines themselves. Each of them clusters results and only shows the first one or two records from any particular site. They provide links to get to other matching records from those sites. HotBot's interface does not provide such links; therefore you will get only the first one or two matching records from any site.

Special Options/Features

HotBot's biggest and most important special feature is its capability for searching several major engines (see earlier discussion). It also provides a Related Searches and a Related Categories option for results pages.

Related Searches

By choosing Related Searches on the Results Preferences page, you can have HotBot results show searches that were done by other searchers using your terms. This feature works on a search in any of the four databases.

Related Categories

HotBot uses a search of Open Directory to identify related categories. The categories appear when you search in any of the four databases.

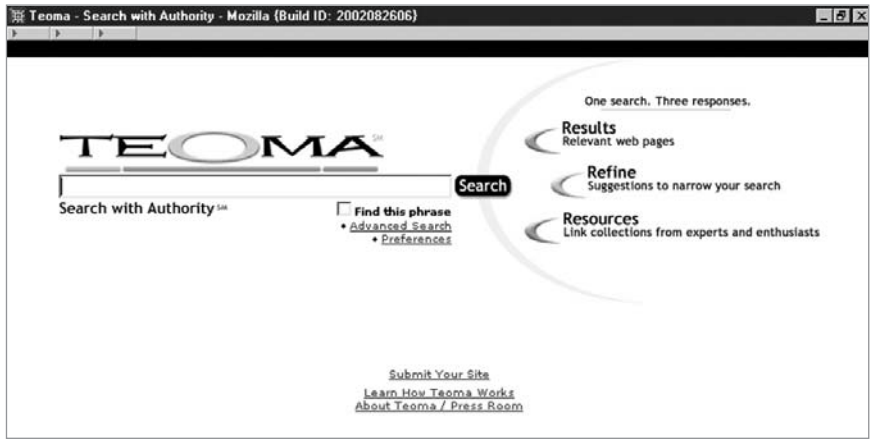
TEOMA

http://teoma.com

Overview

Teoma is among the newest Web search engines. It is growing, but at present typically yields only around one half the number of records that Google finds. As a result, it will probably not be the first choice for most searches. Its greatest strength lies in the Resources section of results pages, where you will find a list of collections of links (metasites, resources guides). These collections are basically specialized directories that Teoma has identified, and the capability of identifying them makes Teoma unique. It also has jumped on the bandwagon

Figure 4.17



Teoma's Home Page

for categorizing results and, like WiseNut (mentioned later), mimics the late Northern Light's approach while providing some variations on the theme.

On Teoma's Home Page

Teoma has a very simple home page on which you will find these items:

- The Search box
- A phrase search option (just use quotation marks, instead)
- A link to Teoma's Advanced Search
- A Preferences link. You can choose the number of results per page (10, 20, 30, 50, or 100).

Teoma's Advanced Search Page

Teoma's advanced page provides options for all of the most typical search engine search features.

The page includes these features, in the order they appear on the page:

- Number of results per page (10, 20, 30, 50, or 100)
- Simple Boolean (must, must not, should) menus
- Search boxes. "Find" and "Include or exclude words or phrases" boxes
- Field menu. anywhere, title, URL
- Language (10 languages)
- Domain/Site
- Geographic region (continent)
- Date

Search Features Provided by Teoma

Teoma provides several field searching options by means of menus on the advanced page or by using prefixes. When you use a prefix, Teoma usually requires that it be in combination with a regular search term.

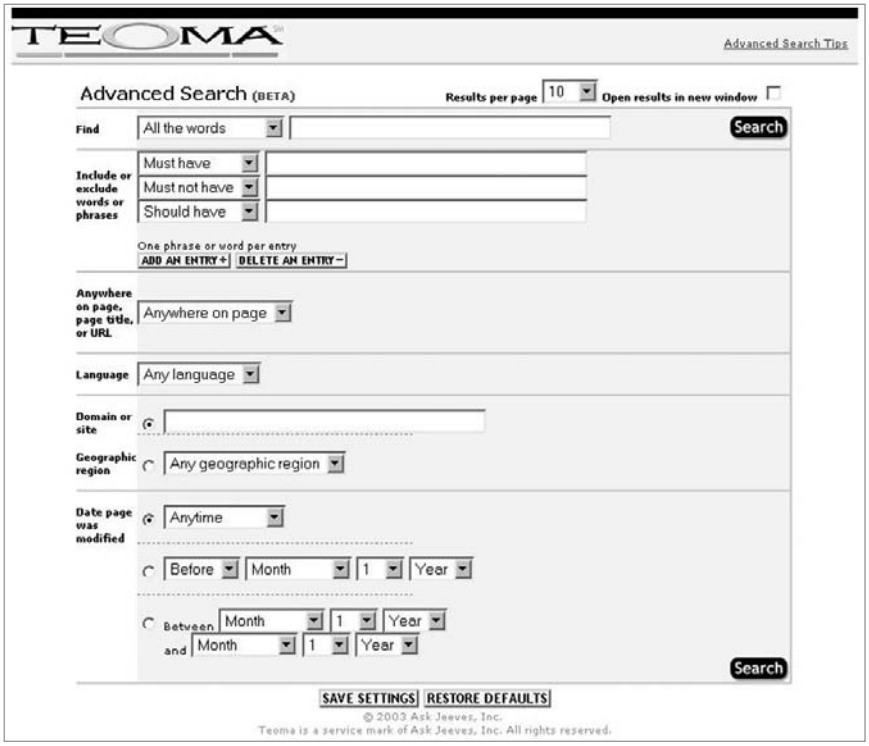
Example: paris lang:french

The following search options are available.

Title Searching

To search for pages with a particular term in the title, you can use either of these methods:

Figure 4.18



Teoma's Advanced Page

1. On the advanced search page, enter your terms in one of the search boxes and then choose "in page title" from the "Anywhere on page, page title, or URL" menu.
2. On the home page, use the "intitle:" prefix.
   Example: intitle:progesterone

URL

In Teoma, to find pages from a specific URL, you can use the following procedures:

1. On the advanced search page, enter the URL in one of the search boxes and then choose "in URL" from the "Anywhere on page, page title, or URL" menu. This will enable you to find all pages from the URL. If you want to

do a "site search" for a particular term or terms, enter the terms in the search boxes and then enter the URL in the "domain or site" box. However, combining terms and a URL in Teoma seems to be significantly less effective so in other search engines.

2. On the home page, you can use the "inurl:" prefix.

   Example: inurl:ssu.edu

If you want to search for a term(s) within a site, use the term in combination with the "site:" prefix.

Example: biology site:ssu.edu

Language

To limit retrieval to one of 10 languages, on Teoma's advanced search page, enter your terms in the search boxes and then choose the language from the languages menu.

You can also use the "lang:" prefix.

Example: lang:swedish

Geographic Region (Continent)

To limit retrieval to pages from a particular geographic location (continent), on Teoma's advanced search page, use the "Geographic region" menu.

You can also use the "geoloc:" prefix,

Example: ibm geoloc:europe

Date Searching

To limit retrieval by the date a page was modified, on Teoma's advanced search page you can use the "Date pages was modified" menu and either choose a time frame such as "Last 3 months," or you can specify before, after, or between the dates you select in the date boxes.

For dates, there are also these prefixes: "last:," "afterdate:," "beforedate:," and "betweendate:," but it is much simpler to use the date searching on the advanced search page.

Boolean

All terms you enter in Teoma's main search box are automatically ANDed, unless you otherwise qualify them. You can use simple Boolean by means of pull-down windows on its advanced page.

OR can be used in the search box, *but* if you try to use it with any terms you wish to AND, using the implied AND, it will not produce meaningful

results. For example, a search expression in the form of "A B OR C" will not give you either combination that might logically be expected.

You can accomplish a NOT by use of the minus sign.

Example: labor OR labour -pregnancy

Teoma Results Pages

Teoma delivers three kinds of results on its results pages:

1. Web pages. These are typical search engine results listings, from Teoma's own database. Because, like other search engines, Teoma clusters results, look for the "More results from …" link to get to additional matching pages from any site.
2. Refine. These are suggested narrower searches.
3. Resources. This section of Teoma results is the most unique, and for many searchers it is the most important part of the results page. Sites listed here are those that Teoma has identified as containing a collection of links on the topic searched. As a result, many or most of these are specialized directories. Because of this feature, Teoma is probably the best place on the Internet to locate specialized directories.

Special Features

Spell-check

Like Google, Teoma does a spell-check. For words that look like they might be misspelled, you will get a suggestion to that effect on results pages.

OTHER GENERAL WEB SEARCH ENGINES

The Web search engines covered in this section are engines that the serious searcher needs to be aware of. However, they either no longer or do not yet offer any particularly compelling reasons to go into the level of detail provided for the more major engines just discussed.

Lycos

http://lycos.com

Lycos has positioned itself as more of a portal than primarily as a search engine. It is a very good portal, providing a good collection of resources, including news, multimedia, and other specialized searches; downloads; job

listings; phone directories; weather; and other features. It provides a search engine, but the database used is the same database as is behind AllTheWeb (FAST), which is more searchable using the AllTheWeb interface. Lycos' search has both a home page and advanced version. The home page version has minimal search features (+word, -word, " "). The advanced search provides more options, using menus. The Lycos home page is personalizable and Lycos also provides over 20 country/language-specific versions. To get to these, click on the Visit Terra Lycos Worldwide link at the bottom of Lycos' home page.

WiseNut

http://www.wisenut.com

WiseNut was one of two new general Web search engines to come on the scene in mid-2001. (The other is Teoma.) Although it claims to search over 1.5 billion pages, WiseNut retrieves fewer records than should be expected. (This assessment is based on some brief benchmarking, and you may see WiseNut catching up.) WiseNut's most outstanding feature is its "WiseGuide Categories" that appear on results pages and are generated based on semantic relationships of words in your search. These categories allow easy and effective narrowing of search results by subject. WiseNut does not have an advanced mode. A Preferences page allows choice of limiting searches to particular languages, number of results per page, unclustering of results, display of WiseGuide categories, and an adult-content filter. Since WiseNut creates its own database, if you absolutely have to find everything on a topic, include WiseNut in your list of engines to be searched.

MSN Search

http://search.msn.com

MSN is Microsoft's entry in the search engine market. The database it uses is Inktomi, the same database used by HotBot. However, all Inktomi database versions and the way they are searched by different Inktomi partners are not the same. Searches on HotBot often yield substantially more than on MSN. MSN Search's advanced search page allows for simple Boolean, stemming (variant word endings), continent, language, domain, document depth (within the Web site), and type of content included (images, JavaScript, etc.). The fact that there is nothing particularly unique in its offering and that a more effective

search of the Inktomi database can be done elsewhere means that it need not usually be thought of as an essential tool in the serious searcher's toolbox.

SPECIALTY SEARCH ENGINES

In addition to the general Web search engines discussed here, numerous specialty search engines are available. Some are geographic, focusing on sites from one country, and some are topical, focusing on a particular subject area. To locate these, try the following category in Open Directory (at http://dmoz.org, or under Google's Directory tab):

Computers > Internet > Searching > Search Engines > Specialized

METASEARCH ENGINES

Metasearch engines are services that allow you to search several search engines at the same time. With one search you get the results from several engines. (They should not be confused with "metasites," which is another term for specialized directories, as discussed in Chapter 3.) Considering the emphasis earlier in this chapter that was placed on using more than one engine, the metasearch idea seems compelling—and it is indeed a great idea. However, the reality is often something else. You may find that you like a particular metasearch engine and have legitimate reasons for using it, but it is important to note some particularly important shortcomings of which you need to be aware.

First, though, it should be noted that this section addresses the free sites on the Web that allow the searching of multiple engines. Additionally, there are metasearch programs (software) that can be purchased and loaded on your computer to aid in the searching of multiple engines. These "client-side" programs do a much more complete job, but involve the downloading (and eventually purchasing) of a program and sometimes several more steps to get to your results. These programs go beyond what the Web metasearch engines do, and can effectively search a variety of Web search engines, sort out the results, allow further local searching, and perform a variety of related tasks. Most frequently noted among these are Copernic and BullsEye. Particularly if you need to repackage search results to deliver to a client, the purchase of one of these programs should be considered.

Back to the metasearch engines on the Web, they are numerous. New ones frequently appear and older ones disappear as quickly. Among the better known

are DogPile, ixquick, vivisimo, MetaCrawler, and Search.com. They can cover portions of large numbers of search engines and directories in a single search and they can sometimes be useful in finding something very obscure.

However, each metasearch engine usually presents one or more, and sometimes all, of the following drawbacks:

1. They may not cover most of the larger search engines. (If you have a favorite metasearch engine, see if it covers Google, AllTheWeb, AltaVista, HotBot, and Teoma.)
2. Most only return the first 10 to 20 records from each source. If record number 11 in one of the search engines was a great one, you will probably not see it.
3. Most syntax does not work. Some metasearch engines may allow you to search by title, by URL, and so on, but most do not. Some do not even recognize even the simplest syntax: the use of quotation marks to indicate a phrase.
4. Many present paid listings first.

Also, by now you know that on search engine results pages, the additional content presented (besides just the listing of Web sites) can often be very valuable. You lose this with metasearch engines.

If you find that a metasearch engine meets your needs, by all means use it. However, they are not the solution for doing an exhaustive—or even a moderately extensive—search.

KEEPING UP-TO-DATE ON WEB SEARCH ENGINES

To keep up-to-date with what is happening in the realm of Web search engines, take advantage of the sites listed in the section "Keeping Up-to-Date on Internet Resources and Tools" in Chapter 1, but also look at the best known search engine news site on the Web, Search Engine Watch.

Search Engine Watch

http://searchenginewatch.com

This site is maintained by Danny Sullivan, a leading journalist in the area of Web search engines, along with Chris Sherman, noted speaker and writer on the topic. The site provides up-to-date news and reports in a clear and readable style. It is a valuable resource for both the search engine user and Web site

developer. Access to much of the content on the site is free, but more in-depth material is available for a small subscription fee. A free bi-weekly newsletter is available. For those who want to keep up on a daily basis, Search Engine Watch also provides SearchDay, a daily update by Chris Sherman.

Search Engines Features Chart

Table 4.2

	AllTheWeb Home	AllTheWeb Advanced	AltaVista Home	AltaVista Advanced
SIZE (pages)	2.1 billion	2.1 billion	1.3 billion	1.3 billion
SIMPLE BOOLEAN	*term term (Defaults to an AND) -term (for NOT)*	(menu) term term (Defaults to an AND) -term (for NOT)	*term term (Defaults to an AND or to a phrase) -term (for NOT)*	*term term (Defaults to an AND or phrase) -term (for NOT*
FULL BOOLEAN OPERATORS	*(term1 term2) equals an OR*	*and or andnot*	OR AND AND NOT ( )	OR AND AND NOT ( )
PHRASE	(menu) " "	(menu) " "	" " some automatic	" " automatic
PROXIMITY				NEAR (=within 10 words)
TRUNCATION			*term** (asterisk - internal or at end of term)	*term** (asterisk - internal or at end of term)
TITLE FIELD	title:*term*	(menu)	title:*term*	title:*term*
DATE FIELD		menu		(date range boxes)
URL FIELD	url:*term* *site:term*	(menu)	url: *term* domain: *term* host: *term*	url: *term* domain: *term* host: *term*
"LINKS TO" A URL	link:*term*	(menu)	link: *term*	link: *term*
LANGUAGE	language:xx e,g., language=fr	(menu)	Radio buttons for "All languages" or Default language	Radio buttons for "All languages" or Default language
MEDIA SEARCHING	Links to News, Pictures, Videos, MP3, FTP searches	News, Pictures, Videos, MP3, FTP all have their own advanced mode	links to media searches image: *term*	image: *term*
CASE SENSITIVE	no	no	yes (if placed in quotes)	yes (if placed in quotes)
SEARCHES ALL COMMON WORDS	yes	yes	yes	yes
WEB DIRECTORY ATTACHED			yes (Open Directory)	
ADDITIONAL OUTPUT FORMAT OPTIONS	Can set # of results etc. by using "Preferences" page	Can set # of results, etc. by using "Preferences" page	Can set # of results, etc. by using "Settings" page	Can set # of results, etc. by using "Settings" page
PAID SITES FIRST	Yes (Identified as "Sponsored Links")	Yes (Identified as "Sponsored Links")	Yes (Identified as "Sponsored Matches")	Yes (Identified as "Sponsored Matches")
SIMILAR PAGES LINK ON RESULTS PAGES			yes ("Related Pages")	yes ("Related Pages")
ALSO SHOWN ON RESULTS PAGES	Headlines Clusters Multimedia matches Similar queries	Headlines Clusters Multimedia matches Similar queries	Headlines Categories ("Prisma") "Others searched for" Links to maps, etc.	Headlines "Others searched for" Links to maps, weather, etc.
OUTSTANDING SPECIAL FEATURES	Multimedia search Adult content filter Extensive news search Search by region Indexes PDF, Flash	Adult content filter Extensive news search Search by region Indexes PDF, Flash	Images/audio/video search Translations Adult content filter Hit terms highlighted Indexes PDF	Images/audio/video search Translations Adult content filter Indexes PDF

Search Engines Features Chart

Table 4.2 continued

	Google Home	Google Advanced	HotBot Home (Inktomi)	HotBot Advanced
SIZE	3 billion	3 billion	2 billion	2 billion
SIMPLE BOOLEAN	*term term (Defaults to an AND) -term (for NOT)*	term term (Defaults to an AND) -term (for NOT)	*term term (Defaults to an AND) -term (for NOT)*	(menu) *term term (Defaults to an AND) -term (for NOT)*
FULL BOOLEAN OPERATORS	OR Defaults to an AND	(text boxes) Defaults to an AND		
PHRASE	" "	" " (text boxes)	" "	(menu) " "
PROXIMITY	*term * term (1 asterisk =1 intervening term)*	*term * term (1 asterisk =1 intervening term)*		
TRUNCATION				
TITLE FIELD	intitle:term allintitle:term1 term2	intitle:term allintitle:term1 term2	title:term	(menu – "Word Filters") title:term
DATE FIELD		(menu)		(Date box)
URL FIELD	inurl:term allinurl:term1 term2 term1 site:term2	inurl:term allinurl:term1 term2 term1 site:term2 (menu - "Domain" window)		(menu) (Region menu)
"LINKS TO" A URL	link:*term*	link:*term* *("Links" textbox)*		(menu)
LANGUAGE	Can set using Preferences page	(menu)		(menu)
MEDIA SEARCHING	Tabs for Images search			(checkboxes)
CASE SENSITIVE	no	no	no	no
SEARCHES ALL COMMON WORDS	yes	yes	yes	yes
WEB DIRECTORY ATTACHED	yes Tab links to Open Directory	no	no	no
ADDITIONAL OUTPUT FORMAT OPTIONS	10,20,30,50,100 results Can set using Preferences page	10,20,30,50,100 results Can set using Preferences page	10,20,30,40,50 results and other user selections (Preferences page)	10,20,30,40,50 results and other user selections (Preferences page)
PAID SITES FIRST	Yes (Identified as "Sponsored Links")	Yes (Identified as "Sponsored Links")	Yes (Identified as "Sponsored Links")	Yes (Identified as "Sponsored Links")
SIMILAR PGS LINK ON RESULTS PAGES	yes	yes	no	no
ALSO SHOWN ON RESULTS PAGES	Open Directory categories and sites Link to cached page Translation option Link to definitions, maps Headlines, stock quotes Addresses & phone #'s	Open Directory categories and sites Link to cached page Translation option Link to definitions, maps Headlines, stock quotes Addresses & phone #'s	Option to transfer search to other engines' databases	Option to transfer search to other engines' databases
OUTSTANDING SPECIAL FEATURES	Popularity ranking Newsgroups/ images Cached pages Adult content filter Covers PDF & other formats News search	Popularity ranking Newsgroups/ images Cached pages Adult content filter Covers PDF & other formats News search	Easy transfer of advanced search to Google, AllTheWeb, Teoma	Easy transfer of advanced search to Google, AllTheWeb, Teoma

Table 4.2 continued

Search Engines Features Chart

	Lycos Home	Lycos Advanced	Teoma Home	Teoma Advanced	WiseNut wisenut.com
SIZE	Uses the AllTheWeb database 2.1 billion	Uses the AllTheWeb database - 2.1 billion	1 billion	1 billion	1.5 billion (?)
SIMPLE BOOLEAN	term term (Defaults to an AND) -term (for NOT)	(menu) term term (Defaults to an AND) -term (for NOT)	*term term (Defaults to an AND) -term (for NOT)*	(menu) *defaults to AND*	*term -term defaults to AND -term (for NOT)*
FULL BOOLEAN OPERATORS	*(term1 term2) equals an OR*	*(term1 term2) equals an OR*	*OR*		
PHRASE	" "	(menu) " "	" "	" "	" "
PROXIMITY					
TRUNCATION					
TITLE FIELD		(box, under "Word Filters")	intitle:*term*	intitle:*term*	
DATE FIELD				(menu)	
URL FIELD		(boxes, under "Word Filters")	inurl:*term* *site:term*	(menu)	
"LINKS TO" A URL		(boxes, under "Link Referrals" tab)			
LANGUAGE		(radio buttons, under "Language" tab)	lang:*language*	(menu)	(menu on Set Preferences page)
MEDIA SEARCHING	Link to "Multimedia" search	(radio button)			
CASE SENSITIVE	no	no	no	no	no
SEARCHES ALL COMMON WORDS	yes	yes	no, but can force with a + (e.g., +the)	no, but can force with a + (e.g., +the)	
WEB DIRECTORY ATTACHED	(link to Open Directory)	(link to Open Directory)	no	no	no
ADDITIONAL OUTPUT FORMAT OPTIONS			Can set # of results, etc. using Preferences page	Can set # of results, etc. using Preferences page	Can set # of results, etc. using Preferences page
PAID SITES FIRST	Yes (Some identified as Sponsored Links")	Yes (Some identified as Sponsored Links")	Yes (Identified as Sponsored Links")	Yes (Identified as Sponsored Links")	Yes (Identified as Sponsored Links")
SIMILAR PGS LINK ON RESULTS PAGES			yes (Related pages)	yes (Related pages)	
ALSO SHOWN ON RESULTS PAGES	Directory hits Links to quotes, news, travel links, etc. Matching news Related searches		Categories Resource guides	Categories Resource guides	
OUTSTANDING SPECIAL FEATURES	Hit terms highlighted Adult content filter	Adult content filter Downloads, MP3, news, newsgroups	Resource guides	Resource guides	Automatic categorization of results

GROUPS AND MAILING LISTS

WHAT THEY ARE AND WHY THEY ARE USEFUL

Groups, newsgroups, mailing lists, and other online interactive forums are tools that are often under-used resources in the searcher's toolbox. Particularly for competitive intelligence (including researching and tracking products, companies, and industries) and for other fields of intelligence (including security, military, and related areas), newsgroups and their relatives can be gold mines (with, analogously, the product often difficult to find and to mine).

Groups, mailing lists, and a variety of their hybrids represent the interactive side of the Internet, allowing Internet users to communicate with people having like interests, concerns, problems, and issues. Unlike regular e-mail, where you need the address of specific persons or organizations in order to communicate with them, these channels allow you to reach people you don't know and take advantage of their knowledge and expertise. This chapter outlines the resources available for finding and mining this information and some techniques that can make it easier.

A major barrier to understanding these tools is the terminology. Newsgroups have little to do with "news" and mailing lists are definitely not to be confused with the junk mail you receive in either your e-mail or traditional mailbox. "Newsgroups," narrowly defined, usually refers to the Usenet collection of groups that actually originated prior to the Internet as we now think of it. "Groups," more broadly defined, includes newsgroups and a variety of other channels, variously referred to as groups, discussion groups, bulletin boards, message boards, forums, and even (by dot.com marketers, primarily) as "communities."

The biggest distinction between groups and mailing lists lies in how the information in them gets to you. With groups, messages are posted on computer networks (e.g., the Internet) for the world to read. Anyone can go to a group and read its content and, usually, anyone can post a message. Mailing list

content goes by e-mail only to individuals who subscribe to the list. With groups, you have to take the initiative each time to go get the messages; with mailing lists, the messages come automatically to you. If you go to some type of Web page to look at a posting, it is probably a group. If you get it in your e-mail, you are probably looking at a mailing list. One further important distinction is that messages that appear in groups are usually more fully archived and, therefore, more retrospectively available than the content of mailing lists.

Both groups and mailing lists can be moderated or unmoderated. With unmoderated groups (and lists), your posting appears immediately when you submit it. If the group or mailing list is moderated, your posting must pass the inspection of someone who decides whether to approve the posting, and, if approved, then submits it for publication to the list. Among other things, this means that moderated groups and lists are more likely to have postings that really are directly related to the subject.

GROUPS

Collections of groups originate from, and are found in, a variety of online collections, including the grandparent of all groups, Usenet; in commercial portals such as Yahoo!; and on professional association sites, among others. The next few pages will give an overview of the nature of these various collections and how you can most easily access them and participate.

Usenet

Usenet is the original and still best known collection of groups, created in 1979 at the University of North Carolina and Duke University by Jim Ellis, Tom Truscott, Steve Bellovin, and Steve Daniel. Usenet (a "*use*rs' *net*work," originally spelled USENET, but now just Usenet) started as a collection of network-accessible electronic bulletin boards and grew quickly both in terms of use and geographic reach. Not only does Usenet predate the Web, it predates the Internet as most of us know it today. With the popularization of the Internet and the Web, however, Usenet access is now, for all practical purposes, through the Internet, and most users use Web-based interfaces rather than the older specialized software known as news readers. (If you bump into any Usenet old-timers, be sure to let them know that you know that Usenet is not "part" of the Internet, but it is accessible through the Internet.)

Usenet groups are arranged in a very specific hierarchy, which at first glance appears a bit arcane. The hierarchy consists of 10 main top-level categories and thousands of other top-level hierarchies, based mainly on subject, geography, and language. Each hierarchy is further broken down (otherwise, they wouldn't be "hierarchies").

Examples: sci.bio.phytopathology
rec.crafts.textiles.needlework

The main top-level hierarchies are:

- **alt.** For "alternative", as in alternative lifestyle or alternative press. This is the "anything goes" category, and the creation of a new group in this category does not require the clearly defined nominating and voting process that is required for other hierarchies.
- **biz.** Business
- **comp.** Computers
- **humanities.** Humanities
- **misc.** Subjects that don't fit neatly into the other main categories
- **news.** Formerly primarily news relating to Usenet, but now that plus a variety of odds and ends
- **rec.** Recreation (sports, games, hobbies, etc.)
- **sci.** Science
- **soc.** Social sciences
- **talk.** Political and social issues, among others

The messages within each individual group are arranged by "threads"—series of messages on one specific topic consisting of the original message, replies to that message, replies to those replies, and so on. Users can post messages to either the original message or to any of the replies, or they can start a new thread.

Accessing Usenet Groups

News Readers

Until probably the late '90s, most Usenet access was through an Internet Service Provider (ISP), and messages were read and posted by means of special software called news readers or through such software built into browsers such as Netscape. ISPs received newsfeeds from the computers that hosted Usenet groups, and made that content available to the ISP's customers. Coverage (those groups that were stored by the ISP) was usually selective, due

to the large volume of Internet traffic. If your ISP did not provide access to the group in which you were interested, you usually merely had to ask and the ISP would add the group. Although this process still happens, some major ISPs today no longer support Usenet access.

News readers are very similar to e-mail programs, allowing you to both read and post messages. You select from your ISP's list those groups to which you wish to subscribe, and when you wish to view postings to a group, or post a message yourself, you click on the name of the list in your news reader, and recent postings are delivered to your computer. ("Subscribe" here means that you wish for that group to be on the list of groups for which your ISP sends you messages. For most groups, there is no official membership.)

The preceding paragraph will probably best be treated as history, but it is useful because you will still run across news readers and, to be conversant in Internet terminology, you should probably know what they are. However, most people who don't have a lot of time on their hands will probably be better off getting their Usenet access by means of the Web through their browser.

Web access to Usenet newsgroups first became widely available through a site called Deja News, created in 1996 and later became "deja.com." It was great—until the people responsible for its design and marketing began to miss the point and decided to make it into a shopping site, with the newsgroup access relegated to a minor position. Deja.com went out of business and can now best be remembered as an early pioneer of the dot.com bust.

To the rescue comes none other than almost-every-serious-searcher's favorite site, Google. In 2001 Google bought Deja's remains, began loading the archive, and quickly added the capability to not just search Usenet postings but to post messages as well. By the end of the year, it had made a 20-year archive of Usenet postings available. By 2002 the argument could be made that Google provided the easiest and most extensive capabilities ever for both the average user and the serious researcher to access and participate in newsgroups.

Other Groups

Although Usenet is the best-known collection of groups, it is not the only one. Groups can be found on commercial sites and portals such as Yahoo!, Delphi Forums, and ezboard. You will also find a lot of specialized groups on association and club sites, such as, for example, the U.S. Bicycle Racing

Association, the Institute of Electrical and Electronics Engineers, and the Welsh Rugby Union. These Web-based groups vary considerably in terms of the appearance of the interface, but they all function in about the same way. You can read, post, follow threads, and so on. Unlike Usenet, you usually have to sign up to use these groups, and to varying degrees may be required to identify yourself, although often your e-mail address is about the only identifying information required. On sites such as Yahoo! Groups, you must sign up for a Yahoo! password, and on most association sites, you must be a member of the association to participate in the discussions.

Resources for Locating and Using Groups

The following resources (and others) can be used for locating *groups* of interest or *messages* on specific topics of interest:

- Google's Groups Search. For searching Usenet groups and postings
- Google's Web Search. For locating some non-Usenet groups
- Yahoo! Groups. Only for groups on the Yahoo! site, but there are hundreds of thousands of them
- Delphi Forums. Over 100,000 active forums on the Delphi Forums site (http://www.delphiforums.com)
- Ezboard. Can be found at http://www.ezboard.com

Sites of Associations Related to Your Topic

Look on the site of an association related your search topic for an indication of a "forum," "discussion," or similar term suggesting the presents of a group. This tip also applies to locating mailing lists.

USING GOOGLE TO FIND GROUPS AND MESSAGES

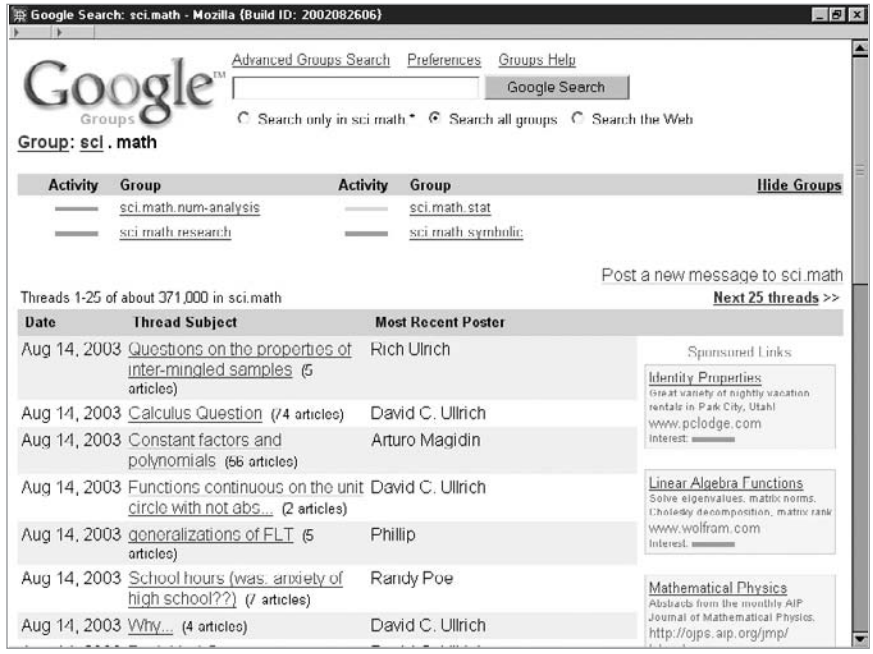
At present, Google (http://google.com) is, by far, the best tool for finding, accessing, and using the Usenet collection. It contains not just current postings, but archives going back over 20 years (to 1981) and over 800 million messages. As with the rest of Google, the online documentation for Groups is exceptionally good. The following is a quick overview and some highlights.

Browsing the Hierarchies

From Google's main Groups page (click on the Groups tab on Google's main page to get to it), you can browse down through the 10 main top-level hierarchies. To get to the other top-level hierarchies (listed alphabetically), use the "Browse complete list of groups …" beneath the top ten list. It will provide you with a pull-down window that allows you to get quickly to the appropriate place in the alphabet. Notice that at the top of the screen a search box with radio buttons allows you to search either all of Google Groups or just within the current hierarchy (see Figure 5.1). Limiting your search to the current hierarchy is a powerful capability, especially when dealing with an ambiguous term. However, don't be too quick to do so if your subject could appear in more than one hierarchy.

As you browse down, at each level you will see links to additional lower levels in the hierarchy and also specific threads posted at that level. Clicking on the latter will take you to the messages themselves.

Figure 5.1



Google Groups: Browsing Within a Hierarchy

Searching Google for Groups and Messages

When you use the search box on either the main groups or advanced groups search page, Google will retrieve all groups that have your term(s) in one of the sections of the name, plus any messages containing that term. When you search from the main Google Groups page, you can, if you wish, use prefixes, such as group:britain, or author:smith to specify a particular group or author. However, for easy yet sophisticated queries, you are probably better off just clicking on the Advanced Groups Search link and using the powerful combination of options found there.

Figure 5.2

Google's Advanced Groups Search Page

The Advanced Groups Search page (see Figure 5.2) allows you to specify the following attributes:

- Simple Boolean expressions:
  - Use the "with all of the words" box to specify a Boolean AND
  - Use the "with at least one of the words" box to specify an OR

- Use the "without the words" box to eliminate records that contain a particular word (NOT)

- Messages from a specific newsgroup only. You can use the full name or any section (e.g., soc.religion.shamanism or religion.shamanism or shamanism)
- Subject, author, message ID, language, date range

Search results are ranked by relevance, but you can click the "Sort by date" link on the results page to view the newest first.

Figure 5.3

Google Groups: Message Thread

Messages

When you view a message, you can see a single message or, by clicking on the "View entire thread" link above the message, a clickable list of all messages in the thread, who posted them, and when (see Figure 5.3).

Retrieved messages contain: title (linked to the entire message), date and excerpt of text around retrieving words, a link to the group in which it was found, and a link to the thread.

Transit in Bratislava

I recently returned from Bratislava, Slovakia, and I must say that their tram system continues to impress me. Clearly, it is one of the best I have seen. Trams ...

misc.transport.urban-transit - 08 Apr 2001 by Christopher Rivituso - View Thread (11 articles)

Posting Messages Using Google Groups

To post a message, browse through the hierarchy until you get to a group of interest. At the top of the list of messages in the group you are viewing, you will see a link that will enable you to start a new message in that group. When viewing a message, a link at the bottom of a message enables you to post a reply. To either start a new message (thread) or reply to a message, you will be required to register with Google. After clicking on the "post new message" or "reply" link, you will be presented with a form to fill in, including your e-mail, the name you wish to use when posting, and a password of your choice. There is also a fairly long Terms agreement. Google will respond with a confirmation e-mail, you click the confirmation link presented there, and you are ready to go. It is a good idea to read the "Posting Style Guide" that you will find linked on message pages. (Note: When you do a search and click on a message, the posting option will not appear. You must either browse the hierarchy or click on a group link somewhere.)

Using Google's Web Search for Identifying Non-Usenet Groups

Although it doesn't advertise the fact, Google's Web search actually picks up over 2 million postings from non-Usenet groups. It doesn't work perfectly, but you can locate some of these groups in Google by using this trick: As part of your search, use the phrase "next thread." Because that phrase is present in many messages from other groups, it will locate messages from a variety of sources that have been indexed by Google because of the existence of a group on a Web site.

YAHOO! GROUPS

After Google, the second place to look to find groups is Yahoo! (at http//:groups.yahoo.com, or look for Groups in the Connect section of Yahoo!'s home page). These groups are actually a hybrid of groups and mailing lists,

because for each group you can receive messages either at the Yahoo! Web site or by e-mail. Yahoo! allows you to search or browse through the groups, to post messages, and to create groups of your own. There are hundreds of thousands of Yahoo! groups, some with thousands of members and many with but a single member. (Join one of the latter and brighten someone's day.) There are 48 "duct tape" groups alone.

Finding Groups on Yahoo!

You can find Yahoo! groups of interest by either browsing through the 16 categories on the main groups page, or by using the search box found there. Be aware that a search there only searches names of groups and their descriptions, not individual messages. Yahoo! does automatically truncate, though, so a search for "environment" will also retrieve groups that have "environmental" in their title or description. Terms you enter in the search box are automatically ANDed. Unlike regular Yahoo!, you cannot use "-word" to exclude a word. Also, unlike almost every other search box you will find on the Internet, you cannot search a phrase by using quotation marks.

Whether you use the search box or browse the categories to find groups on Yahoo!, the listing of groups that results will contain the name of the group, the description, whether there is an archive, and if the archive is public or not. If it is public, you can browse through the messages without joining the group. (Click on the word "Public" to view the list of messages.) Clicking on the name of the group from the group listing will show you more detail about the group, including when it was founded, whether membership is open, whether it is moderated, and so forth, plus a calendar showing numbers of messages posted each month (see Figure 5.4). The number of members and volume of postings are usually important indicators of the potential usefulness of the group.

Joining a Yahoo! Group

Registration is free. After identifying a group of interest, check first to see that the membership is open. To join, click on Join This Group. You will be asked for your Yahoo! Password, and if you do not have a Yahoo! password, you can get one at this point. After joining, Yahoo! Groups will send you an e-mail message containing a link for you to click that confirms that the e-mail address you used is really your own. Once you have confirmed this and selected delivery methods, you can go to the home page for the group and read

Figure 5.4

Yahoo! Group Description Page

and post messages. Also, after you have registered, you can click on My Groups to set your e-mail preferences:

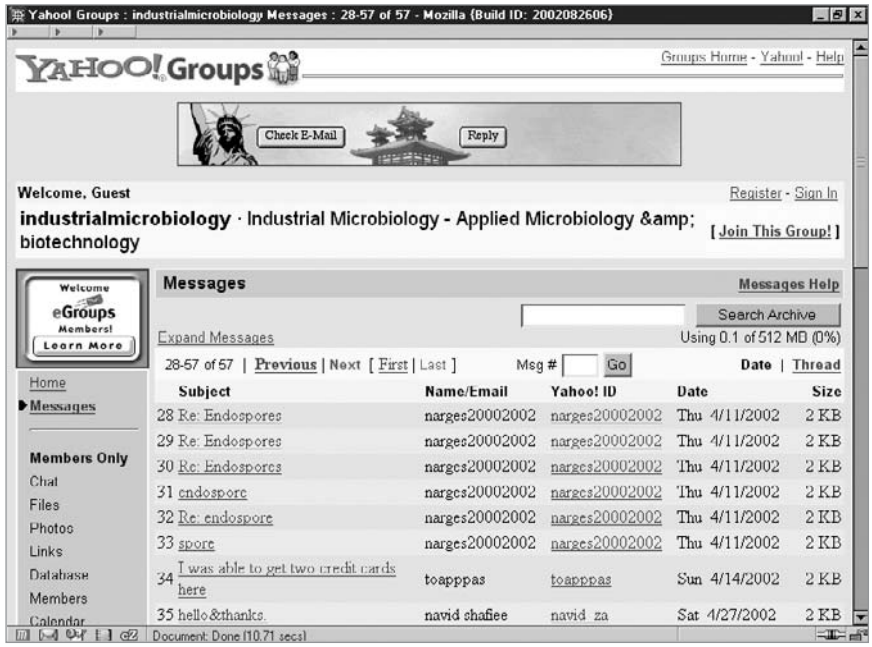
- Individual e-mails. To receive individual e-mail messages
- Daily digest. To receive multiple e-mails in one message
- Special notices. To receive update e-mails from the group's moderator
- No e-mail. The true groups approach, where you go get the messages, rather than receiving them by e-mail

Once you have joined one or more groups, when you sign on, you will be presented with a page providing links to all of the groups to which you belong. You also use this page to unsubscribe to groups. Look for the Edit My Groups link.

On both the pages that list messages and on the message pages themselves, you have a good variety of navigation and viewing options, such as moving backward and forward in a thread, viewing by thread or by date, viewing the message in brief (collapsed) form or expanded (full message) form, viewing the sender's profile, searching the archive, among others. The latter is where you have the opportunity to search the actual content of messages. Each group

also is provided with an option for the following features and content: Chat, Files, Photos, Polls, Links, Database, Members list, Calendar. The database, particularly, is a very powerful tool for sharing information.

Figure 5.5



List of Yahoo! Group Messages

Starting a Group on Yahoo!

Yahoo! is definitely one of the easiest—and perhaps actually *the* easiest—place on the Web to set up a group. A group of your own can be great for a course you are teaching, networking and support groups, family, and so forth, and you can get one set up in 10 minutes or less.

You choose the category (although Yahoo!'s staff may change the category if they see it and feel the category is inappropriate), name the group, decide if it is to be public, moderated, and so on. Basically, all you have to do is fill in the blanks.

With Yahoo!'s large number of users and members, large number of groups, ease of use, and accessibility to both those who want to use and those who

want to sponsor groups, Yahoo! is an indispensable resource for those who can make use of the Internet as a communications channel.

OTHER SOURCES OF GROUPS

There are numerous other places where you will find groups, some large and some small, but most have considerably less reach and less content than those available through Yahoo! or Google. Nevertheless, the group that may precisely meet your needs may be in one of the smaller collections. Two additional sources, not as large as Yahoo!, but having both a large number of groups and members are Delphi Forums and ezboard.com.

Delphi Forums

http://www.delphiforums.com

According to Delphi Forums itself, this site has over 4.5 million registered users, 100,000 active forums, 50,000 messages each day, and over 40 million total messages. As with Yahoo! Groups, with Delphi you can read most messages without registering, but to post messages you must register. Registration is easy, but asks for more information than you may feel is Delphi's business. (I, myself, would never, of course, advocate dishonesty of any type, but let me just say that when some sites ask for information than is inappropriate, it inclines some people to use some "latitude" in their response to questions.)

Delphi's lists are browsable using the 21 categories on the home page, and searchable by using the search box just above the category list. Terms entered in that box will search messages, but, interestingly, the names of groups containing your term in the group name are not retrieved. Also be aware that the search results may pop up in a window that may show in minimized form, causing you to think that nothing happened.

You can create a free forum on Delphi, but you have to have one of Delphi's advanced memberships to do so (so it's "almost" free). Membership fees, though, are extremely inexpensive. If it is really important that you find as many groups out there on your topic as possible, don't ignore a search on Delphi Forums.

ezboard

http://www.ezboard.com

ezboard has 1 million forum "communities" and over 10 million registered users. It is searchable and browsable using a very detailed multilayered

categorization. (Click on the "find" or "forum find" links to get to the list.) Terms entered in the search box will be searched both in forum titles and in the messages themselves. Registration is free, and there are inexpensive upgrade options. Maintaining a forum of your own here requires one of the upgraded memberships.

MAILING LISTS

Most of what can be said about the usefulness and nature of groups also applies to mailing lists. As mentioned earlier, the biggest differences are that with mailing lists, (1) the message arrives in your e-mail rather than you having to request to see messages, with every message sent to the list coming to you, (2) you have to subscribe, often providing identifying information (and may need to be a member of the sponsoring organization), (3) the content of mailing lists is less likely to be archived and searchable than for groups, and (4) although the e-mail delivery mode makes it easier to access and ensures that you don't miss anything important, mailing lists postings can fill up your mailbox and can be a nuisance to deal with. The comparison is analogous to a company bulletin board compared to the inbox on your desk. Some information is more appropriately accessed by your going to the bulletin board periodically, whereas for some information, you would prefer to get a copy on your own desk. If, on a particular topic, you want to make sure you don't miss anything, a mailing list may better serve you.

The receipt and distribution of messages on mailing lists are controlled automatically by "listserver" software, and lists are often referred to, inappropriately, as "listservs." Listserv® is a registered trademark for listserver software produced by the L-Soft company, and the term (legally) should not be applied generically. The other most frequently encountered mailing list managers are Majordomo and Listproc.

Once you have located a mailing list of interest, you need to subscribe in order to participate. (How to find lists is coming up soon.) Some sites, for example, lots of association sites, provide a nice Web interface where you just have to fill in the blanks. Other sites provide instructions for sending an e-mail message to the mailing list administrative address and tell you what command you need to put in the header or message in order to join. For example, you might be instructed to send a message to majordomo@alektorophobia.org with the message *subscribe fearofchickens* in the body of the message. The

instructions will vary primarily dependent upon the listserver software being used. You will usually receive a reply confirming your membership to the list and referencing an information file explaining how to use the list, ground rules, and so on.

The following are other important points to know about using mailing lists:

1. The e-mail address to which you send administrative messages is different from the one you use for posting messages. It is a great annoyance to list members to see administrative messages in their mailboxes.
2. Many lists offer delivery of a "digest" form in which a number of messages are bundled on a regular basis (e.g., daily or weekly). This is especially useful for lists that have a lot of traffic and digests can avoid clogging up your e-mail inbox. They may also have an option where you can suspend delivery while you are on vacation.
3. Many (probably most) lists will provide an FAQ (Frequently Asked Questions) file or Web page. It will usually be worth scanning this.
4. Some lists provide archives, many of which are searchable.
5. Before you sign up, note (from descriptions you find of the list) the level of traffic. If you subscribe to several high-volume mailing lists, you will end up not being able to read them because of the hundreds of messages you receive. For high-volume lists, consider taking advantage of (1) digest versions and (2) "on vacation" options.

Tools and Techniques for Locating Mailing Lists

For many people, their first experience in using mailing lists is through organizations to which they belong. Numerous other lists of interest may be out there and, fortunately, some online sites make them easy to find. Among these are Topica.com and L-Soft Catalog. Yahoo! can also be added to this list because, as pointed out earlier, its groups also have an e-mail option.

Topica.com

Topica.com's thrust is providing mailing list services to companies, associations, and individuals. Many readers who use mailing lists may have noticed that instead of associations managing their own lists, many are taking advantage of this service. Topica (formerly liszt.com) hosts thousands of e-mail newsletters. In addition to association lists and lists created by individuals, many

of Topica's lists are commercial, but keep in mind that these latter are opt-in lists—you only join if you want to. They can be valuable for competitive intelligence purposes, as well as keeping up to date on products and special deals from your favorite suppliers.

You can search by list topic without signing up, but signing up will enable you to participate and to search individual messages as well as list descriptions. To sign up, you provide your e-mail address and a password of your choice. You then receive a confirming e-mail to which you need to reply in order to complete the process. Sign up for some lists and, thereafter, when you go to Topica and log in, your page will show your lists. From there you can subscribe, unsubscribe, and set mail preferences. The latter includes options to receive your mail in digest form or receive messages by going to the Topica Web site (more like groups).

Lists of interest can be identified either by using the search box or browsing through the Topica categories. To browse, click on one of the 16 categories at the bottom of Topica's home page, or better, click the More option there. The resulting page will give you a better idea of coverage of the categories.

You can search using the search box on the main page or on the categories pages. Once you are two levels down in the categories or on search results pages, the search box provides an option of searching either Lists (names and descriptions) or the content of messages themselves. Topica allows you to use AND, OR, and NOT (capitalization is not necessary). If you do not use any operator between words, Topica defaults to an OR. You can also use quotation marks to search for phrases. You may want to narrow your search using these techniques, because Topica returns a maximum of 200 matches.

The list descriptions given usually make it easy to determine if this is a list for you (see Figure 5.6). The description pages also make it easy to read and subscribe to (join) the list. On those and other pages you will find how to (very easily) start a list of your own. (First ask yourself, "Does the world really need my list?")

L-Soft CataList, the Official Catalog of LISTSERV® Lists

As the name says, L-soft (http://www.lsoft.com/lists/listref.html) is the official catalog for the 61,000 public lists that use LISTSERV software. In addition to searching list names and descriptions, you can also view lists by host country, view only those with 10,000 subscribers or more, or those with 1,000 subscribers or more.

Figure 5.6

Topica List Description

ONE MORE CATEGORY—ONLINE INSTANT MESSAGING

Online instant messaging, pioneered by AOL as Instant Messenger, with variations by MSN, Yahoo!, and others, is another incarnation of online interaction of people and is a hybrid of groups and e-mail. Although at first populated mainly by teenagers and an extension of the historic evolution from hanging out on the street corner to occupying the family phone, instant messaging has now gone beyond the teenage realm. In some corners of the corporate world, it is beginning to take the place of popping your head into someone's office. It promises to potentially be one more significant productivity enhancer.

If you haven't used or seen it, the way it works is that participants create a buddy list of people they want to interact with online on an immediate basis. You send a message to someone on your list, and it will pop up on their screen. People who use the same instant messaging service who are not yet a buddy but who want to talk to you can send you a message asking to talk. You also have the option of creating a chat room in which multiple people are invited to join the conversation. This technology is well developed, but not yet in the

forefront of professional communications. It is, however, moving in that direction. Borrow a local kid to show you how it works.

SOME NETIQUETTE POINTS RELATING TO INTERNET GROUPS AND MAILING LISTS

Readers of this book most likely already have a good sense of Netiquette (Internet etiquette), but some may profit by these selected points relating to groups and mailing lists:

1. Lurk before you leap. Lurking, or hanging around just observing a discussion without participating, is definitely a good idea. It may involve just reading a few messages or a few threads, and you may find yourself ready to leap in and join the conversation in a matter of minutes. Read enough messages (and preferably the FAQ or similar documentation) to be sure that the conversation is at the level appropriate to your needs and knowledge. If a group is very technical, they get annoyed at beginners asking extremely simple questions. If there is a searchable archive, check it out. Don't be caught trying to start a discussion of a topic that has already been beaten to death.
2. Don't use either newsgroups or e-mail lists for advertising. Depending on the group or list, there might be times when it would be acceptable to respond to a posting that may have requested a service you might provide, but be careful. You can easily make a lot of people mad at you. In such a case, you can play it safer by responding directly to the poster by e-mail, rather than responding to the group or list as a whole.
3. Don't get sucked into a flame war, an angry or unnecessarily strongly worded series of messages (a.k.a. flaming). Remember the sad truth that there are people out there who have nothing better to do than wasting their time being nitpicky, rude, and generally obnoxious. The advent of groups and lists has become a wonderful channel for their frustrations and repressed feelings.
4. Only forward messages if allowed. Some associations, particularly, have rules regarding privacy of messages, often relating to such things as client privilege and competitive intelligence. Follow those rules very carefully. This mistake can cause you to be banned from a group—and worse.
5. Use cross-posting (posting the same message to multiple groups or lists) advisedly. It clutters up peoples' mail and time.

AN INTERNET REFERENCE SHELF

All serious searchers have a collection of tools they use for quick answers—the Web equivalent of a personal reference shelf. The challenge is to make sure that you have the right sites on your shelf. This chapter provides a selective collection of such sites that should be in most researchers' collections. Different researchers have different quick-reference needs requiring different tools. For many of us, we may have found out about most sites through a friend or through just stumbling across them. The attempt here is to highlight reference tools that provide quick answers to some of the most frequently requested information, from the mundane to the esoteric.

This chapter goes hand-in-hand with Chapter 5. Indeed, for subject areas of interest to you, many of the resource guides of the types covered in Chapter 5 should be in your reference collection, in the same way that the reference section of a library usually contains a good collection of resource guides. On the other hand, in addition to quick-answer sites, a number of resource guides for reference tools in particular areas such as statistics, government information, and companies are also included.

Going from general to specific, we look first at some prime general tools, such as encyclopedias, and then move in the direction of tools that can provide very specific bits of information. For many of the categories, as well as listing specific sites, some suggestions will be provided about effectively using the resources.

Remember that all of the links presented here, as well as links for all sites covered throughout this book, are available at http://extremesearcher.com.

THINKING OF THE INTERNET AS A REFERENCE COLLECTION

Especially for people with broadband connections, going to the Internet instead of printed resources for frequently sought-after information is

becoming more and more common. With practice, it becomes quicker and easier (and in some cases such as telephone directory assistance, a lot cheaper). The biggest trick is, first, simply understanding the range of quick-reference tools that are out there, and, second, getting in the habit of using them—remembering to use them and bookmarking them. Another trick is not to fall into the trap of always going to the Internet first. (I have an encyclopedia right behind me that I often grab rather than grabbing the keyboard.)

The tools listed in this chapter provide a start in making sure the reader has a sense of the breadth and variety of quick-answer sites. The next step in understanding the range of these tools is to spend some time browsing one of the several reference resource guides listed at the end of this chapter. Plan to spend at least 20 minutes poking around in these sites. Almost anyone can find something new and interesting in them.

SOME SITES ALL RESEARCHERS SHOULD KNOW ABOUT

Criteria Used for Selecting the Following Tools

Selection of the tools covered here was based on several factors. The first factor is the experience as a long-time Internet user and former reference librarian as well as experience in observing and talking with thousands of Internet users from a variety of kinds of organizations and many countries. The second factor is the measure of a site's utility to a wide range of users. Some sites were chosen because they provide good examples of the range of these tools, and others were chosen because they provide examples of particular features to look for when examining and using reference sites. In several instances, multiple sites serve basically the same function (such as the travel reservations sites). More than one of such types are included in order to point out the differences and the utility of using more than one, rather than choosing a favorite and always going there.

Traditional Tools Online

A number of online tools are electronic versions of common printed tools, including encyclopedias, dictionaries, almanacs, and the like. These

are excellent for quick answers and for background relating to more specific research. In these (and many other tools), a number of factors contribute to their usefulness. These factors are important to know in some circumstances, irrelevant in other circumstances, and often are the same ones to be considered when using printed reference tools.

- Does the tool contain everything that the printed version contains? Encyclopedia.com contains everything the printed version does (and more), whereas the free online version of Encyclopedia Britannica contains only a small portion of the printed content.
- Does it contain things the printed version does *not*? Many of these tools provide collections of links and often news headlines that the hardcopy version cannot provide.
- How current is it? Bartlett's Familiar Quotations (available as part of Bartleby.com) is the 1919 edition.
- Is all of it free? Or is a fee required to access part of the content? For many of the tools that require a subscription, the subscription fee is not too costly, and you may find the expenditure worthwhile.

The annotations for the sites discussed here are purposely brief and not intended to be reviews of the sites, but they do include the major points that researchers should consider when determining whether to use the tool. ("… too great brevity of discourse tends to obscurity; too much truth is paralyzing." Blaise Pascal. Quote located by using Bartleby.com.)

ENCYCLOPEDIAS

Encyclopedia.com

http://encyclopedia.com

Encyclopedia.com includes 57,000 frequently updated articles from the Seventh Edition of the Columbia Encyclopedia. Unlike Encarta and the britannica.com (Encyclopedia Britannica), all articles are free. Articles can be located either by browsing alphabetically or by searching (using the search box). When searching, terms are automatically "ORed," but you can specify an AND between words or use quotation marks for phrases. Encyclopedia.com also provides links to news and magazine articles through eLibrary, but these involve a fee.

Figure 6.1

Article from Encyclopedia.com

Encarta

http://encarta.msn.com

The more than 42,000 articles here are usually lengthier than the articles found in encyclopedia.com. Some are free, but access to many articles requires a subscription. When searching, terms are automatically ANDed and you cannot use Boolean operators or quotation marks.

Voila Encyclopédie avec Hachette

http://encyclo.voila.fr

This extensive French encyclopedia contains 50,000 articles and a dictionary—all free. You can limit your search to articles, media (photos, maps, video), the atlas, or the dictionary.

Encyclopedia Britannica

http://britannica.com

The online version of the renowned Encyclopedia Britannica, this site provides *very* short articles for free, but the vast majority of the content requires a subscription. Considering the quality of this encyclopedia, you may find that

buying a subscription is well worth the price. You can either browse or search, and results include the encyclopedia articles, many free magazine articles, and carefully selected Web sites.

DICTIONARIES

YourDictionary.com

http://www.yourdictionary.com

YourDictionary.com is a resource guide that provides links to over 600 dictionaries in 260 languages as well as to a variety of resources relating to languages. It includes multilingual dictionaries and specialized subject dictionaries, including technical and scientific dictionaries. The quality and extensiveness of the dictionaries varies, but for most languages you will have a number of dictionaries to choose from. When you consider that few libraries in the world have as many language dictionaries on their shelves as this site brings to your fingertips, you can better understand the potential of the Web as a reference resource.

Dictionaries—Selected Examples

In addition to taking advantage of the resource guide YourDictionary.com, it may be worthwhile to bookmark one dictionary for each language you are most likely to encounter. Here are some recommendations:

Merriam-Webster Online

http://www.m-w.com

A full-featured English dictionary with pronunciation (with audio), part of speech, etymology, and a thesaurus. Give the word games a try, as well. An unabridged version with an atlas and other tools is available for a subscription fee.

Dictionnaire Universel Francophone En Ligne

http://www.francophonie.hachette-livre.fr

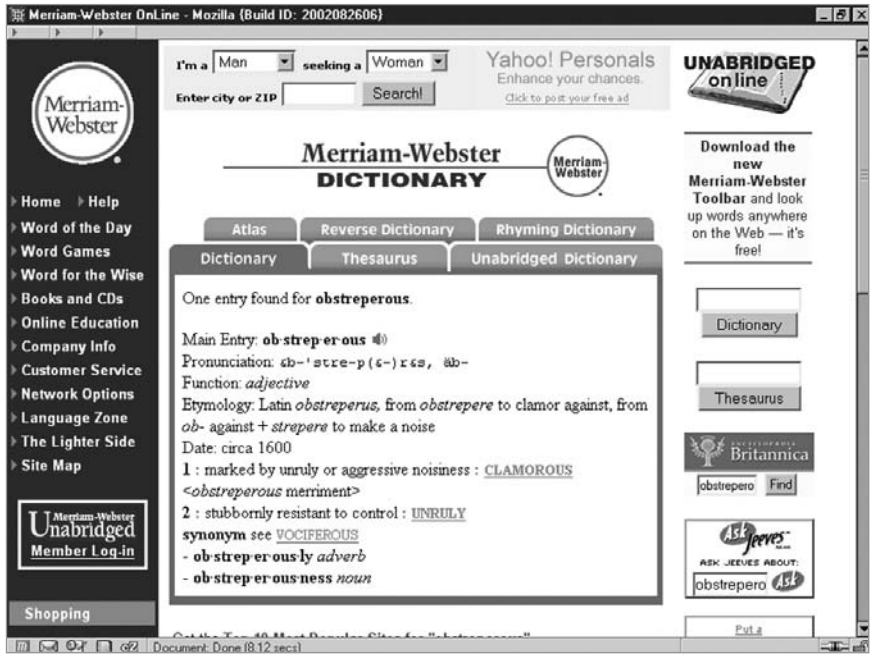
This French dictionary includes 45,000 words, 116,000 definitions, and 13,000 proper names.

diccionarios.com

http://www.diccionarios.com

This general Spanish dictionary also includes translations between Spanish and English, French and Catalan. It contains 95,000 entries.

Figure 6.2



Definition from Merriam-Webster Online

LEO—Link Everything Online

http://dict.leo.org

LEO contains over 300,000 entries, and although it does not provide a complete dictionary entry for a term, it does provide a quick English/German, German/English lookup. You will also find here a nice display of usage and idiomatic expression examples, with links to such things as declension tables.

ALMANACS

InfoPlease

http://www.infoplease.com

No brief description can substitute for spending time exploring this site, which is much more than just an almanac. Explore each of the main sections: Daily Almanac, World, United States, History & Government, Biography, Sports, Arts & Entertainment, Business, Society & Culture, and Health & Science. The site contains the almanacs, an encyclopedia (Columbia Encyclopedia, 6th ed.), Infoplease Dictionary (125,000 entries), Infoplease Atlas, biographies (30,000 of them), and more. Lots of little gems such as the (greatly abridged) extracts from

the Encyclopedia of Associations, chronologies, statistics, calendars, country profiles, and so on. For non-U.S. users, the "World" link will move you away from the U.S. orientation of the home page. One of the many interesting features is the Cite link, which shows how to cite the item being viewed. When using the main search box, the search is performed automatically on all of the almanacs, the encyclopedia, and the dictionary. Terms you enter are ORed, but items with all of your terms (AND) will be listed first. Quotation marks can be used to search phrases. By using the pull-down window by the search box, you can limit your search to specific almanacs, biographies, the dictionary, or the encyclopedia.

ADDRESSES AND PHONE NUMBERS

There are numerous places to go on the Web for phone numbers and addresses—worldwide. For a specific country, start by identifying the available directories by using a resource guide such as Wayp International White and Yellow Pages. Do not expect 100 percent success (or even 70 percent) in any of the directories. Some of the directories may be incomplete or a bit dated. However, if you ordinarily use telephone directory assistance, these sites will allow you to find lots of people a lot easier and a lot less expensively. For U.S. phone numbers, Google may be your best bet (http://google.com). When searching, remember that names may be listed in a variety of ways, such as with first initial, instead of the full first name.

Wayp International White and Yellow Pages

http://www.wayp.com

This site is a straightforward worldwide collection of white and yellow pages by country. Some of the yellow pages are Internet-only (without an equivalent printed version) and may be fairly limited. Some are quite extensive. Searchability and extensiveness of white pages listed here likewise vary considerably.

Yahoo! People Search

http://people.yahoo.com

For finding addresses, phone numbers, e-mail addresses, and so forth, the People link on Yahoo!'s main site covers U.S. numbers and addresses and also has an e-mail address search that works reasonably well.

For non-U.S. numbers, use the relevant country-specific Yahoo! site and look under the "Communication" or "Connect" links for the telephone directory (e.g., Communica Cerca persone; Kommunicera Personsök)

AnyWho

http://www.anywho.com

This site has U.S. yellow pages and white pages plus a good collection of links to international directories (under International Links). For U.S. numbers and addresses, when you get the results for a person's phone number or address, click on the street name, and you will be given a list of other people on the street. Clicking on Maps and Directions will lead you to a MapBlast map for the address. Also, this site is where you can find a reverse phone lookup. If you have a phone number and don't know who it belongs to, click the Reverse Lookup tab, enter the number, and you will probably get the owner.

QUOTATIONS

Quote Links

http://www.quotationspage.com

Quote links is a resource guide that provides a searchable database of over 15,000 quotations, a subject directory, and links to other sites for various kinds of quotes: famous, humorous, scientific, and others. This furnishes a good source when preparing a talk, an article, or a paper. Quote something from Lucius Accius, and people may think you have actually read his works.

Bartleby

http://www.bartleby.com

Bartleby also belongs in the category of "just plain amazing" sites. Chief among its quotation sources are Bartlett's Familiar Quotations (1919 edition), Columbia Quotations, and Simpson's Quotations, but it also contains a wonderful collection of other sources of quotes, handbooks, anthologies, collected works of famous authors (including Shakespeare), and other reference tools. The content is primarily humanities, but Bartleby even throws in some science. The contents of all of these resources can be searched together, or you can use the pull-down windows from the main page or on the Reference, Verse, Fiction, or Nonfiction tabs to individually search over 200 full-text works. The following list is just a selection of what is available at Bartleby.com:

Columbia World of Quotations. 1996.
Simpson's Contemporary Quotations. 1988.
Oxford Book of English Verse. 1919.
Yale Book of English Verse. 1919.
Columbia Encyclopedia, 6th ed. 2001.
Columbia Gazetteer of North America. 2000.
The World Factbook. 2001.
American Heritage® Dictionary of the English Language, 4th ed. 2000.
Roget's II: *The New Thesaurus*, 3rd ed. 1995.
Roget's International Thesaurus of English Words and Phrases. 1922.
Bartlett, John. *Familiar Quotations*, 10th ed. 1919.
The Columbia World of Quotations. 1996.
Simpson's Contemporary Quotations. 1988. ("The most notable quotations since 1950.")
American Heritage® Book of English Usage. 1996.
The Columbia Guide to Standard American English. 1993.
Fowler, H. W. *The King's English*, 2nd ed. 1908.
Quiller-Couch, Sir Arthur. *On the Art of Writing*. 1916 and 1920.
Quiller-Couch, Sir Arthur. *On the Art of Reading*. 1920.
Sapir, Edward. *Language: An Introduction to the Study of Speech.* 1921.
Strunk, William, Jr. *The Elements of Style*. 1918.
The Bible. King James Version. 1999.
Brewer, E. Cobham. *Dictionary of Phrase and Fable*. 1898.
Bulfinch, Thomas. *The Age of Fable*. 1913.
Frazer, Sir James George. *The Golden Bough: A Study in Magic and Religion*, abridged ed. 1922.
Cambridge History of English & American Literature (18 vols.). 1907–1921.
Eliot, Charles W., ed. *The Harvard Classics and Harvard Classics Shelf of Fiction*. 70 volumes. 1909–1917.
Eliot, T. S. *The Sacred Wood*. 1920.
Shakespeare, William. *The Oxford Shakespeare*. 1914.
Van Doren, Carl. *The American Novel*. 1921.
Gray, Henry. *Anatomy of the Human Body*, 20th ed. 1921.
Farmer, Fannie Merritt. *The Boston Cooking-School Cook Book*. 1918.
Post, Emily. *Etiquette*. 1922.
Inaugural Addresses of the Presidents of the United States. 1989.
Robert, Henry M. *Robert's Rules of Order Revised*. 1915.

Figure 6.3

Bartleby.com

TIP:

To identify quotations, if it sounds like a famous quotation, try Bartleby first. If you don't find it there, try a search engine and search for the quote as a phrase. Bartleby has the advantage of greater authority, while the search engines have wider reach and more current material.

FOREIGN EXCHANGE RATES/CURRENCY CONVERTER

If you travel internationally, have family or friends living in other countries, or purchase items from locations outside the U.S., you may frequently need to know the U.S. dollar equivalent of a certain foreign currency. There are numerous sites on the Web that do these calculations. Yahoo!'s is one of the best.

Yahoo! Finance—Currency Conversion

http://finance.yahoo.com/m3?u

Although Yahoo!'s Finance page does not make this part of the site easy to find, and only seven people in the world are likely to memorize the obscure exact address, if you bookmark this site, you will discover that it provides a convenient table for some major currencies and a conversion calculator that handles over 150 currencies. In the chart, if you click on the name of the currency, you will get a graph showing fluctuations.

WEATHER

Both for local weather and for travel planning, a good weather site is essential. A good option is your own personalized portal, such as My Yahoo! or My Netscape, that allows you to automatically have in front of you the weather for the cities you specify. If you use a news site as your start page, look there to see if you can select weather for specific cities. If you want a weather-only site, try Weather Underground.

Weather Underground
http://wunderground.com

Click on the appropriate region of the world map to get weather for a particular continent. One of the advantages of this site over other weather sites is that you can personalize it by choosing which map (U.S., Europe, etc.) will be shown automatically and to choose which cities to be shown (worldwide). You can also choose temperature to be in either Fahrenheit or Celsius, or both.

MAPS

The Perry-Castañeda Library Map Collection
http://www.lib.utexas.edu/maps

The Perry-Castañeda Library Map Collection is a tremendous collection of maps plus links to gazetteers and so forth. Most of the over 5,000 maps available here are in the public domain and no permission is required to copy or distribute them. A large portion of the maps are produced by the CIA. The site also has a fascinating collection of historical maps. Take time to read the FAQ, especially for the useful tips on printing the maps. The General Libraries at the University of Texas should be thanked profusely for providing this resource.

GAZETTEER

Global Gazetteer
http://www.world-gazetteer.com

Latitude, longitude, current population, size rankings, and other statistics are available at the World Gazetteer site for countries, administrative divisions, cities, and towns. The pronunciation table it provides for dozens of languages will be useful not just here but for other applications.

ZIP CODES

U.S. Postal Service

http://www.usps.com/zip4

If you have the street address, the U.S. Postal Service site will give you the nine-digit ZIP Code. The City/State/ZIP Code Associations page link at the bottom of the home page provides a list of ZIP Codes associated with a particular city.

STOCK QUOTES

As with many other frequently asked reference questions, numerous places on the Web can help you find stock quotes. For the searcher who needs stock quotes frequently, it will be worthwhile to investigate several sites and determine the best for you by looking at such things as ease of use, clarity of presentation, detail provided, personalized portfolio features, types of charts and graphs available, and presence of associated news stories. As with weather information, consider using a general portal, such as Excite or Yahoo!, that can integrate selected stock information and a personalized portfolio into your start page. Remember that, for free, you are usually going to get quotes that are delayed by 20 minutes. If you use a major brokerage house or an online trading service, look at their site. You may qualify to sign in as a client and receive real-time data and order capabilities. CNN is the example chosen here of one of the many places to get free stock information online.

CNNMoney

http://money.cnn.com

CNN's site is very rich, with not just detailed stock quotes but financial news, company backgrounds, a currency converter, e-mail newsletters, financial tools, and many other kinds of market-related information. You can set up and track your own portfolios for free. Streaming real-time data is available for a subscription fee.

STATISTICS

Although not every statistic you might want will be available on the Internet, finding a statistic by using the Internet makes locating a needed statistic amazingly easier than just a few years ago. The expanse of statistical information is

immense, as is the amount that can be said about finding statistics on the Internet. A few very basic hints and resources are provided here. For more detail, there are excellent books available on the topic. Although it is a bit dated, still one of the best books on the topic is *Finding Statistics Online: How to Locate the Elusive Numbers You Need*, by Paula Berinstein (CyberAge Books, 1998). For help in finding business statistics, refer to Berinstein's more recent book, *Business Statistics on the Web: Find Them Fast—At Little or No Cost* (CyberAge Books, 2003).

In general, because the topic of statistics is so broad, you are often best off by starting with one of the numerous resource guides. Other than resource guides, only a couple of specific sources for the most commonly sought statistics are given here.

In the way of a few hints for finding statistics, keep the following in mind:

- There are three main ways of finding statistics on the Internet:
    1. Go to a site you think may contain the statistic and search or browse. For example, try the relevant governmental department (e.g., Department of Agriculture for agricultural statistics). Think about what agency or other organization would have an interest in collecting the kind of data you are looking for.
    2. Go to a collection of links to statistics sites (such as those listed later).
    3. Use a general Web search engine such as AllTheWeb or Google. More and more statistical material is indexed by search engines than was the case one or two years ago, especially because of the indexing of PDF files, Excel spreadsheets, and other document types. A search strategy can often be very straightforward. Try a combination of one or two *subject terms* plus the *place* and perhaps, the *year*.

  Example: avalanche fatalities norway 2002
- Good news: There is plenty of redundancy of identification and access; in other words, there are many routes online to the same statistic.
- When you find a statistics site you might use again, *bookmark it!* To make it a lot easier to use bookmarks, create folders to organize similar types of sites.
- On statistics sites, take advantage of site search boxes and site maps.
- Watch out for terminology. Unless you are familiar with the topic, the terminology may not be obvious. Terms such as "housing starts" may not be what you immediately think to look for.

Statistical Resources on the Web—Comprehensive Subjects

http://www.lib.umich.edu/govdocs/stcomp.html

Statistical Resources on the Web—Comprehensive Subjects is one of many excellent specialized directories maintained by the University of Michigan Documents Center, and is one of the best starting places for finding collections of statistics.

Statistics.com

http://www.statistics.com

Another good resource guide, Statistics.com is a searchable directory of statistics available on the Internet, plus a variety of other statistics-related links.

The Directory of Online Statistics Sources

http://www.berinsteinresearch.com/stats.htm

Hosted by the author of the books mentioned previously, this site is an especially useful resource guide because it lists not just free Internet resources, but also commercial online sources (which contain a lot of statistics not available free on the Internet).

U.S. Statistics

USA Statistics in Brief

http://www.census.gov/statab/www/brief.html

This site contains selected tables from the venerable *Statistical Abstract of the United States* and includes summary tables for a broad range of subjects plus basic state population data.

FedStats

http://www.fedstats.gov

FedStats contains links to statistics produced by over 100 U.S. Federal agencies. You can browse or use the search feature to search across agencies.

Books

Most book searches on the Internet fall into one of two categories: (1) finding information *about* books—in other words, what books are available on a particular topic or by a particular author—or verifying bibliographic information, or (2) trying to locate the entire book online. Unless the book was

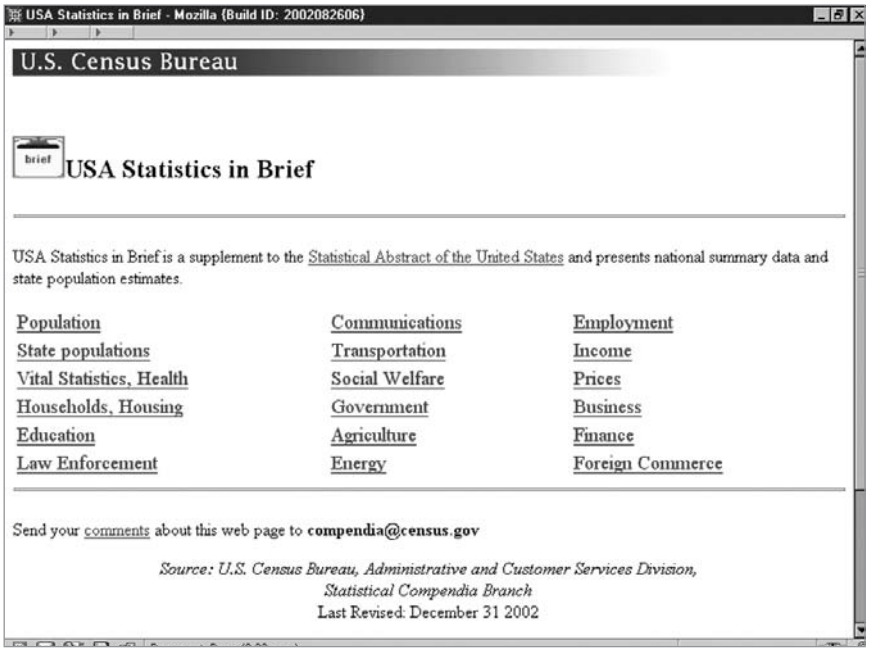


Figure 6.4

USA Statistics in Brief

published close to a century ago, you shouldn't expect to get it in full text online. Nevertheless, thousands of books are out there in full text and the number is growing rapidly. If the book is by a famous pre-20th-century author, you have a pretty good chance of finding the full text online.

Finding Information About Books—Bookstores

Keep in mind that the large online book vendors' sites are good not just for buying books, but also for identifying books currently in print on any topic—or for that matter, out of print, but still available for sale by used- or rare-book dealers.

Amazon.com

http://www.amazon.com

Amazon lists millions of titles for sale and offers good discounts. The site is searchable by author, title, subject, ISBN, or publication date and is browsable by subject. Click on Books, then on More Search Options, for this advanced searching. Amazon now also includes millions of used, rare, and out-of-print

books, from hundreds of booksellers, that were previously available at http://bibliofind.com. For rare books, click on the Books tab, then look for the Around Our Bookstore section.

Barnes & Noble

http://www.barnesandnoble.com

Competing nose-to-nose with Amazon, Barnes & Noble also provides access to millions of books (and other merchandise). The Quick Search box on the home page allows you to search by author, title, or subject. Click the Search button to get to a larger selection of searchable fields. It also has a collection of 25-million out-of-print, used, and rare books from dealers around the world (click on the Out-of-Print tab).

Finding Information About Books—Bibliographic Databases

To find what books have been published at any time on any topic, go to the online catalog of one of the major national libraries. For English materials (although, of course, they are not limited to English materials), you might start with either the British Library or the U.S. Library of Congress.

Library of Congress

http://catalog.loc.gov

The Library of Congress online catalog includes 12-million records for books, serials, computer files, manuscripts, maps, music, and audio-visual materials. The Basic Search option allows a search by title, author, subject, call number, and keywords and by LCCN, ISSN, or ISBN numbers. Guided Search provides limiting to date of publication, language, type of material (e.g., book, serial, music), location of the collection within the Library of Congress, place of publication, and 17 other fields. OR, AND, and NOT can be used by means of the pull-down windows and radio buttons. Also, you can truncate by the use of a question mark at the end of the word.

British Library

http://blpc.bl.uk

The British Library site provides not just a search of the 10 million items in the British Library Public Catalogue and other British Library online catalogs (serials, manuscripts, newspapers, sound archive), but also lets you

order photocopies or a fax or request loans. The main page allows searching by author/editor, organization, title, subject, publisher, ISBN/ISSN, and date of publication. You can use an asterisk to truncate a term. The Advanced Search provides the same searchable fields, but with Boolean capabilities added.

Your Local Library's Online Catalog

For something more local, check the online catalog of your local library. If it has a Web accessible catalog, you'll find it easily through a search engine, or you can find it by going to the Library of Congress' Gateway to Library Catalogs at http://lcweb.loc.gov/z3950/gateway.html.

Full-Text Books Online

If you are trying to locate a specific work online, a search engine usually works quite well. However, it may be easier to use a site that compiles a large number of such works and that will usually enable you to browse by title or author. Bartleby.com provides over 200 full-text books, including a number of useful reference works. But for over 6,000 available books, consult Project Gutenberg, and for over 20,000 titles, look at The Online Books Page. In these collections, the vast majority of works are no longer under copyright. Therefore, with a few exceptions, they are from before the 1920s. (Unfortunately, the increased availability of 20th-century texts threatens to be slowed by attempts by both the EU and the U.S. Congress to extend copyright virtually into perpetuity). The sites discussed here are definitely sites to which the word "amazing" must be applied. Whether you want to find Cicero or The Bobbsey Twins, these are good places to start.

The Online Books Page

http://digital.library.upenn.edu/books

This site contains links to over 20,000 books in English. The creator and editor of the site, John Mark Ockerbloom, founded it in 1993 and has been adding to it constantly since then. For books to be included, they must be in English, in full text, and must qualify as significant by being listed in the online catalog of a major library or be otherwise recognized. The site is easily searchable by title and is browsable by author, title, subject, and serial title.

Figure 6.5

The Online Books Page

Project Gutenberg

http://www.promo.net/pg

Project Gutenberg, begun in 1971, aims to place online, in easily accessible format, as many public domain electronic texts (eTexts) as possible. So far, it has provided over 6,000 texts, and the collection is growing rapidly. Although the majority of books covered are in English, Project Gutenberg contains (a few) books from 14 other languages. The breadth of texts available makes this an excellent research site, but also consider it as a source of eTexts for reading on your laptop. Because most of the books are stored in ASCII text, many are small enough to be loaded on a floppy disk.

Bartleby.com

http://www.bartleby.com

For a list of the books covered by Bartleby.com, see the earlier discussion related to quotations.

Historical Documents

EuroDocs: Primary Historical Documents From Western Europe

http://library.byu.edu/~rdh/eurodocs

A resource guide, the EuroDocs site provides links to Western European documents that are online in transcribed, facsimile, or translated form. They are arranged by country and, within that, chronologically.

A Chronology of U.S. Historical Documents

http://www.law.ou.edu/hist

The Chronology of U.S. Historical Documents site contains links to over 150 full-text documents from the pre-colonial period to the present.

University of Virginia Hypertext Collection

http://xroads.virginia.edu/~HYPER/hypertex.html

The University of Virginia offers a collection on its site of classic and other texts in the area of American Studies, including books and journals.

Governments and Country Guides

In lots of situations, information about specific countries is needed—basics such as population, names of leaders, flags or maps, or more detailed information on economics, geography, and politics. Numerous resources provide this type of information, and those resources differ primarily in terms of amount of detail and categories of data covered.

Governments on the WWW

http://www.gksoft.com/govt

Governments on the WWW is an excellent resource guide. Arranged by continent and country, the links on this site connect you to official government sites (including individual sites for parliaments, offices, courts, embassies), banks, multinational organizations, and political parties.

Foreign Government Resources on the Web

http://www.lib.umich.edu/govdocs/foreign.html

The resource guide Foreign Government Resources on the Web provides links to government sites by country, but also by topic, such as constitutions, embassies, and flags.

CIA World Factbook

http://www.odci.gov/cia/publications/factbook

This annually revised work provides easily usable and quite detailed country guides. Each country's data is arranged in the following sections: Geography, Communications, People, Transportation, Government, Military, Economy, and Transnational Issues. Also notice the Chiefs of State link on the main page. This is an extremely rich site, and even if you do not think you will use it frequently, you will find spending some time exploring it worthwhile. As an indication of how widespread the respect for this site is, the "Basic Facts on Iraq" section of the official site of the former Permanent Mission of Iraq to the UN (http://www.iraqi-mission.org) was mostly taken word-for-word from the CIA World Factbook!

CountryWatch.com

http://www.countrywatch.com

CountryWatch.com contains country-specific geopolitical intelligence on 191 countries. In addition to the background information on the country, the site also displays current newswire headlines relating to the country. Part of the content is free, but part requires a subscription. Use the pull-down window or click on a continent, then the country, to display data and the current headlines relating to that country (from Radio Free Europe and other sources). The free information you will find here is in considerably less detail than that provided by the CIA World Factbook.

U.S. GOVERNMENT

FirstGov

http://firstgov.gov

FirstGov is the official Internet gateway to U.S. Government resources and is a good starting place for locating information from or about government agencies. Note the four main divisions of the site: Citizens, Businesses and Nonprofits, Federal Employees, and Government-to-Government.

GPO Access

http://www.gpoaccess.gov

Use this site to search the Federal Register, Code of Federal Regulations, Commerce Business Daily, Congressional Record, Government Manual, and other databases, either singly or together.

THOMAS: Legislative Information on the Internet

http://thomas.loc.gov

THOMAS consists of a variety of detailed and easily searchable databases containing information on Federal legislation. It also contains the Senate and House Directories and links to other government information. This is an excellent place to start a search on what legislation is in process, legislation on a specific topic, or tracking a particular current bill.

U.S. STATE INFORMATION

Library of Congress—State and Local Governments

http://lcweb.loc.gov/global/state/stategov.html

The Library of Congress State and Local Government directory is a resource guide containing a convenient collection of links to state government information, state maps, and so forth.

TIP:

For official state sites, use the following URL "recipe":

http://www.state.pc.us (pc=postal code)

Example:

http://www.state.md.us

U.K. GOVERNMENT INFORMATION

UK Online

http://www.open.gov.uk

Ukonline is a searchable and browsable collection of information, news, and links to U.K. public sector information, including both central and local government information and links.

BASIC RESOURCES FOR COMPANY INFORMATION

Whole books—many of them—have been written on finding company information on the Internet. Anyone who searches for company information frequently will want to spend some time with one of those books and may already be familiar with the quick-reference company sites included here. For those who have only occasional need for company information or who are just getting into the area, the following sites will provide a start (see also Chapter 5).

First, we cover a few basic pointers about tools for finding company information. To know where to go for company information, it helps to start by thinking about what kinds of company information you might reasonably expect to find on the Internet. You might think of three categories:

1. Information that a company *wants* you to know, such as its stature, its products or services, and any good news about the company.
2. Information that a company *must* let you know, such as that required by government laws and regulations (for example, Securities and Exchange Commission filings in the U.S. and Companies House filings in the U.K.).
3. *What others are saying* about the company.

For what a company *wants* you to know, start with the company's home page. Depending upon the company, you will probably find detailed background, products and services, company structure, press releases, and so on. To find a company's home page, you can usually count on just entering the name in any of the largest search engines. The company home page will usually be among the first few items retrieved.

For what a company *must* let you know, first keep in mind that this applies only to publicly held companies. Others typically do not *have* to divulge very much information publicly. For U.S. publicly held companies, SEC filings are available through several sites, but Hoovers, a major company directory discussed later, makes these filings available very conveniently along with a lot of other useful data about a company. For public companies in other countries, the amount of mandated information is usually much less than that required of U.S. companies, but start by looking in CorporateInformation.com.

For the third category of company information, *what others are saying* about a company, some items to keep on your Internet reference shelf are newsgroup resources (especially Usenet groups as available through Google Groups and other groups sources discussed in Chapter 5) and news stories (such as through MSNBC, CNN, and BBC). For some key news sites, see Chapter 8.

These resources, however, are primarily useful for finding information about a specific company you already have in mind. Many company questions are along the lines of: "What companies are there that match a particular set of criteria." For example, What are some of the largest seafood packers in Maryland? What is the name of a plumber who serves my neighborhood? These questions are often answerable by the use of company directories or online yellow pages of the types listed earlier in this chapter.

Company Directories

Company directories on the Web differ in terms of:

- Number and type (public, private, U.S., non-U.S.) of companies included

- Free, paid subscription, or pay-per-view
- Searchability (name, industry location, ticker symbol, size, etc.)
- Amount of information provided about each company (usually the more companies included, the less information about each)

CorporateInformation

http://www.corporateinformation.com

This site, from Wright Investor's Service, provides tens of thousands of company research reports, profiles, and analyses for over 20,000 companies around the world, including data from SEC filings for U.S. public companies. Perhaps the most useful and unique part is the links to company directories and other resources arranged by country, for over 50 countries. Use the pull-down window on the main page and choose the country. Full use of the site requires registration, but registration is free.

Hoovers

http://www.hoovers.com

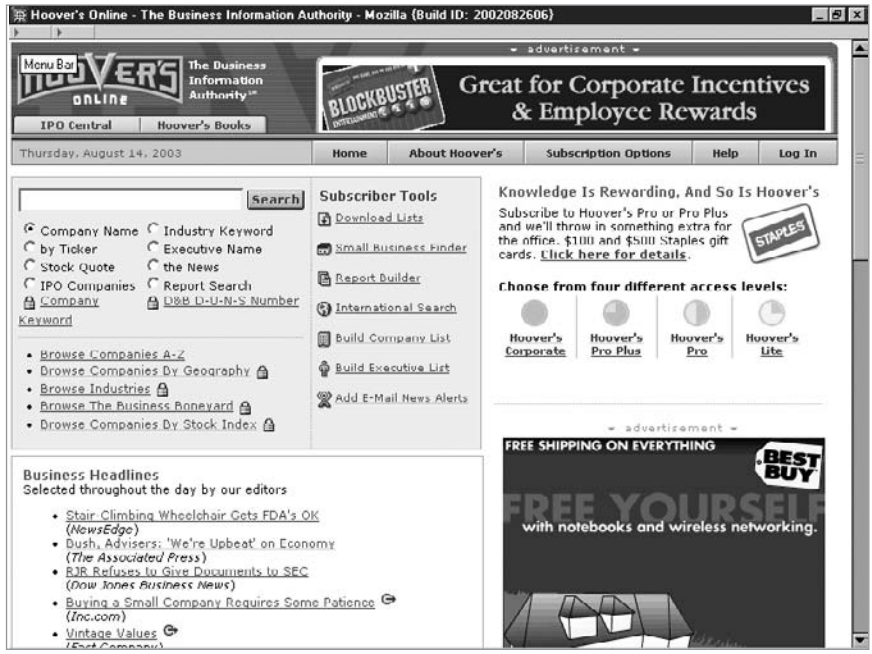
Hoovers provides information on 12 million companies. The site includes company profiles for over 18,000 companies, plus news, lists, IPO Latest Pricings and Filings, and other data. Much of the information is provided free, but a number of Hoovers' features are available by subscription only. The free portion is searchable by company name, ticker symbol, keyword, and executive name, and includes both U.S. and non-U.S. companies. Spend some time exploring this site. There are many unexpected gems, such as executive biographies, links to political contributions, pollution reports, and much more.

D&B Small Business Solutions

http://sbs.dnb.com

D&B Small Business Solutions, from Dun & Bradstreet, furnishes addresses, phone numbers, some trade names, industry, type of ownership, and revenue (often estimated) for over 13 million companies. Links to companies' home pages are available. The Search Options page allows a search by company name, city, state, ZIP, telephone number, or DUNS number. Other services and data are available for a fee.

Figure 6.6



Hoovers

Thomas Register

http://www2.thomasregister.com

This online version of the well-known print version of Thomas Register allows you to search by company name, product/service, or brand name. It covers 170,000 U.S. and Canadian companies (manufacturers only) with products listed under 72,000 headings. You will also find links to 7,800 product catalogs, plus company Web sites. You need to register to use the site fully, but registration is free.

Company Phone Numbers and Addresses

For companies, don't forget that the company's home page will usually provide phone numbers. Also, check the phone directories listed earlier in this chapter.

ASSOCIATIONS

If you know the name of the association and need further information, usually the best place to start is with the association's home page. From the other

direction, if you need to find the names of associations that relate to a particular topic, there are a couple places to consider as starting points:

1. Use a search engine and search for the subject and terms such as association society, organization.

   Example (in Google): "solar energy" association OR society OR organization

   or,

   solar energy OR power association OR society OR organization

   or, just

   solar association OR society OR organization

2. For U.S. associations, take advantage of the directory provided by the American Society of Association Executives.

American Society of Association Executives Gateway to Associations
http://info.asaenet.org/gateway/OnlineAssocSlist.html

This ASAE Gateway provides links to over 6,500 association sites. You can search by term, category, city, or state.

Professional Directories

To locate directories for a particular profession, try a search on the name of the profession and the word directory. It works sometimes; sometimes it doesn't. Two of the most widely useful directories, for physicians and lawyers, are listed here.

AMA Physician Select—Online Doctor Finder
http://www.ama-assn.org/aps/amahg.htm

This AMA site includes "information on virtually every licensed physician in the United States and its possessions, including more than 690,000 doctors of medicine (MD) and doctors of osteopathy or osteopathic medicine (DO). All physician credential data have been verified for accuracy and authenticated. …" It also contains a reference library of information on specific conditions.

Lawyers.com
http://lawyers.com

Lawyers.com allows a search of law firms or attorneys in 164 countries by practice area, name, and location. For more search power, click on More Search

Options. Lawyers.com now uses the Martindale-Hubbell database that will be familiar to any legal researcher.

LITERATURE DATABASES

As great as the resources on the Internet are, they still cover only a tiny portion of what we think of as the world's literature. In addition to only a tiny part of 1 percent of the world's books having their full text available through the Web, the vast majority of journal articles (especially those more than a few years old) are not available on the Web in full text. But, just as even a very large library owns only a small portion of extant literature, both a library and the Internet at least provide pointers to the broader corpus.

You will find numerous bibliographic databases on the Web that enable you to identify at least portions of what has been published on a particular topic, by a particular author, and so on. Many of these databases are available only through subscription, but many are available free, particularly on some large government-sponsored databases, such as for books, major national libraries' catalogs, and for journal literature, databases such as Medline, ERIC, and others.

To identify bibliographic databases on the Web for a particular subject, take a look at Gary Price's Direct Search site, or for a more specifically bibliographic list, try "A grab bag of (mainly) free bibliographies and bibliographic databases on the Web" from the Universiteitsbibliotheek at the University of Leiden. For single-site access to a broad range of journal literature, try Ingenta.

Direct Search

http://www.freepint.com/gary/direct.htm

Direct Search is Gary Price's collection of Invisible Web links and includes many bibliographic databases. Make use of the search box.

A grab bag of (mainly) free bibliographies and bibliographic databases on the Web

http://www.leidenuniv.nl/ub/biv/freebase.htm

From the Universiteitsbibliotheek at the University of Leiden, this site contains links to over 2,000 bibliographic databases and specific bibliographies.

Ingenta

http://www.ingenta.com

When you search the Ingenta site, you have access to 28,000 publications, mainly journals (from all fields). These include trade, scientific, and technical journals with coverage going back to 1988. In all, the site covers over 15 million articles. Ingenta is searchable by keyword, author, or journal title. When searching, keep in mind that you are searching titles and article summaries, not the full text, and therefore you may need to be a bit more general in your choice of terms.

COLLEGES AND UNIVERSITIES

Peterson's

http://petersons.com

The Peterson's site allows a search by the name of the institution, by keyword, location, major, tuition, size, GPA, type of college (e.g., four-year), and religion. You successively narrow your search by these criteria. It also provides resources for finding graduate programs, for test preparation, and for financial aid.

College Search

http://www.collegeboard.com/csearch

From The College Board, the College Search site provides a variety of resources relating to the Scholastic Aptitude Tests, finding a college, and financing an education. It contains information on 3,500 schools and presents a useful side-by-side comparison option.

TRAVEL

Travel is one area where you definitely need to know and use more than one Web site. Especially for travel reservations sites, don't count on any one always providing either the lowest cost flight or the itinerary that best suits your needs. On the other hand, loyalty to one site, and consequent heavier usage of that site, may get you special deals and discounts. Even if you don't book your own flights, it is useful to use these sites before calling your travel agent, because if you use these sites to select your flight first, you have more time to consider your itinerary than you may feel that you have when you are on the phone to the travel agent.

Fodors.com

http://www.fodors.com

Fodors, the print publisher, has a reputation for publishing what many travelers consider to be the best travel guides out there. Their Web site is an extremely rich resource for a tremendously useful collection of travel information, from what to see in a particular city to tipping practices worldwide.

Lonely Planet Online

http://www.lonelyplanet.com

The Lonely Planet site is a down-to-earth online guide to world travel from another well-known publisher of travel guides. If you are looking for something more off-beat, try the Themes Guide section.

Reservation Sites

Travelocity

http://travelocity.com

To find timetables, use a search engine and search for something along the lines of "timetable vienna prague rail"

As with most other travel reservation sites, Travelocity provides not just airfare, but rail fares, car rentals, hotel reservations, cruises, and more. It also provides travel guides and advice. In Travelocity, read the tips for identifying lowest fares.

Expedia

http://expedia.com

Expedia sometimes has lower prices than Travelocity (and vice-versa). Some users will prefer the way in which Expedia allows you to search for fares and itineraries and the way in which the results are presented.

Orbitz

http://orbitz.com

The newest of these three reservations sites, Orbitz provides differences in navigation and display of results. Compare the three to see which best suits your needs, but if you want the lowest price and best itinerary, check out all three. The Orbitz Travel Watch section provides a good selection of travel news, tips, forecasts, and so on.

FILM

Internet Movie Database

http://www.imdb.com

Whether you are looking for current show time, or a list of all of the movies in which Kevin McCarthy appeared, the Internet Movie Database is the place to go. It is not just a database of movies, but a movie portal with many resources, including movie and TV-related headlines.

REFERENCE RESOURCE GUIDES

The sites discussed in this chapter only scratch the surface in terms of what is available. For other reference-shelf type sites, consult the general reference directories (resource guides) discussed in Chapter 3. For a good printed reference tool covering the kinds of sites mentioned in this chapter, consult *The Web Library: Building a World Class Personal Library with Free Web Resources*, by Nicholas G. Tomaiuolo (CyberAge Books, Medford, NJ, 2004).

This page intentionally left blank

SIGHTS AND SOUNDS: FINDING IMAGES, AUDIO, AND VIDEO

"Amazing" is about the only word that adequately describes the collection of multimedia (images, audio, video) resources available on the Web. For images, not only are they there, but they are searchable—not as searchable as we would like, but still searchable. Whether you need a picture of the person you are about to meet, or photos of the streets of a specific town in a remote country or of an obscure microorganism, you have a moderately good chance of finding it on the Web. For audio and video, whether utilizing open sources for military intelligence purposes or for a discussion of Winston Churchill's "Finest Hour" speech in a history classroom, audio and video files can be tremendously useful. This chapter summarizes what is available, provides some basic background and terminology for understanding and using these resources, points to the tools for finding what you need, and presents some techniques for doing so most effectively.

THE COPYRIGHT ISSUE

Prior to using—or discussing here—any of the resources themselves, the overarching issue of copyright must be considered. Although the issue and its implications are already known to most people using the Internet for research, teaching, and other professional applications, the importance of the issue should be highlighted. The good news is that hundreds of millions of images, audio, and video files can be found easily on the Web. The bad news is that you may not be able to use those images as you might like. Whenever using images (and any other original works) in any way, remember first of all that the vast majority of images on the Web belong to someone. They are copyrighted. There still exists among some people (even some who should know better) the attitude that "I found it on the Internet, I can use it any way I want." As most readers of this book know: Not so! This does not mean that you cannot use these types of files in various ways. It does mean that the ways in which you use them must fall within "Fair Use" and other provisions of copyright law.

If you have found an image of interest and wish to use it in a report, on your own Web page, or for other purposes, be very aware that in most cases you cannot legitimately do so without getting the permission of the copyright owner. First, look on the site in which you found the image. You may be lucky and find a copyright statement that specifies when, where, and how you may use images from that site. (For a good example of such a statement, look at the NASA statement at http://nix.nasa.gov/copyright.html, but don't expect most sites to have such a clear statement with such minimal conditions.) For people in companies, universities, school systems, and many other organizations, you may find that your organization has published copyright guidelines for your use. For the layperson, trying to understand and interpret the actual laws will probably be more of a challenge than your time allows. For a very basic understanding of copyright issues, look at the article on copyright in the Legal Encyclopedia section of the "Nolo: Law for All" site at http://www.nolo.com.

IMAGES

A Tiny Bit of Technical Background

To view images on your screen, no technical knowledge at all is needed. If, however, you plan to save images and use them ("remember copyright") on a Web page, or print the image you save, a few points are in order.

Digital Image File Types

Web browsers can typically display only three image file formats: Joint Photographic Experts Group format (jpeg or jpg file extensions), Graphics Interchange Format ("gif" file extension), or Portable Network Graphics (png) format. The latter is relatively rare at present. Some search engines will allow you to narrow down your image search by these file types, but you are unlikely to need to do so.

Image Size

You will usually see image size referred to in "pixels." Pixels ("picture elements") are the space-related elements that make up a digital image. You can think of them as the "atomic" level of an image —- the smallest unit of a digital image. The ordinary Internet user can think of a typical monitor (with typical settings) as displaying "approximately" 72 or 95 pixels per inch (ppi). Depending on a number of factors, you can expect an image that has

dimensions of 140 pixels by 140 pixels to take up around 2 inches by 2 inches on a typical screen.

Capturing Images

To save an image file that you have found to your disk:

1. Hold your cursor over the image you wish to capture.
2. Click the right mouse button.
3. From the menu that pops up, choose "Save Image As" (in Netscape) or "Save Picture As" (in Internet Explorer).
4. Select the folder in which you wish to place it and rename the file if you wish. (Do not assign or change a file extension. It is important that the original file extension (either .gif, .jpg, jpeg, or png) be retained.

Editing Images

A discussion of image editing is beyond the scope of this book. However, since the object of an image search is often to have at print copy of the image, searchers are often confronted with the need to do a least some very minor editing of what they find. Operations such as cropping (trimming) and resizing, are fairly common and fairly easy to do. Anyone who has purchased a scanner or digital camera probably received software that provides these functions. There are numerous image-editing programs often packaged with scanners and digital cameras, the most common probably being Adobe's PhotoDeluxe or PhotoShop Elements. Almost any photo-editing software that you have will provide the basics. Windows operating systems also often include an image-editing program such as Imaging for Windows, or Paint, but some of these, surprisingly or not, may not offer some of the basic operations such as cropping. For the Internet user who wants to get into a bit more high-powered image editing, two of the most obvious choices are PaintShop Pro and PhotoShop. (PhotoShop Elements, one of the programs often packaged with scanners, goes a surprisingly long way toward the fuller capabilities of PhotoShop.) The main problem with photo editing is that it quickly becomes additive. When you have decided on one to use, do a quick search in one of the search engines for the program "AND" the term "tutorial." There are dozens of good photo-editing tutorials out there.

Types of Image Collections on the Web

Many, many image collections are available on the Web. Some are collections of images found on various pages throughout the Web, such as the image collections found on Google and AllTheWeb. Some are specialized by topic and represent the collection of a specific organizations, such as the Australian National Botanic Gardens' National Plant Photographic Index (at http://www.anbg.gov.au/anbg/photo-collection). Some are specialized by topic and represent the holdings of multiple institutions or sites, such as The Digital Scriptorium of . . . Medieval and Renaissance Manuscripts (at http://sunsite.berkeley.edu/Scriptorium). Some collections are by format or application, such as the numerous clipart collections. Another category, especially important for those who need good images that they can safely (legally) re-use in publications or elsewhere are the commercial collections, such as Corbis (corbis.com).

When searching for images, start by limiting your query to one or two words. Most images only have a very few words of indexing associated with them—if you search for "Boeing 747" then you will get substantially fewer good pictures of the plane than if you searched for just "747."

Searchability of Images

Though there are now hundreds of millions of images that can be searched on the Web, the search capabilities are fairly limited and rather "approximate." This is primarily due to the fact that the amount and quality of indexing that can be done at present by search programs is quite limited. Technologies are currently in development that will be able to see a picture of a tree, and without any text already attached to the image, be able to tell that the tree is a tree, maybe even to identify it as a spruce and maybe as a blue spruce. Implementation of this on a large scale for Web applications may take a while. In the meantime, except for relatively small collections, Web search engines do not have much to work with to identify and index what a picture is about. In most cases, the most that can used for indexing is the name of the image file (e.g., sprucetree.jpg), the ALT tag that may be included, a caption if the image is in a table, and text that is near the picture. Indexing the latter becomes somewhat of a gamble and can account for many of the false hits that may occur in image search results. That said, with a little imagination and a little patience and tolerance, the searcher can usually find a useful image quite quickly and easily using the collections and search techniques currently available.

Directories of Image Resources on the Internet

As with almost any other type of Internet content, there are numerous specialized directories (resource guides) that provide easy identification of image collections. The three that follow are well-known and useful examples of these directories and lead you not directly to individual images, but to sites that contain collections of images. For all three of these sites, the directory is on a single long page, so if you want to quickly find a specific topic on the page, you may want to take advantage of your browser's "Find in this page" option under the Edit menu.

Finding Images Online: Directory of Web Image Sites
http://www.berinsteinresearch.com/fiolinks.htm

This site is created by Paula Berinstein, author of *Finding Images Online* (CyberAge Books, Pemberton Press, Wilton, CT. 1996). It contains over 1,000 links to collections of images, arranged alphabetically by category.

Digital Librarian: A Librarian's Choice of the Best of the Web: Images
http://www.digital-librarian.com/images.html

Here you will find over 800 very well-annotated links to image collections. For maps, also see the companion "maps" collection at www.digital-librarian.com/maps.html. Be aware that the search box on the page is not a search of the page, but a search of Amazon.

BUBL LINK: Image Collections
http://bubl.ac.uk/link/types/images.htm

BUBL LINK's Images page has links to around 140 collections of images, with good, and often very extensive descriptions of each site. In addition to the obvious usefulness of such annotations, this means that, by using your browser's "Find in this page" option, you can do a very effective search of the page by topic.

Search Engine Image Collections

Image collections available through the major general search engines (Google, AltaVista, and AllTheWeb) are the largest collections of images on the Web. Getting their images from the billions of Web pages

they cover in their "Web" databases, they provide not just access to hundreds of millions of images, but also provide easy searchability (given the current limitations on image searching discussed above). Google's Images Search provides access to around 400 million images. AltaVista has almost as many. As for AllTheWeb, because of the combination of database size, indexing, and retrieval techniques used, you can expect retrieval of approximately half the number of images as Google and AltaVista. (These things change rapidly, so this proportion may change by the time you are reading this.) Keep in mind that the numbers retrieved do not necessarily reflect the relevance of the images to your specific search. Searchability and display of image results will also differ significantly among these three sources.

Google's Images Search

http://google.com

Google has the Web's largest collection of images, with approximately 400 million of them. To get to it, either click on the "Images" tab on Google's main page, or go directly to images.google.com. Once in Google's Image Search, you can simply enter your terms in the search box or you can click on "Advanced Image Search" to go to the advanced version. In either case, you will probably want to limit your search to only from one to three terms, since the number of words under which an image is indexed is rather limited.

Image Searchability in Google—Main Image Search Page

On Google's main image search page all terms are automatically ANDed. If you enter "temple esna" (without using the quotation marks) you will get only those images indexed under both terms. Quotation marks can be used for phrases and a minus sign in front of a term can be used to eliminate items indexed under that term. You can also use the OR as with a regular Google Web search. To retrieve all images indexed under the term "temple" and also under either esna or chnum, search for:

temple esna OR chnum

You can also use any of the prefixes that can be used in Google's Web search. In the case of images, the "site:" prefix is the most likely useful, in order to limit retrieval to a particular Web site. This can be used in combination with other operations such as the OR. For example, to get

images of either a "corn" or "maize" kernel from the U.S. Department of Agriculture site, search for:

corn OR maize kernel site:usda.gov

Searchability in Google—Advanced Image Search Page

Using the Advanced Image Search page (see Figure 7.1), you can:

- Use the "Find results" boxes to do simple Boolean by use of the "all the words," "any of the words," "Not related to the words" boxes.
- Specify a phrase search by using the "related to the exact phrase." (Using quotation marks around the phrase in any of the boxes works just as well.)
- Use the Size box to specify images of the following sizes: "any size", "icon sized", "small", "medium", "large", "very large", "wallpaper sized."
- Specify either jpg or gif formats, using the "Filetypes" box (default is "any filetype."
- Specify "any colors", "black and white", "grayscale" or "full color" images.
- Retrieve things only from a specific domain (such as gov or fda.gov).
- Use the SafeSearch option to set adult content filtering at "No filtering", "Use moderate filtering" (the default), or "Use strict filtering." (This is available only in the English version of Google.)

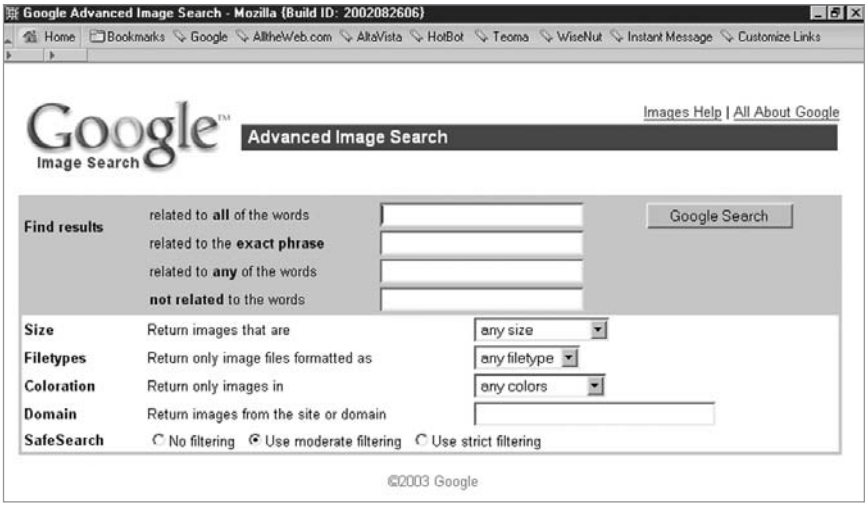


Figure 7.1

Google's Advanced Image Search Page

Google Image Results Pages

As the result of a search, Google will return a page containing thumbnail images for the first 20 images retrieved (with links at the bottom of the page to get to additional results). Included with each thumbnail is the file name, the dimensions in pixels, the size of the file (e.g., 16k), and the URL for the page on which it was found. As with Web results, image results are "clustered," and only the first (ranked by "relevance") image from a particular site will be displayed. If there are more matching images from that site, a "More results from …" link will be shown.

When you click on the image on a results page, it will take you to a split screen, with a larger (usually) image at the top and the original page at the bottom. The image at the top will usually be larger than the thumbnail, but scaled down from the original size, in which case there will be a link to "See full-size image."

AltaVista

http://altavista.com

AltaVista, in the late 1990s, had, by far, the largest image collection on the Web. It then fell substantially behind Google and AllTheWeb, but has more recently climbed back up to where you will often get almost as many, and sometimes more, results from an image search here than in Google.

Image Searchability in AltaVista's Images Search Page

Unlike Google and AllTheWeb, AltaVista's image search usually ORs all of the terms you enter, rather than ANDing them. This will vary, however, mainly because of AltaVista's algorithm that automatically identifies some common phrases and treats them differently. (AltaVista frequently changes its default operation, so you may want to keep an eye out for a change in this.) You can use quotation marks to specify an exact phrase and you can use a minus in front of a term to exclude images indexed under that term (this, however, does not work consistently).

The Image Search page (see Figure 7.2) offered by AltaVista allows you to limit your retrieval to:

- Photos, Graphics, or Buttons/Banners
- Color or Black and White
- The Web or AltaVista "Partner Sites" (e.g., RollingStone.com)

Figure 7.2

AltaVista's Image Search Page

AltaVista Images Results Pages

AltaVista image search results pages include thumbnails for the first 15 images with links at the bottom of the page to get to additional results pages. For each image, the file name, dimensions in pixels, and file size in kilobytes are displayed. Clicking on the image will take you to the Web page on which the image was found. Clicking on the "More info" link will display a page containing the preceding information, plus file type (e.g., jpeg), whether it is in color or black and white, a link to the original page, an "All media from this page" link that will show other images found on the original page, and a list of other pages that contain the same image.

AllTheWeb

http://alltheweb.com

AllTheWeb's "Pictures Search" provides one of the three largest image collections on the Web, but usually retrieves significantly fewer images than either Google or AltaVista. You get to it by clicking on the "Pictures" tab on AllTheWeb's main page.

Image Searchability in AllTheWeb—Main Picture Search Page

In AllTheWeb's Pictures Search, all terms you enter are automatically ANDed. You can use quotation marks for a phrase and a minus in front of a term to exclude items indexed under that term, but you cannot use the other search features that you can use in AllTheWeb's Web search (prefixes and parentheses for an OR). To get images that are indexed under the term "glacier" and also under "Alaska," but not those that mention the word "bay," you would enter the following:

glacier alaska -bay

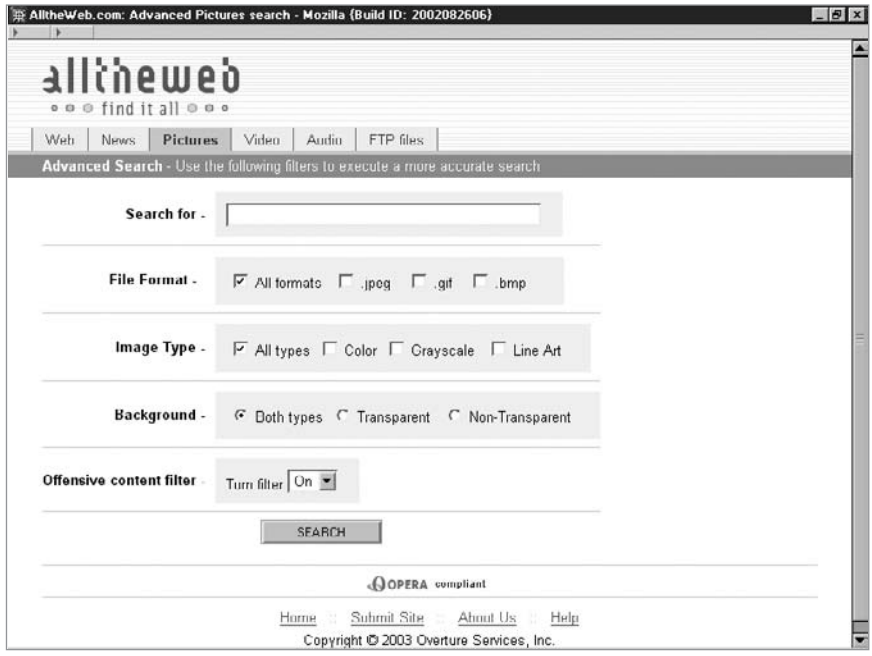
Image Searchability in AllTheWeb—Advanced Picture Search Page

From the "Pictures" search page, you also have the option of an advanced search page (see Figure 7.3).

This page does not allow any more Boolean or field-searching capabilities than does the main Picture Search page, but does allow you to limit your retrieval to the following:

- File format (all, jpeg, gif, or bmp)
- Image Type (all, color, grayscale, line art)
- Background (transparent, nontransparent, both)

Figure 7.3



AllTheWeb's Advanced Pictures Search Page

- Offensive content reduction either on or off (default is "on")
- Number of thumbnails shown on results pages (9, 12, 15, 18, 21, 24)

AllTheWeb's Images Results Pages

AllTheWeb will (by default) display nine thumbnail images on each picture search results page with links at the bottom to get to additional results. With each thumbnail, you will see the title of the image (often truncated), the type of file (jpeg, gif), image dimensions in pixels, and file size in kilobytes. There is also a magnifying glass icon which, when clicked, will display the full-size image. Clicking on either the thumbnail itself or on the title will take you to an "information page" for that image. There you will be shown, additionally, an enlargement of the image, the text (from the original page) from which your search term was drawn, a link to the page on which it was found and a link that displays the image by itself, in full size. When on the information page, there are also convenient "previous image" and "next image" links that allow easy browsing through your retrieved items.

Other Searchable Collections

There are a number of other searchable collections that contain images found on Web pages. The three general Web search engines above contain, by far, the largest collections, but you may want to examine the three directories of image resources listed earlier to identify searchable collections in particular subject areas.

Commercial Collections

If you are looking for a high-quality image that you wish to use in a publication, and not have to worry about possible violation of copyright or having to track down an owner to ask permission, consider going to a commercial collection of images where you can straightforwardly buy the right to use an image. Though not the only such collection, the best known collection is from Corbis.

Corbis

http://corbis.com

Drawing upon a range of collections such as the Bettmann Collection, the Hermitage Museum, UPI, and 3,000 other collections, Corbis collects and

makes available for sale a variety of photography, fine art, illustrations, etc. In addition to its own site, Corbis makes images available for purchase through other sites such as Yahoo!'s Picture Gallery (at gallery.yahoo.com).

Exemplary Individual Collections

As will be seen by browsing through the directories of image resources discussed earlier, there are hundreds or perhaps thousands of sites that contain useful collections of images. The following are just two examples of specific collections that exemplify the possibilities.

American Memory Project

http://memory.loc.gov

From the U.S. Library of Congress, this collection contains over 7 million digital items from over 100 historical collections at the Library of Congress. It contains Maps, Motion Pictures, Photos & Prints, Sound Recordings, Written Materials (Books & Other Printed Texts, Manuscripts, Sheet Music). Note that, even though this is a government site, most of the material contained on this site is protected by copyright. Use the Collection Finder section to browse by collection or topic, and the Search page to search across collections.

WebMuseum, Paris

http://ibiblio.org/wm (more specifically, www.ibiblio.org/wm/paint)

This truly impressive collection of artwork is a collaborative project headed by Nicolas Pioch. It is searchable by artist (around 200 of them) and by theme/period (from Gothic to the 20th century, plus Japanese art of all periods).

ClipArt

Still in the category of images, but addressing a somewhat different function and requiring different sources, is the area of clipart. In the Web context, this refers usually to artwork that is available on the Web, usually but not always free, for use on Web sites or printed documents. There are numerous collections and directories for theses resources, two of which are listed below. Users should read the fine print carefully. Most of the artwork is free, but you may be required to give a specific acknowledgment of the source.

Barry's Clipart
http://barrysclipart.com
This collection is both searchable and browsable by topic. The tabs at the top of the page lead to other large clipart collections.

Yahoo Directory > Graphics > Clipart
http://dir.yahoo.com/Computers_and_Internet/Graphics/Clip_Art
This section of Yahoo!'s directory provides links to over 100 collections of clipart, arranged by category and also alphabetically.

AUDIO AND VIDEO

Although less frequently used by researchers than the image resources on the Internet, audio and video files have a variety of applications beyond just entertainment (though all work and no play makes the extreme searcher a dull person). Accessing these resources is much easier than it was only a very few years ago, since most computers come with the necessary players, or they at least make it easy to identify and download the necessary player. For most types of files that will be encountered, the same players can be used for both audio and video. One of the greatest advances that has made the use of these files easy was the advent of "streaming" audio and video players that allow you to begin hearing or seeing the file without having to wait until the file downloads completely, and consequently, to make use of files of almost any length. The current remaining drawback in using larger files is for those users who still do not have a broadband connection. For them, the slowness of loading may make many files, especially video, virtually inaccessible.

As with viewing images, hearing and viewing sound and video files is easy. Searching them is the challenging part, mainly because of lack of indexing. Most audio and video files are indexed only under a very few words. Software is under development that will allow, on a large scale, detailed indexing (and searching) of these kinds of files, but it is not there now. In the meantime, keep the scarcity of indexing in mind as you search for sound and video files on any particular topic.

Players

For most of the older sound and video file types you are likely to encounter, (wav, au, avi, midi, etc.) your computer probably came equipped with the software necessary to play them. Likewise for more recent file types, particularly the

currently dominant, highly compressed, but high-quality sound and video file format, "mpeg" (Moving Pictures Expert Group format, with mpeg, mpg, mp2, mp3 file extensions). If you encounter a file type not currently supported, there is a good chance that there will be a link on the page that leads you to an easy free download of the necessary file. Among the players that many users are likely to encounter are Windows Media Player (pre-installed with all recent Windows operating systems), RealPlayer (a free download for the basic, but very adequate, version and upgrades available for a fee), Winamp, Musicmatch, QuickTime (essential for Apple users, but also available in a Windows version) and Divx - "The Playa" (for DVD). There are a number of others.

Audio

Music, historic speeches, online radio stations, and other sound resources can be valuable for numerous purposes, but in terms of frequency of use, the most frequently accessed type of audio content on the Internet is music, and unfortunately much of that access is widely recognized as being illegal due to the violation of copyright. However, there is ample opportunity for legal access to music and also access to other types of useful audio content.

Especially since the unaware serious searcher (and their employer) could easily become the target of copyright infringement suits, the copyright issue should be foremost in the minds of those who download audio and video material from the Internet. The popularity of file sharing ("peer-to-peer" or "P2P") among computer users on the Internet became very popular very quickly with the advent of the Napster program. Though Napster's own life was short (1999–ca. 2001), the concept begat a number of other P2P programs such as Kazaa, Grokster, Morpheus, and Gnutella that allowed listeners to continue to avoid paying for music. The intent of this book is neither to sermonize nor editorialize, but the serious searcher must be aware of the copyright issue.

In the next several pages, you will find a directory of audio resources, information of using the audio searching capabilities of major search engines, and three sites that focus on some specific types of audio resources (radio, speeches, and movie sound clips.)

World Wide Web Virtual Library: Audio
http://archive.museophile.sbu.ac.uk/audio

The World Wide Web Virtual Library: Audio is a directory of audio resources that provides over 150 links to audio resources on the Internet, including general repositories, newsgroups, radio, software, and other sites.

Audio Searching Using Web Search Engines

AltaVista, AllTheWeb, and HotBot all provide audio search options, with databases composed primarily of sound files they have found on the Web pages included in their main Web database. Each, of course, does it differently, and as with Web page searching, you need to be familiar with using more than one engine when searching for audio files.

AltaVista

To get to AltaVista's audio search, click on the MP3/Audio tab on the main page. Usually, the terms you enter in the search box are automatically ORed, unless AltaVista finds your terms in their master list of phrases, in which case your query may automatically be treated as a phrase (even narrower than ANDing). You can force an AND (in order to narrow down results) by placing a plus sign in front of each term.

Example: +churchill +"finest hour"

On audio search results pages, AltaVista will tell you, for each hit, the title of the sound clip, the author (if identifiable), the URL of the page on which the clip was found, whether it is in mono or stereo, and the duration. You need to go to the page itself to hear the clip. To get more information about any of the results, click on the More Info or Open in New Window link.

AltaVista also has an advanced audio/MP3 search that allows you to specify file type (MP3, wav, Windows media, Real, or other) and duration (less than or greater than one minute). For either the main or advanced searches, there is a link to turn the Family Filter on or off. The default is "on."

AllTheWeb

AllTheWeb's Audio Search (accessible from that tab on the main page) is limited to MP3 files. As a result, you may not find much of the older archival material here. All terms entered in the search box are automatically ANDed, and there is no option for an OR operation or for phrase searching (but because

all terms are ANDed, just entering two or three desired terms will usually be adequate).

On Audio Search results pages, the display for each retrieved record shows the reliability of the source, the title of the track, the size (in KB), and the date of the file. The reliability is indicated by the number of stars shown by each record and is an indication of "the reliability of the connection to the computer on which the file is located." The retrieved files are ranked by this criterion. A file folder icon in the Title indicates that a directory containing your search term was found on a site. Clicking the file folder icon takes you to the directory in which the file was found, clicking on the file folder with an arrow in it takes you to a listing from the parent directory, and clicking on the icon with the musical note symbol downloads the file (and usually begins playing it).

AllTheWeb has no advanced Audio search.

Lycos

On Lycos' home page (http://lycos.com), you will find a Multimedia link that allows you to search for pictures, audio, and video. The audio search usually retrieves fewer than half the number of items retrieved by AltaVista and AllTheWeb, but you may want to give it a try.

Audio Resources: Radio Stations

With thousands of radio stations now providing audio archives of their programs and/or streaming audio of their current broadcasts, great possibilities are open to Internet users. Besides just the recreational possibilities, these radio resources provide not just another channel for news (see Chapter 8) but can provide answers along the lines of "Who said what, and when?", "Did so-and-so really say what she was quoted as having said?", and "What have people been saying about a particular topic?" Although recent interviews may not be available in transcribed form, the audio may be there, whether on a well-known source such as BBC or on a local radio station. These radio stations can also be of value to those who are learning a foreign language. To easily locate a particular station, the following site will be useful:

Radio-Locator (formerly The MIT List of Radio Stations on the Internet)
http://www.radio-locator.com

Radio-Locator provides links to over 10,000 radio station sites worldwide and includes 2,500 that have live, streaming audio (for continuous listening). From this site you can search for radio stations by country, by U.S. state or ZIP code, by Canadian province, by call letters, and by station format (classical, rock, etc.). The advanced search page provides searching by multiple criteria, but limits your results to only the U.S. or Canada.

Other Audio Resources

The History Channel: Speeches

http://www.historychannel.com/speeches

A search in the search box will deliver links to a variety of resources available on the History Channel site. To get to the audio, look on results pages for the Related Speeches section and click on the audio (speaker) icon. You can also click on the Speech Archives link to get to an alphabetic listing of available speeches.

The Movie Sounds Page

http://www.moviesounds.com

This is a source for sound clips from over 80 major movies, mostly post-1950. The Sound Tools page has a very good collection of links to audio editing tools.

Video

In terms of usefulness and applications, most of what can be said about audio resources on the Internet is also true for video resources. In most cases, the same players that can be used for audio are used for video. You may find that your computer is not equipped with Apple's QuickTime Movie Player (available also for PCs), and it will be worthwhile to download if you run across a file that requires it.

To look for video, try the following places:

- For news, try news services such as BBC, CNN, and MSNBC, plus local radio and TV station Web sites.
- Use the video search capabilities of AltaVista and AllTheWeb.
- Look around in subject-specific sites such as The History Channel and American Memory (discussed above under "Audio").
- Use the BUBL LINK video resource guide described next.

Directory of Video Resources on the Internet

BUBL LINK / 5:15 Catalogue of Internet Resources: Video
http://bubl.ac.uk/link/v/video.htm

The BUBL LINK page is actually a directory of directories, providing annotated descriptions and links to over a dozen sites, each of which, in turn, provides collections of links to video resources for a variety of subject areas.

Video Searching Using Web Search Engines

Two of the major Web search engines, AllTheWeb and AltaVista, have databases of video content.

AllTheWeb's Video Search

To get to AllTheWeb's video search, click the Videos tab on AllTheWeb's home page. All terms entered in the search box are automatically ANDed, and there is no OR capability. There is an Advanced Video Search page that allows you to specify the following attributes:

- Formats (All, AVI, AVI/DivX, MPEG, Real, QuickTime)
- Streams or Downloads (Both, Streams Only, Downloads Only)
- Offensive Content Reduction On or Off

AltaVista's Video Search

Click on the Video tab on AltaVista's home page to get to its video search. As with AltaVista's audio search, terms you enter in the search box are usually automatically ORed. However, if AltaVista finds your terms in its master list of phrases, your query may automatically be treated as a phrase (even narrower than ANDing). You can force an AND (in order to narrow down results) by placing a plus sign in front of each term.

AltaVista will list up to 15 items on results pages, for each of them giving the file name, the file type (mono or stereo), the URL of the page on which the file is found, and links for "More info." The latter gives more information as to file size, duration, number of channels, sample bits, and sample rate.

NEWS RESOURCES

Once more, the word "amazing" has to be used. To be able to read the headline stories from a newspaper 10,000 miles away, sometimes before the paper appears on local residents' doorsteps, is indeed amazing. This chapter covers the range of news resources available (news services and newswires, newspapers, news aggregation services, etc.) and how to most effectively find and use them. Very importantly, the chapter emphasizes the limitations with which the researcher is faced, particularly in regard to archival and exhaustivity (comprehensiveness) issues.

TYPES OF NEWS SITES ON THE INTERNET

Understanding news resources on the Internet is challenging not just because there is such a broad and rich expanse of news available there, but because almost every news site is designed differently from every other and tends to serve somewhat different functions and missions. In ancient times, it was fairly easy to group news resources into categories such as newspapers, magazines and journals, radio, and TV. Today, it is harder to definitively categorize the types of places to go on the Internet for news. Although many typologies of news sources are possible, using the following categories can prove helpful in sorting things out (while recognizing that there is considerable overlap and that many sites fit in more than one category):

- Major news networks and newswire sites. Sites that are original sources for news stories, but may also gather and provide stories from other sources
- Aggregation sites. Sites that serve primarily to gather news stories from multiple sources
- Newspaper and magazine sites. Sites that serve as the online version for a printed newspaper or magazine
- Radio and TV Web sites
- Multi-Source News Search Engines. Sites that provide extensive search capabilities for a broad range of news sources

- Specialized News Services. Sites focusing on news in a particular subject area
- Alerting Services. Sites that provide, on a regular basic, a personalized selection of current news stories

FINDING NEWS—A GENERAL STRATEGY

A good starting point when one thinks of utilizing news on the Internet is to ask the question, "What kind of news are you looking for?"

1. Are you interested in breaking news (today's headlines)?
2. Do you need older news stories?
3. Do you want to be automatically kept up-to-date on a topic?

For breaking news, you might start with virtually any of the categories listed earlier, depending upon the breadth of your interests, both with regard to subject and with regard to the local, national, or international perspective needed. If you want to browse headlines, consider bookmarking and personalizing a general portal (such as My Yahoo!) and perhaps using it as the start page for your browser. Headlines in categories of your choice will show up every time you open your browser (or click Home). Alternatively, you might choose a news network site (BBS, MSNBC, etc.) or your favorite newspaper as your start page.

For older news stories, the choice is much more limited. If you are interested in the last few weeks, one of the search engines may serve. For international or high-profile news going back a few years, BBC may be a good choice, because it provides searching of all stories covered on its site back to 1997. If your interest is more local, check to see if the local paper has searchable archives.

If you need to keep up-to-date on a particular topic, take advantage of one of the alerting services and have headlines relating to your interests delivered to you by e-mail.

Characteristics to Look for When Accessing News Resources

For a research project or question, particularly when it is important that you know what you have and have not covered in your research, it is imperative that you be aware of exactly the kinds of items and time frames particular news sites

include. You certainly do not need to know this for every search, but the following factors are among the major content variables encountered among news sources on the Internet:

- Time frame covered. Some sites cover only today, others go back weeks, months, or years.
- Portion of original actually included. Particularly for newspapers and magazines, there is great variation as to how much of the print version is available online.
- Sources covered. Some sources may draw only from a single newswire service, others may include thousands of sources.
- Currency. Although "old news" can be tremendously valuable, "news" often implies "new." Depending on the site, the stories may be only minutes old, whereas for other sites the delay in including stories may be considerably more.
- Searchability. Many sites only allow you to get to stories by browsing though a list or by category. Other sites allow searching by keyword, date, and other criteria. Look around on any news site for a search box.
- Availability of alerting services. Although it may not be emphasized, on many sites, if you dig around a bit, you may find that a free e-mail alerting service is available. Some sites specifically exist as alerting services.
- Personalization capabilities. Some sites may allow you to personalize the site, so that when you go to it, categories of headlines of your choice and your local news, weather, and sports are displayed.

News Resource Guides

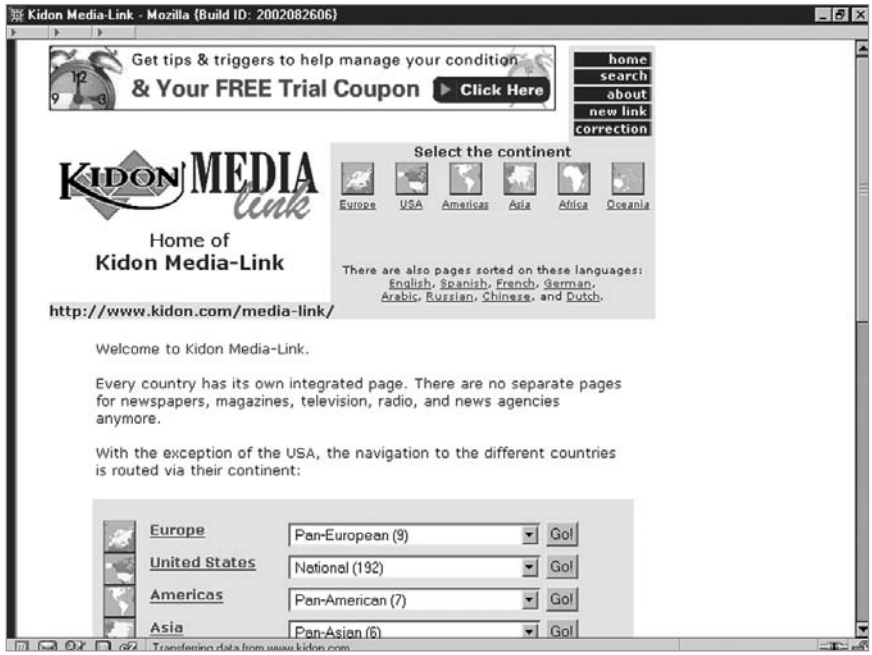
Many thousands of news sites are out there, and this chapter can only include a few selected sites. For knowledge of other sites, take advantage of one of the several good news resource guides. The three listed here are among the more highly regarded. Each provides somewhat different options in terms of coverage and searchability or browsability. One of the most important uses of the following sites is the easy identification of newspapers and other news resources for virtually any country and large city in the world. If you need to know the Web site for the local newspaper in Kathmandu, these resource guides will lead you there. You will find it worthwhile to go to one of these guides, choose a country, and spend a few minutes browsing through the sites for that country.

Kidon Media-Link

http://www.kidon.com/media-link

Kidon Media-Link is arranged to allow you to browse media sites by continent and by country, but also has a search page that enables you to search by a combination of media type (newspaper, radio station, etc.) and either by city or by words in the title of the site. It will also display sites by language (English, Spanish, French, German, Arabic, Russian, Chinese, and Dutch). Symbols indicate the presence of streaming audio and video for each site.

Figure 8.1



Kidon Media-Link

NewsLink

http://newslink.org

In addition to browsing newspapers worldwide by country, this site allows you to browse U.S. newspapers by the following categories: national papers, most-linked-to, state, type, major metros, dailies, nondailies, business, alternative, specialty, campus papers by state. You can also search by city and state, and specify All, Newspaper, TV, or Radio. It covers considerably fewer sites than does Kidon Media-Link and dead links are a problem.

Metagrid

http://www.metagrid.com

Metagrid covers not just newspapers but magazines, and for the magazines, it provides a nice browsable directory by subject. It covers fewer newspaper sites than does Kidon Media-Link.

Major News Networks and Newswires

Major news networks and newswires have sites that primarily provide news items that they themselves have produced, although they may utilize and incorporate other sources as well. Sites such as BBC, CNN, and MSNBC are the choice of many Internet users for breaking news, because the headlines are updated continually. They also typically provide a number of other items of information beyond news headlines, such as weather. These are sites for which the "click everywhere" principle emphatically applies. By spending some time clicking around on the page, clicking through the index links at the bottom of the main page, and browsing through the site index, you can get an idea of the true richness of these sites.

Newswire services such as Reuters, UPI, AP, and Agence France Presse are primarily in the business of providing stories to other news outlets. Their sites may contain current headlines, but may also be more a brochure for the service.

BBC

http://news.bbc.co.uk

A large portion of searchers throughout the world consider this the best news site on the Internet. It is particularly noted for its international coverage (BBC "World Edition"). In the international section of some U.S. services, "international" seems to be defined as "news from abroad that is of particular interest to the U.S." BBC's international coverage, though, is much more truly "international." Among its other strengths are its easy browsabilty, its extensive search capability, and the availability of free searchable archives going back to November 1997. The BBC news site is only one small portion of what the overall BBC site offers. Browse through the "A–Z Index" to find things from the Arabic Language News to Zoos. On the news home page, look for the languages options, the Country Profiles, and the free e-mail service.

All content comes from BBC writers, though they may utilize other sources such as Reuters in writing their stories.

The Advanced Search page allows searching by using multiple keywords, news section, and date, and have your results sorted by date or relevance. In the "Search for" box, you can use quotation marks for phrases and an asterisk to truncate (e.g., portug*, for portugal, portuguese). Terms you enter are automatically ANDed. To get to the Advanced Search Page, you must use the search box on the main page, then click on "Advanced search" on the results page.

Figure 8.2

BBC News Advanced Search Page

CNN

http://www.cnn.com

CNN.com, an AOL Time Warner company, has been displaying an increasingly international perspective, partly in connection with CNN's strong presence on European TV. It has European, Asian, and international versions,

and has interfaces in six languages (Spanish, Portuguese, Italian, Korean, Arabic, and Japanese). The Preferences page allows you to set the edition, personalize your weather, and receive e-mail alerts. Transcripts are available for most of its TV news shows for the last week and selected transcripts are available back to 2000. CNN.com also offers daily e-mail alerts on breaking news and weekly e-mail alerts on selected topics. For business news from CNN go directly to CNNMoney (*money.cnn.com*).

MSNBC

http://www.msnbc.com

The MSNBC site has an excellent menu for browsing by category and it also provides a search box, but no advance search option. In addition to MSNBC's own stories, you will find stories from local NBC stations, Associated Press, Newsweek, and other sources. Most stories are held online for a few weeks, some for many months. U.S. users can personalize this site by entering their ZIP Code, which will result in local news, weather, and sports headlines appearing at the bottom of the main page. There is also a free e-mail option. The MS in MSNBC means that this is one more opportunity to have Bill Gates influence your life.

Reuters

http://reuters.com

Reuters.com provides content that comes from over 2,000 Reuters journalists around the world. The site, which was significantly expanded in 2002, allows you to browse through general, financial, and investment news for the last day or so, and the search box allows retrieval of stories going back about two months. The site is searchable by keyword, company name, or stock symbol, and you can browse using eight main news categories. Do a search in the Quote search box and you are taken to the Company Search page, which provides not just stock quotes for the company, but also excellent company profiles, news, and other information on the company. Reuters also provides a free e-mail alert.

Newspapers

Thousands of sites for individual newspapers are available on the Internet. There may still be a few newspaper sites that contain an insignificant number

of actual stories, but most contain at least the major stories for the current day, and most contain an archive covering a few days, a few months, or even several years. Many online versions of newspapers do not contain sections such as the classified ads (or display ads) that appear in the print version. Some online versions contain things that are not in the print version, such as profiles of local companies.

Although most people are not likely to desert the print version of their favorite newspaper for a long time to come, the online versions do provide some obvious advantages, such as the searchability and archives. Some also provide greater currency, with updates during the day. Perhaps the most obvious advantage is simply availability—the fact that newspapers from around the world are available at your fingertips almost instantly. Take advantage of the availability of distant papers particularly when doing research on issues, industries, companies, and people. For industries, take advantage of specialization of newspapers dependent upon their location. For example, the San Jose Mercury is strong on technology because of its location in Silicon Valley, the Washington Post is strong on coverage of U.S. government, and Detroit papers are strong on the auto industry. For companies and for people, the local paper is likely to give more coverage than larger papers.

More and more newspaper archives are available online. In some cases, you can get recent stories for free, but have to pay for earlier stories. The price is usually quite reasonable, especially considering the cost to obtain them through alternative document-delivery channels.

Use the news resource guides mentioned earlier to find the names and sites for papers throughout the world. For availability of newspaper archives, check the site for the particular paper. Keep in mind that commercial services such as NewsLibrary, Factiva, LexisNexis and Dialog may have archives for newspapers that predate what is available on the newspaper's Web site.

RADIO AND TV

Sites for radio and TV stations are excellent sources for breaking news and may also contain audio (and sometimes video) archives of older programs. The next site mentioned, Radio-Locator, makes it easy to locate radio stations, but also take a look at Chapter 7 for further information on finding and using audio and video resources. The second site, NPR, is particularly valuable for archives of National Public Radio shows.

Radio-Locator (formerly The MIT List of Radio Stations on the Internet)
http://www.radio-locator.com

Radio-Locator's site provides links to over 10,000 radio station sites worldwide and allows you to search for radio stations by country, by U.S. state or ZIP Code, by Canadian province, by call letters, and by station format (classical, rock, etc.).

NPR
http://www.npr.org

This site provides easy access to National Public Radio stations throughout the U.S., but also provides a searchable audio archive of NPR stories and a facility for ordering transcripts.

Aggregation Sites

There are a number of sites whose main function is to gather news stories from a variety of newswires, newspapers, and other news outlets. Also, the three largest general search engines (Google, AllTheWeb, and AltaVista) provide extensive news searches of thousands of news sources. There are numerous other sites, for example, general portals such as Yahoo!, Lycos, and Excite, for which news aggregation is one function among many. Among the following six sites are three that are the most prominent sites focusing specifically on news aggregation. The other three are search engine sites (see Table 8.1 for a comparison of search features for the three search engine news sites.) These are all good places to go to make sure you are covering a wide range of sources, and each does it in a somewhat different way, with differing content and differing browsing and searching capabilities.

World News Network
http://www.wn.com

World News Network is an extremely impressive network of over 1,000 sites for individual countries, industries, religions, and so forth. The main page provides headlines and a list of categories for Regions and for Business, Countries, Entertainment, Environment, Politics, Science, Society, and Sport. The regional categories lead to the individual country news sites and the subject categories lead to news for a tremendous variety of subjects from nuclear waste to cocoa.

Table 8.1

Search Engine News Search Features

	AllTheWeb News	AltaVista News	Google News
No. of sources covered	3,000	3,000	4,000
"front page"	no	yes	yes
Archive	7 days	30 days	30 days
Advanced search page	yes	no	no
Limit by date	yes	yes	no
Limit by category	yes	yes	no
Limit by region	no	yes	no
Limit by domain	yes	yes	yes
Limit by language	yes	no	no
Truncation	no	yes	no
Results sortable by date	yes	yes	yes

The search options on the main page (see Figure 8.3) allow a search by a combination of keyword(s), language, and date and also allow you to specify how you want results sorted (source, language, word frequency, date). Consider taking advantage of the free e-mail alert services that allow you to choose from a list of geographic or topic choices. For this service, click on Site Map on the home page and look for WN by e-mail.

Moreover.com

http://moreover.com

Moreover.com primarily provides newsfeeds to organizations for their internal use or for use on their Web sites, but individuals can search the Moreover public database of over 2,700 publications by registering. Moreover provides the news for a large number of sites, including some major news sites such as AltaVista's News Search.

Newsnow.co.uk

http://newsnow.co.uk

Newsnow, like Moreover, is in the business of providing newsfeeds to other organizations and sites, and it was the first major site providing news aggregation dedicated to a U.K. audience. Like Moreover, anyone can search it, but unlike Moreover, Newsnow does not require registration. From its home page

World News Network

Figure 8.3

you can either search or browse by category. The categories are particularly useful due to the detailed breakdown provided.

Aggregation Sites—Major Web Search Engines

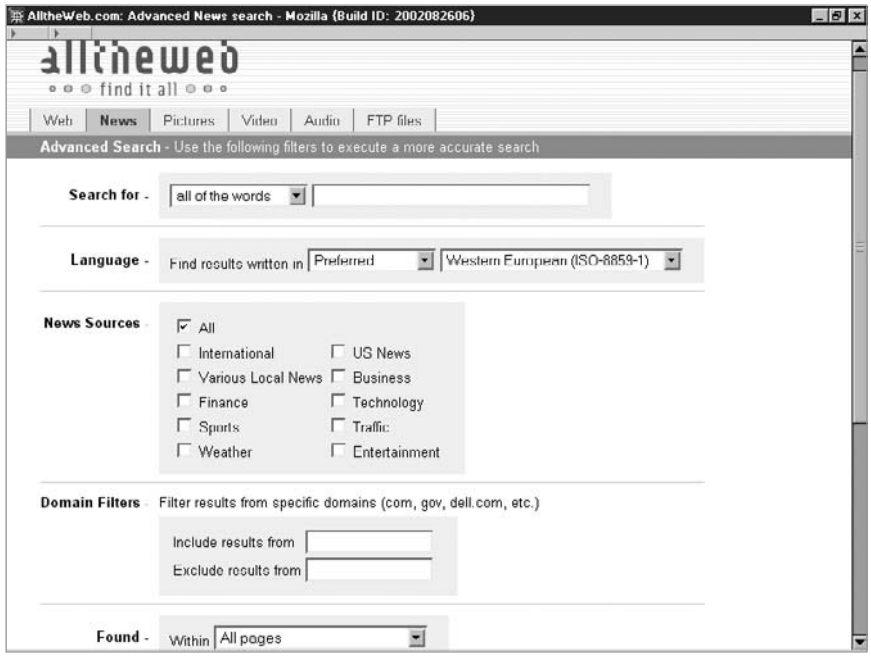
AllTheWeb News Search

http://alltheweb.com

To get to AllTheWeb's News Search, click the News tab on AllTheWeb's home page. Unlike the Google news page, AllTheWeb's news page is basically a search box and has no browsing capabilities (other than browsing the results of a search). It covers 3,000 top news sources, indexed on a near real-time basis, and records are retained for one week.

In the main page's search box, all terms are ANDed and you can OR terms by putting them in parentheses (just as with AllTheWeb's Web search). AllTheWeb does have an Advanced News Search page (see Figure 8.4) that allows specification of language, type of source (International, U.S. News,

Figure 8.4



AllTheWeb Advanced News Search Page

Various Local News, Business, Finance, Technology, Sports, Traffic, Weather, Entertainment), domain restriction, language (49 of them), Boolean (all the words, any of the words), and more. You can also choose to see 10, 25, 50, 75, or 100 results per page and limit your results to only those indexed in the last 2, 6, 12, or 24 hours; two days; or one week.

On results pages, there is an option that allows you to sort by relevance (the default) or by date.

AltaVista News Search

http://altavista.com

AltaVista's News search covers 3,000 publications, including sources from Moreover.com, other news sites, and stories found by AltaVista's own Web crawlers.

On its main news page, AltaVista provides a "front page" look with headlines of top stories and stories from four other categories. For the first two stories in each category, it shows the title (linked to the article itself), a two-line excerpt or description of the story, how long ago the story was found, a link to enable

translation of the story into any of eight languages, and a link to more information about the article.

Although it does not have an advanced search page, AltaVista has built extensive search functionality into its main news page (see Figure 8.5). In the main search box, terms you enter are automatically ANDed, but you can also use the Boolean OR, AND NOT, or NEAR. The NEAR (within ten words) is particularly powerful because it means you can allow for a few intervening words but still be sure that the words probably do have a meaningful relationship to each other. Also a minus can be used to NOT a term and you can use quotation marks to specify a phrase. Unique among the three search engine news sites, you can truncate a term (by using an asterisk). Prefixes can be used as in AltaVista's Web search, for example, url:nytimes to limit to New York Times stories. Pull-down windows are provided that allow you to limit results to a particular category (Top Stories, Business, Entertainment/Culture, Finance, Lifestyle/Travel, Science/Health, Sports, Technology), to a region of the world, to one of 13 of the major news sources, and to a date range (today/yesterday, last two weeks, last 7 days, last 30 days, or to a specific date range). For a searcher

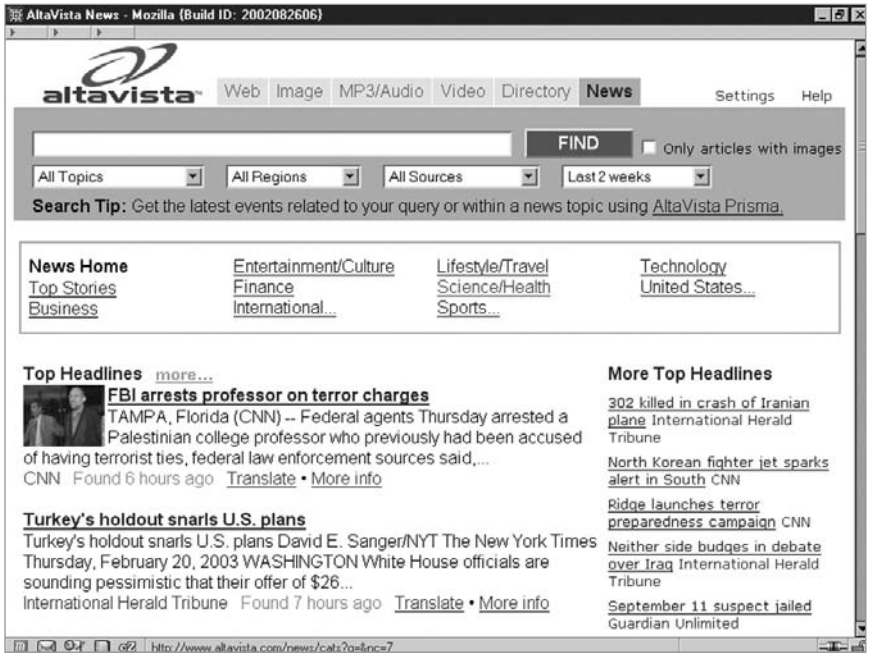


Figure 8.5

AltaVista News Search

who takes advantage of the Boolean, the NEAR, and truncation, AltaVista's news search provides the greatest control over precision and recall of all three search engine news sites.

Results are sorted by relevance, but there is a link allowing you to sort them by date. For stories that contain a picture, a thumbnail of the picture is shown next to the item.

Google News Search

http://news.google.com

Google's news search covers about 4,500 sources, with sites crawled continually, meaning that you may be able to find some things on Google only minutes after they appear in the original source. Items are retained in Google's news database for 30 days, and Google now provides a free alert service.

On Google's news page (http://news.google.com) you will find a browsable newspaper-type layout, with titles and brief excerpts for Top Stories and three records for each of the following sections: World, U.S., Business, Sci/Tech, Sports, Entertainment, Health (see Figure 8.6). Each news record contains the title, an indication of how long ago the story was indexed, a 30- to 40-word

Figure 8.6

Google News Search

excerpt, and links to related stories from other sources. If the story has a photo, a thumbnail appears beside the story summary. The small In the News section provides links to 10 hot topics.

On the left side of the page, links for each of the eight news categories will take you to a full page of 20 top stories for that category. Below that is a link that takes you to a text version of the page.

Importantly, of course, there is a search box. At the moment, Google has no advanced news search page, but in the main news search box you can use prefixes such as "intitle:" and "inurl:" (However, for the latter, only use the main part of the URL: "inurl:reuters" works well, but "inurl:reuters.com" misses most of the Reuters stories.) Search results look very similar to Web search results, but you will also find a Sort by Date link that conveniently arranges results by latest first.

Although news records are retained on Google for 30 days, for some sources the article may not be there when you click, especially for newspapers that have dynamic pages that change frequently, or that keep older articles in a separate archive database (mainly for fee-based access). Unlike regular Google, there is no cached copy of news pages.

Specialized News Services

Having a site for specialized news for a particular industry, area of technology, and so on, can be not just useful, but sometimes critical for those who need to make sure they are not missing important developments in that area. Such sites exist for a tremendous variety of subjects. In some cases, they are news-only sites, but in some cases specialized news is just one function of the site. For a good idea of the possibilities, go to WorldNews.com (discussed earlier) and click on Site Map. There alone, you will find over 200 specialized news sites. One very simple, yet effective approach to finding a specialized news site is to use a Web search engine and search for the industry or topic and the word "news."

Example: paper industry news

Weblogs

Fitting, somewhat, into the category of specialty news sites is the Weblog phenomenon. These sites began to appear in very large numbers around 2001. Weblogs (also known as "blogs," with the verb form "blogging") are, according to Dave Winer who runs the Weblogs.com site, "often-updated sites that

point to articles elsewhere on the Web, often with comments, and to on-site articles." These often focus on topics of very specialized interest and are a good way of keeping up-to-date on such specialized topics. One excellent example is Gary Price's "The Resource Shelf" (at http://resourceshelf.freepint.com), which covers news items of interest to reference librarians and other researchers. For a good list of Weblog sites, check out the Weblog category in Open Directory:

Open Directory: Computers: Internet: On the Web: Weblogs
http://dmoz.org/Computers/Internet/On_the_Web/Weblogs

ALERTING SERVICES

Among the most underused news offerings on the Internet are the numerous, valuable, and easy-to-use news alerting services. These are services that automatically provide you with a listing of news stories, usually delivered by e-mail and sometimes are very personalizable according to your interests. You don't have to go to the news, it comes to you. Although the concept has been around for decades, it has gone through many incarnations, ranging from mailings of 3 × 5 cards in the 1960s through the over-hyped "push" services in the mid-1990s to the more typical (free) e-mail mailings that have now stood the test of Internet time. If you are not familiar with this concept, the way it works is that you find a site that provides such a service, you register and, in most cases, pick your topic, and thereafter, you will receive e-mails regularly that list news items on that topic. Many newspapers provide alerting services, some allowing you to receive just selected categories of headlines. Some alerting services cover a number of sources and allow you to be very specific with regard to the topic. The best way to find out about these is simply to keep an eye out as you visit sites. Several sites already mentioned in this chapter provide alerting services. The following is one site that epitomizes the possibilities presented by this kind of service.

NewsAlert
http://www.newsalert.com

This is one of the most powerful free alerting services available on the Web and covers Businesswire, PR Newswire, Reuters, UPI, and over a dozen other sources, some of those sources themselves covering scores of sources. You can construct your profile using virtually as many terms as you like and

using Boolean and truncation features if you wish (see Figure 8.7). To set up e-mail alerts, first sign up, sign on, then go to News Manager.

Google News Alerts

http://www.google.com/newsalerts

Though still in Beta mode as this book goes to press, Google has begun providing a free alerting service for the 4,500 news sources it covers. You can enter your search and then specify the delivery frequency (daily, or "as it happens"). Multiple alerts can be established.

NewsAlert Topic Construction

Figure 8.7

This page intentionally left blank

FINDING PRODUCTS ONLINE

Whether for one's own or one's organization's actual purchase, or for competitive analysis purposes, many searchers frequently find themselves searching for and comparing products online. The Internet is a rich resource of product pages, company catalogs, product directories, evaluations, and comparisons. From the rather mundane purchase of a pair of slippers to the identification of vendors of programmable servo motion controllers, the Internet can make the job quicker and easier. This chapter takes a look at where to look and how to do it efficiently and effectively. As with other chapters, the intent is not to be exhaustive, but rather to provide the reader with a bit of orientation and some tips, point the reader in a useful direction, and provide examples of some leading sites.

CATEGORIES OF SHOPPING SITES ON THE INTERNET

A wide variety of types of "shopping" sites on the Internet serve a wide variety of functions. Most sites could be considered to fall into one (or more) of the following categories:

- Company catalogs
- Online shopping malls
- Price comparison sites
- Product evaluations
- Buying advice sites
- Consumer rights sites

Used in combination, these types of sites enable the user to find the desired product, check on the quality of both the product and the vendor, and feel confident and safe in making a purchase. The first site listed here, ShoppingSpot, is a good place to start if you want to explore, in an organized way, the variety of shopping resources available on the Web. Many of the sites covered in this chapter serve multiple functions. They are placed in the category that seems to best fit the site's primary function.

Online Shopping Resource Guide

ShoppingSpot

http://shoppingspot.com

ShoppingSpot will not only point you in a good direction as to where to shop, it also has a lot of links related to how to shop, with review sites, price comparison sites, consumer protection sites, coupon sites, and other resources. It has an excellent directory of specialized sites, from Antiques to Travel.

LOOKING FOR PRODUCTS—A GENERAL STRATEGY

The all-purpose rule of "keep it simple" works very well when looking for products online. If you know who you want to buy from, go directly to their site. If you have a specific brand, product, or set of characteristics, jump into a general Web search engine and get a quick (and perhaps a bit random) feel for what information is out there about the product. In the first 20 or so records, there is a good chance that you may get some links to vendors, some pages on specific models, links to some reviews, and, often, for popular items (for example, photo printers), links to sites about selecting that kind of product.

Then move on to a more systematic approach. For a business-related purchase, you might next go to Thomas Register of Manufacturers to identify vendors and specific products. For consumer products, you might go to one of the online shopping malls such as Yahoo! Shopping or eBay. Once you begin to focus on a likely choice, you can check out some reviews of the product itself at one of the review sites, do a search engine search on the specific model or products ANDing the word "review" to your search, use one of the merchant rating sites, and look around in newsgroups to see what individuals may have said about it.

COMPANY CATALOGS

If you know the name of the company you might want to buy from, and don't know their Web address, put the name in a search engine and you usually will be at their site in seconds. If you don't know who manufactures or sells the product, and it is more of a business or industrial product than a consumer product, go to Thomas Register (which also does include consumer products). There you will find a list of who produces what products, detailed categories of products, and links to the manufacturers' catalogs online.

ThomasRegister

http://www.thomasregister.com

The ThomasRegister site is the online equivalent of what library users and librarians recognize as that shelf full of thick green books that for decades has been the starting place in a library for identifying products and manufacturers. ThomasRegister contains millions of product listings, placed under 72,000 product headings, and over 170,000 U.S. and Canadian company listings. You must register to use it, but registration is free. Once registered, you can search by the product or the company (using the Boolean AND, OR, and NOT, if you wish), browse through very helpful and detailed product categories, narrow your list of manufacturers by state or province, get a brief profile of a company, send an RFQ (Request for Quote), and buy an item. ThomasRegister offers two other sites to consider. ThomasRegional (http://www.thomasregional.com) is a site covering more than 550,000 local industrial distributors, manufacturers and service companies. Thomas Global Register (http://www.tgrnet.com) includes 500,000 manufacturers and distributors from 26 countries.

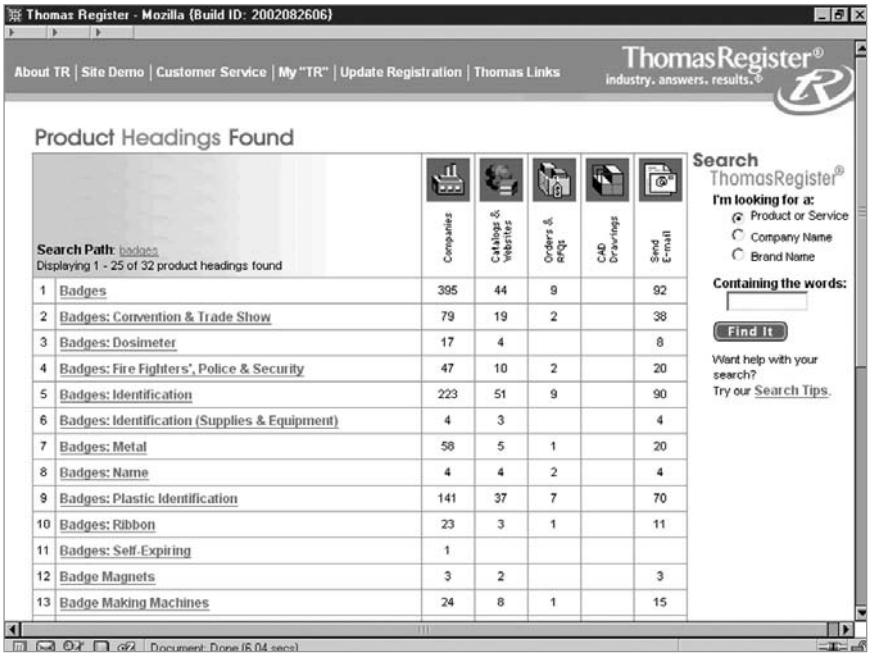


Figure 9.1

ThomasRegister Category Listing

SHOPPING MALLS

You don't have to look hard to find sites that enable you to purchase an item online from hundreds or thousands of online stores through a single site. eBay, Amazon, and Yahoo! Shopping are among the most widely used of these malls, but there are many, many more that serve the same function or may be specialized for a particular category of product (see ShoppingSpot, earlier). These offer the advantages of not just being able to locate the best products from a variety of suppliers, but to purchase the product online easily and securely. Although the interfaces are all different, you will notice a number of commonalties. Most have a directory that allows you to browse by category, most have a search function, and most use Shopping Cart technology, enabling you to gather multiple items and then check out for all items at once. Four representative and well-known sites are described here. For additional sites, see ShoppingSpot.com.

Yahoo! Shopping

http://shopping.yahoo.com

To get to Yahoo! Shopping, click the Shopping link on Yahoo!'s home page or go directly to http://shopping.yahoo.com. From Yahoo!'s main Shopping page, you can browse through 27 product categories or use the search box. When searching, all terms you enter are ANDed, but you can use the minus sign in front of a term to NOT a term. By selecting one of the 22 categories in the pull-down window to the right of the search box (see Figure 9.2), you can limit your retrieval to just that category of products.

Yahoo! provides a number of features that can make purchasing easier, including Yahoo! Wallet (where you can store your credit card, shipping, and billing information and make checkout easier), Shopping Account (displaying past and currently selected purchases), My Stores list (stores you have purchased from or that you want to add to the list), multiple Shopping Carts, and Research and Compare (a comparison feature for selected popular categories of items such as computers, digital cameras, and watches). You will also find a rating system for merchants, based on buyers' feedback. In addition to buying, individuals as well as stores can sell items through Yahoo! Shopping. Yahoo! Shopping makes both buying and selling easier, and the reliability of the Yahoo! brand provides a high level of confidence in the process.

Figure 9.2

Yahoo! Shopping Page

Amazon.com

http://amazon.com

Initially just an online bookstore, Amazon has expanded to a full shopping mall, where you can buy almost anything, from rare books to sweaters and software. The main page provides both a detailed directory for browsing and a search box (terms you enter are automatically ANDed, but you can use a minus to eliminate terms). Because of the richness of the site, both in terms of shopping breadth and shopping features, you will find it worthwhile to try the "click everywhere" approach to exploring the Amazon site. Among other things, you will find an advanced search page for many of the categories (click on the category, then look under the search box for a link to the advanced search page); personalized recommendations based on your previous purchases; sites for Canada, the United Kingdom, Germany, Japan, and France; shipment tracking; gift registries; selling options; and more. Amazon also throws in some unexpected extras, such as your local movie showtimes.

eBay

http://ebay.com

Although many people think of eBay as an auction site where almost anything but body parts are auctioned off, it is not just an auction, but also a shopping mall where you can buy things outright, avoiding either the fun or effort (however you see it) of having to go through the auction process. When you do a search or browse through the categories, you will see tabs that take you to All Items, Auctions, or Buy It Now. The latter is for items that can be purchased without the auction process. eBay has one of the most sophisticated sets of search features of any of the shopping sites. Look for the "Smart Search" or "Advanced Search" links on the main page and other pages. eBay's Advanced Search allows you to search by simple Boolean (all the terms, any of the terms), phrase, category, price range, location, seller, etc. Take advantage of the very good Help section to get a good feel for the possibilities and procedures.

Froogle

http://froogle.com

Froogle (clever name, eh?) was introduced by Google in 2002, a cousin of Google's under-recognized Google Catalogs (http://catalog.google.com) that includes the content of over 5,000 catalogs. However, Froogle goes beyond just listing the content of catalogs, and includes content that (1) is the result of Google's crawling of the Web to identify product sites, and (2) content submitted by merchants. On Froogle's home page, you will see a search box, a link to the Advanced Froogle Search page, and a directory that allows you to browse for products by category.

Ranking of results is not dependent upon payment for listings, but relies on the same ranking technology used at Google.com. Merchants cannot buy search results listings but can buy Sponsored Links that are placed elsewhere on the results page (see Figure 9.3). Unlike most other shopping sites, no purchases are made through Froogle directly. Actually, you will find that Froogle results may include items from other shopping sites such as eBay, Amazon, Barnes and Noble, and others. (Note that only one matching item per store is displayed, but you can click the "all products regardless of store" link to see others.)

Figure 9.3

Froogle Results Page

Froogle's Advanced Search page allows you to search by simple Boolean, price range, category, and also limit your search to product name or description. On results pages, you can also narrow your search by category and price range.

PRICE COMPARISON SITES

Basically any time you look at the same product from two different suppliers, you are doing a price comparison. In that sense, most of the sites discussed in this chapter are price comparison sites. Some sites though, put emphasis on the comparison aspect. These types of sites are discussed here, and the ones that put the emphasis on consumers' own reviews and opinions are grouped together as a separate subcategory. Likewise for merchant evaluation sites. This division is somewhat arbitrary and reflects more a matter of emphasis of the site than on a definitive distinction.

Directory of Price Comparison Sites

Open Directory: Consumer Information: Price Comparisons
http://dmoz.org/Home/Consumer_Information/Price_Comparisons
This section of Open Directory gives over 20 subcategories of price comparison sites (Appliances, Automobiles, etc.) and a listing of over two dozen price comparison sites that cover shopping in general.

MySimon
http://www.mysimon.com
Many online malls such as Yahoo! Shopping and Amazon, allow a price comparison, but you may see featured sites emphasized, or only sites from merchants who pay to be a part of that online mall. MySimon, one of the earliest online shopping sites, puts emphasis on comparison. It, like Froogle, crawls the Web to collect information from online stores. You can browse by category (look for the Browse pull-down window), or use the search box to search either the entire site or a selected category. (All terms you enter in the search box will be ANDed.)

PRODUCT AND MERCHANT EVALUATIONS

Some of the sites discussed here, such as Amazon, may build both product and merchant reviews into their results. Other sites on the Internet specialize in reviews and evaluations, including consumer opinion sites and merchant rating sites. Among these are Epinions, bizrate, Consumer Reports, Consumer Search, and Consumer Review.

In addition to using these sites, Web search engines can also be used effectively to find reviews and evaluations by simply doing a search on the name of the product (e.g., Olympus c700), or the type of product (digital cameras), in combination with the terms "evaluations" or "reviews."

Examples: (in Google) "digital cameras" reviews OR evaluations
(in AllTheWeb) "digital cameras" (reviews evaluations)

Going one step further, especially if you are tracking your own or competitors' products, take advantage of the frequent comments that appear in newsgroups regarding products. Look both at Google Groups (http://groups.google.com) and Yahoo! Groups (http://groups.yahoo.com) (see Chapter 2).

Epinions

http://epinions.com

On the surface, Epinions looks much like other shopping sites, with a search box and over 30 browsable categories that include over 2 million products or services. What differs is that the emphasis in Epinions is on the reviews. For each product, you will find links to further details about the product and to reviews written by Epinions users. To provide reliable reviews, even the reviewers can be reviewed by Epinions' "Web of Trust" system. For various products, you will also find advanced search options, buyers guides, and store ratings.

BizRate.com

http://bizrate.com

At BizRate, you can browse by category or you can search (either the entire site, limited to a particular category). Once you identify a particular product, you will typically have access to details about the product (often detailed specifications in the case of electronic and other technical products), reviews of the product, and the list of stores and their prices. For each store you will see a rating, based on feedback from BizRate users.

Consumer Reports

http://consumerreports.org

Consumer Reports, the publisher of the well-known product review journal, has its evaluations available online, but only to paid subscribers.

ConsumerReview.com

http://consumerreview.com

ConsumerReview.com, one of the specialized product review sites, specializes in reviews of outdoor, sporting goods, and consumer electronics products.

Consumer Search

http://consumersearch.com

Consumer Search takes a different approach to providing reviews by having its editors "scour the Internet and print publications for comparative reviews and other information sources relevant to the consumer." The reviews on the site are based on those sources and a set of criteria developed by Consumer Search.

BUYING SAFELY

Although many Internet users quickly began to take advantage of the benefits of online purchasing, many users are still quite shy about giving up their credit card numbers to a machine. Having a healthy skepticism is indeed a reasonable approach. Knowing where caution ends and paranoia begins is the problem. In general, following a few basic rules should keep the online purchaser fairly safe. There are few guarantees, but there are also few guarantees that the waiter to whom you gave your credit card in the restaurant did not do something illegal with it. If the following cautions are kept in mind, online purchasers should be able to feel reasonably secure:

1. Consider who the seller is. If it is a well-known company, there is some security in that. (Yes, I do remember Enron.) If you don't recognize the seller, do you know the site? Sites like Amazon and Barnes & Noble are respected and want to protect their reputation. If you are buying through an intermediary such as eBay, it likewise has a reputation to protect and builds in some protections, such as providing access to feedback about sellers from other customers. On some merchant sites, you will see symbols displayed indicating that the merchant is registered with organizations that are in the business of assuring that member merchants meet high standards. Two of the leading such organizations are BBBOnline (from the Better Business Bureau) and ePublicEye (http://epubliceye.com). On the BBBOnline site, you can search to see if a company is a member. On ePublicEye you can look up member companies to see their customer satisfaction rating, on-time delivery record, and other information. For various legitimate reasons, even large and reputable sites may not participate in programs such as these, so the lack of a seal of approval alone should certainly not keep you from buying.
2. When you get to the point of putting in payment information, check to see that the site is secure. Look for the closed padlock icon on the status bar at the bottom of your browser, or the https (instead of http) in the address bar of the browser.
3. As with traditional purchases, look at the fine print. Look for the payment methods, terms, and return policy. Also look around for seller contact points, such as phone number and address.
4. Print and keep a copy of the purchase confirmation message you receive when you complete the purchase.

5. Pay by credit card to be able to take advantage of the protections this provides regarding unauthorized billings. Some sites, such as eBay, will also provide services. These charge the seller a fee and may cause a slight delay, but hold the money until the product is received. Payment services such as PayPal also build in some safeguards.

For additional advice, take advantage of the http://safeshopping.org site created by the American Bar Association. If you encounter problems with a purchase, you may want to consult the Federal Trade Commission's site for E-Commerce at http://www.ftc.gov/bcp/menu-internet.htm. For cross-border complaints, consult eConsumer.gov.

This page intentionally left blank

BECOMING PART OF THE INTERNET: PUBLISHING

The Internet is, obviously, a two-way street. So far, this book has been discussing using the Internet to find information. The other direction is providing information to be found. Newsgroups and mailing lists, discussed in Chapter 5, are one way of contributing to the content on the Internet, but the more systematic way of providing information to others is to have your own Web site, not necessarily your own domain (e.g., yourname.com), but at least a page or two that you have produced and are responsible for. (For simplicity in this discussion, the term "Web site" will be used to refer to the page or pages you might build, whether they might be a part of another site or have a domain of their own.) The number of reasons why you should consider such a step is virtually (there is a pun there) unlimited. Indeed, building Web pages isn't just for Webmasters anymore. Anyone who has information they feel is worth sharing with others is a candidate. You may indeed find you want to put up a Web page for a course you are teaching, a conference paper you are presenting, your school, your family, or as an online resume. You may realize that Web pages are useful for lesson plans, for demonstrations, and presentations in a broad range of contexts. Also, you may have noticed that, throughout the book, you have run into pages that were produced by individuals not for monetary gain, but for their love of their subject.

Having created a page or site of your own is also useful for another reason. For those who are involved in contributing input to their organization's site, or to someone else's site, having done your own page or site can provide a healthy perspective. It can, on one hand, take away a lot of the mystique (you won't be unnecessarily awed by some of the cute little things you see), and on the other hand, you will have a better appreciation for the more sophisticated things you see. Also, if your time and inclinations permit such, building your own site can be a lot of fun.

This chapter does not intend to teach you how to do so, but intends to provide an overview of what is involved in order to help answer the questions, Can I do it (build my own Website)? What is involved in doing so? What will it cost?

WHAT'S NEEDED

The main things needed for building a Web site of your own are: a purpose, time, software, skills, and a place to publish. Depending upon what you want to produce, each of these can either be minimal or extensive.

Purpose

The introductory paragraphs to this chapter mentioned some of the reasons for creating your own Web site. Before you start, though, it is advisable to give a fair amount of consideration to why you are doing it and what you want to accomplish. Your aims may change continually, but the more direction you have to begin with, the less you may have to go back and change later. Write down your purpose. The main purpose of almost any page is "communication." What do you want to communicate and why?

Tied in closely to your statement of purpose will be an analysis of your intended audience. Who are you addressing? What background are they likely to have in connection with your topic? What age level are you addressing? How skilled are they likely to be in using and navigating through Web pages? What is their level of interest? For the latter point, if your page is the syllabus for a course you are teaching, users have a high level of interest in that they may be required to use the page. If you are selling something, you need to design a page that will do a good job of attracting and keeping the readers' attention.

Time

If you are using a free Web site service such as Tripod or GeoCities (discussed later), and you take advantage of their templates and already know what information you want to put on the site, you can have a Web site created and available for use in an hour or so. The time required to build and maintain a site goes up from there, depending upon how fancy you want to get, how much content you want to include, and how much maintenance the site will require (updating, etc.).

Software

If you are building a site using a free Web site service such as GeoCities or Tripod, you will not need any software other than your browser. These sites provide what you need to make a basic but at the same time very attractive

site, with room for lots of content and many pages. Beyond that, unless you decide to learn how to write HTML (HyperText Markup Language) code, you will need a Web page editing program (HTML editor) such as Dreamweaver, FrontPage, Homesite, Claris Home Page, or Netscape Composer. (There are many, many more.) These are basically word-processor-like programs that convert what you enter, and the features you choose, into HTML code.

The cost of these can range from free (Netscape Composer) to several hundred dollars. If you are using the editor for educational purposes, you may find an educator's rate for some programs that will be substantially less than the full price.

Netscape Composer, which comes as a part of Netscape Communicator or later versions of Netscape such as Netscape 6, provides the basics of what you needs to build a Web page. Parts of the program can be a bit clunky, it does not provide the more sophisticated features such as forms and cascading style sheets, and its uploading feature really doesn't work. It does, though, provide what most beginners need, and the fact that it is free is significant.

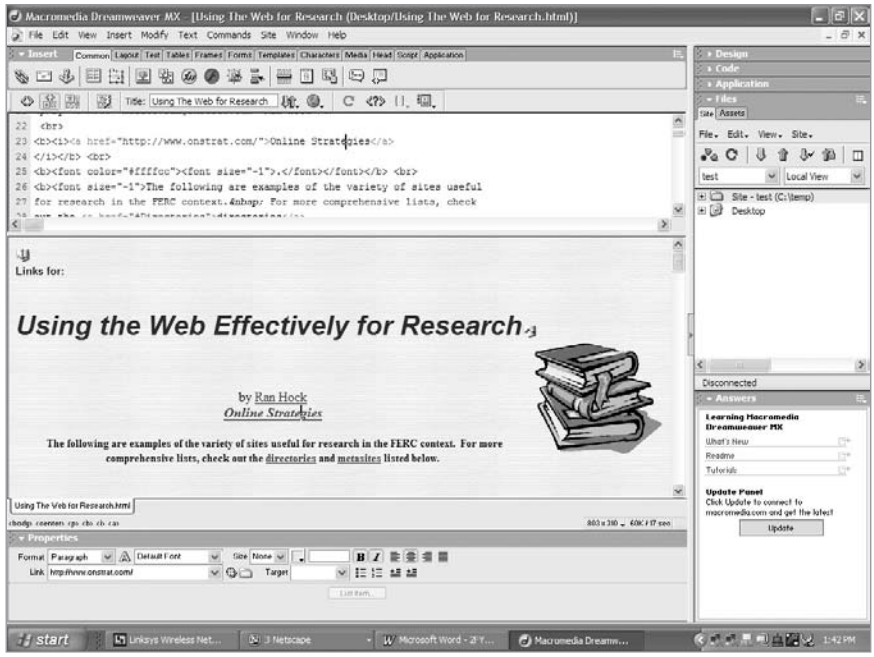
If you think you are going to want to get more sophisticated, have many pages on your site, and make it interactive, you may want to start with a sophisticated, but still easy-to-use program such as Dreamweaver (see Figure 10.1).

Uploading your finished pages to a Web server will require file transfer software. Most of the HTML editors build in this feature, but if you use Netscape Composer, you will want to use some standalone file transfer software such as WS_FTP. (Noncommercial users can download a free version of WS_FTP.) For Macs, a popular file transfer program is Fetch.

Graphics Software

It is likely that you will want some images on your site and unlikely that you will want to put them on your page without making some modifications, such as cropping and some other easy changes that will improve the image. Chances are that you already have graphics software that will do what you need. If you have purchased a scanner or a digital camera, it probably came with a program such as Adobe PhotoDeluxe or Adobe PhotoShop Elements, or any one of several other graphics programs. These programs are surprisingly robust and adequate for most operations that need to be performed on images to make them ready to be placed on a Web page. If you want to get fancier, consider a heavier-duty program such as PaintShop Pro or Adobe PhotoShop.

Figure 10.1



Dreamweaver

Skills

To build a Web site with the minimalist approach (using templates on a free Web site service) requires only the ability to follow step-by-step instructions. Beyond that, the ability to use (or learn how to use) an HTML editor will be needed and ability to work with graphics will be useful. Be aware that the use of graphics software can be addictive, and, as well as using it for your professional work, you may find yourself up at 3 A.M. fixing the cracks and tears in that photo of your great-grandfather and adding feathered edges, drop-shadows, and other special effects to your pictures.

If you are new to using HTML editors and graphics software, there are a number of ways to learn. Your choice of ways will probably depend upon your own learning styles. Most programs you purchase will have a built-in tutorial, and if you commit an hour or so you can be on your way. If you are willing to commit several hours, you will probably find yourself in quite good control

of the program. There are also tutorials available on the Web for most popular programs, and they sometimes provide a more simplified, yet effective, approach to Web page editing and graphics software. Do a Web search for the name of your program and the word "tutorial" and you will probably find several. There are also numerous books and classes available for the more popular programs.

The alternative to using an HTML editor is to learn to write HTML code. Most people would probably consider this the hard way, but it can actually be fun. (Then again, some people also consider jumping into an icy river on New Years Day "fun.") For most people, starting with a Web page editing program makes the most sense, but as you get into Web page building, you eventually may want to learn the basics of HTML because of the added control it can give you. (In the interest of full disclosure, the author admits to having had fun writing HTML code.)

Where to Publish

Among the main options for places where the individual Web site builder may place a Web page are the following: on a Web hosting service with your own domain name, on your organization's server, or on one of the "free Web site" sites.

Your Own Domain on a Web Hosting Service

For someone who owns a company and/or needs to make the most professional impression, having one's own domain name is the way to go. The easiest way to get started at this level is to choose a Web hosting (virtual hosting) company and place your site on their server. These companies can easily be located through their ads in computer magazines, a yellow pages directory, or a Web search. There are numerous directories specifically of Web hosting services. To locate these directories, use the following Open Directory category (at http://dmoz.org or use the Directory tab on Google):

Computers > Internet > Web Design and Development > Hosting > Directories

Web host services typically charge from $15 to $20 per month for basic service and will also lead you through the process of getting your own domain name, which requires a registration fee of around $70 for the first two years. One of the big advantages of these services is that they handle most of the paperwork of the domain name registration. Compare the ads, call their toll-free numbers, and talk to two or three of them, partly to get a feel for their degree of customer service orientation.

Putting Your Site on Your Organization's Server

If you are in an academic institution, there is a good chance that your institution may provide free Web space for you. For other organizations, there may be similar possibilities depending upon your purpose and the nature of the organization. Do not be surprised if you are presented with a list of criteria that must be met, with regard to both content and format. If you are a faculty member at a university, you may easily be assigned Web space with minimal restrictions and the permission to upload your pages when and as you like. At the K-12 level, there is a very good chance that there will be cooperation and enthusiasm for teachers or others to create school and classroom pages. In other situations, it may not be as easy, and there are situations where you will encounter institutional Webmasters with requirements that make little sense. Fortunately, a larger proportion of people in charge of organizational sites are realistic and helpful. If you are in a commercial environment, do not expect to have a page of your very own loaded on a company Web site.

Free Web Page Sites

For many people who want to get started, using a free Web site service is an excellent starting place. Even if you are planning to move up to placing your site on your organization's server or to having your own domain name on a hosting service, these free Web site services provide a good initiation. Free Web sites are available from a variety of sources. The ISP (Internet Service Provider) you use at home may provide a free site for subscribers. There are also commercial sites that specialize in providing free space. You pay for these by putting up with the ads that will come along when your page is displayed, but it is often a good bargain. They usually also offer upgrades (that avoid the ads) for a relatively small monthly fee. These are the leading free Web site services:

GeoCities (a part of Yahoo!)—http://geocities.com

Tripod—http://tripod.com

Angelfire—http://angelfire.com

Each of these provides 15–20 megabytes of storage, enough for a very substantial Web site. They also provide templates that can be used, HTML editors, and uploading capabilities, and they allow you to upload pages you have created elsewhere, such as in another HTML editor. These sites also make it easy to place features such as the following on the pages you create: photos,

a counter, news headlines, weather for places you choose, online messages, and guest books. In most cases, you will have at least a little control over the kinds of ads that appear by your choice of the interests or communities that you select as part of the sign-up procedure.

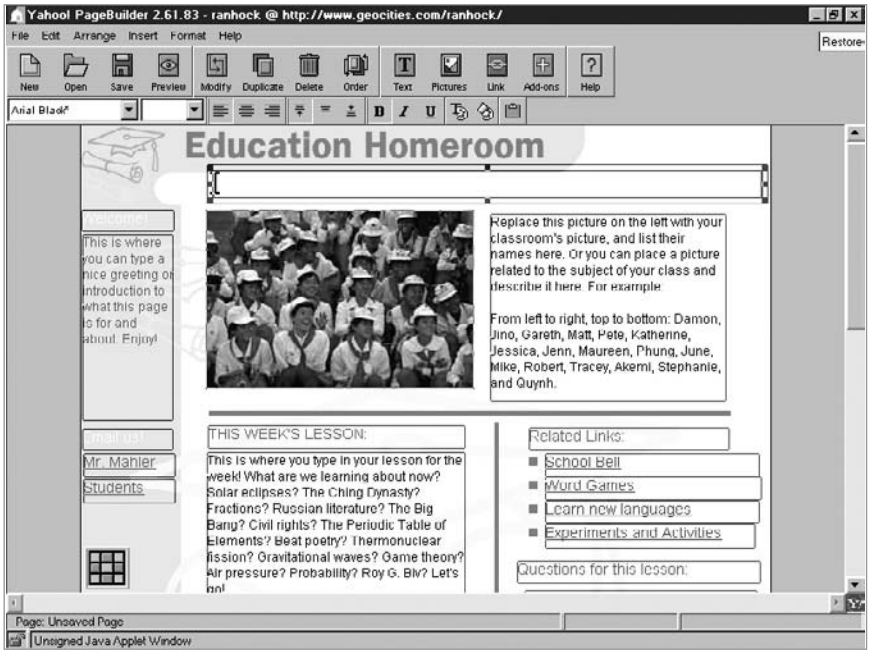


Figure 10.2

Example of a Geocities Template

SITES TO HELP YOU BUILD YOUR WEB SITES

There are thousands of Web sites that provide help in building Web pages. They range from the tutorials already mentioned to sites that provide specific features that you can place on your pages (such as graphics and JavaScript scripts) to sites that bring together a wide collection of a variety of tools. The following three representative sites are sites that the beginner may want to explore, particularly to get a feel for the kind of help that is out there.

Webmonkey

http://hotwired.lycos.com/webmonkey

Webmonkey is especially strong on tutorials for a wide variety of things you might want to place on your page. Look particularly at the Beginners page. Most of the content of this site is written by the Webmonkey staff, and you typically will not find links here to other resources.

Figure 10.3

Webmonkey Beginners Page

Reallybig.com: The Complete Resource for All Web Builders

http://Reallybig.com

Reallybig.com contains over 5,000 links of use to both the beginner and the advanced builder, including resources for "free scripts, CGI, counters, fonts, HTML, Java, clipart, animation, backgrounds, icons, HTML editors, buttons, photographs, site promotion, easy-to-follow Tips and Tricks, and much more."

About.com: Web Design

http://webdesign.about.com

This section of the About.com site contains articles, tips, tutorials, and an excellent collection of links to resources such as clip art collections, JavaScript collections, Web hosting services, legal issues, and so on.

ALTERNATIVES TO YOUR OWN WEB SITE

Two alternatives to easily communicating with large numbers of people are to create a group (see Chapter 5) or to create a Weblog.

The Weblog ("blog") alternative has found much favor in the last few years and requires no more effort (perhaps less) than a free Web site. Discussed earlier (Chapter 8), these tools provide an easy means to gather and distribute news, commentary, and so forth. The main intent is to provide a place for short and frequently updated postings. Although they lack the graphic attractiveness of a Web site, their ease of use has been a major factor in their popularity. For a site that provides free, easily established blogs, try Blogger:

Blogger

http://Blogger.com

Blogger.com provides Weblog space for free, and you can provide the template for your page or use a predesigned one from Blogger. Once you establish a Weblog on Blogger, to publish an item, you just fill out a form and click Publish.

This page intentionally left blank

It is hoped that the preceding chapters have provided some new and useful ideas, information, and sites, even for the very experienced Internet user. My final bit of advice is: "Explore!" As you use the sites I've mentioned, or any site, take a few extra seconds to look around. Poke into the corners of a site, and if it looks very promising, "click everywhere."

—Ran Hock
"The Extreme Searcher"

This page intentionally left blank

The following definitions are in the context of the Internet and are not intended to be more generally applied.

algorithm. A step-by-step procedure for solving a problem or achieving a task. In the context of search engines, the part of the service's program that performs a task such as identifying which pages should be retrieved or ranking pages that have been retrieved.

ALT tag. Text associated with an image, in the HTML code of a page, that can be used to identify the content of the image or for other purposes. Standing for "alternate text," it initially served the purpose of providing a description while waiting for the image to load, but is now used more for other purposes, such as providing a description of the image that can be read by screen-reader applications designed to assist sight-impaired users. In some browsers, you will see this text pop up when you hold your cursor over an image.

AND. The Boolean operator (or connector) that specifies the intersection of sets. When used between words in a search engine query, it specifies that only those records that contain both words (the words preceding and following the "AND") are to be retrieved. For example, the search expression "stomach AND growling" would only retrieve records containing both of those words.

AOL. America On-Line, the most well-known consumer-oriented online service.

applet. A small Java program used on a Web page to perform certain display, computational, or other functions. The origin of the term refers to "small applications programs."

blog. See "Web logs."

bookmark. A feature found in Web browsers—analogous to bookmarks used in a book—that remembers the location of a particular Web page and adds it to a list so the page can be returned to easily. Netscape refers to these as "bookmarks," whereas Internet Explorer uses the term "favorites."

Boolean. A mathematical system of notation created by 19th century mathematician George Boole that symbolically represents relationships between sets (entities). For information retrieval, it uses AND, OR, and NOT (or their equivalents) to identify those records that meet the criteria of having both of two terms within the same record (AND), having either of two terms within the records (OR), or eliminating records that contain a particular term (NOT).

broadband. High-speed data transmission capability. In the home or office context, usually referring to DSL (Digital Subscriber Line), cable, or T1 (or higher) Internet access.

browser. Software that enables display of Web pages by interpreting HTML code, translating it, and performing related tasks. The first widely used browser was Mosaic, which evolved into Netscape. Internet Explorer is the browser developed by Microsoft.

browsing. Examining the contents of a database, Web site, or other electronic document by scanning lists or categories and subcategories. When a site provides this capability, it is referred to as having "browsability."

case-sensitivity. The ability to recognize the difference between uppercase and lowercase alphabetic characters. In information retrieval, it means the difference between possibly being able to recognize White as a name versus white as a color, or AIDS as the disease versus aids as something that provides assistance.

channels. Term used by some online services to organize their services, functions, and Web pages by subject area, often providing selected tools (e.g., calculators), news, links, and other resources relevant to the specific topic.

classification. Arrangement of Web sites by subject area, often using a hierarchical scheme with several levels of categories and subcategories.

concept-based retrieval. Retrieval based on finding records that contain words related to the concept searched for, not necessarily the specific word(s) searched for.

co-occurrence. Occurrence of specific different terms within the same record. Analyzing the frequency of co-occurrence is one technique used to find records that are similar to a selected record.

Cookies. Cookies are small files of information generated by a Web server and stored on the user's computer that are used mostly for personalization of sites.

crawler. See "spider."

dead links. Links that, when clicked, do not work (usually because the page is no longer there or has moved to another URL, or because the URL is incorrect).

diacritical marks. Marks such as accents that are applied to a letter to indicate a different phonetic value.

directory (Web). Collection of Web page records classified by subject to enable easy browsing of the collection. "General" Web directories are those sites that selectively catalog and categorize the broad range of sites available on the Web, usually including only sites that are likely to be of interest to a large number of users.

domain name. The part of a URL (Web address) that usually specifies the organization and type of organization where the Web page is located, e.g., in www.microsoft.com, "microsoft.com" is the domain name. Domain names always have at least two parts, the first part usually identifying the organization or specific machine, the second part ("com" or "uk") identifying the kind of organization or the country.

domain name server. A computer that converts the URL you enter into the numerical address of a domain and identifies the location of the requested computer.

field. A specific portion of a record or Web page, such as title, metatags, URL, etc.

file extension. In a file name, such as letter.doc or house.gif, the part of the file name that follows the period, usually indicating the type of file.

flame wars (flaming). Angry or strongly worded series of messages in Internet groups or mailing lists.

FTP (File Transfer Protocol). Computer protocol (set of instructions) for uploading and downloading files.

Gopher. A menu-based directory allowing access to files from a remote computer. Gophers were supplanted in the mid-1990s by Web tools such as directories and search engines.

home page. The main page of a Web site. Also, the page designated by a user as the page that should be automatically brought up when the user's browser is loaded.

HTML (HyperText Markup Language). The coding language used to create Web pages. It tells a browser how to display a record, including specifications for such things as font, colors, location of images, identification of hypertext links, etc.

Internet. Worldwide network of networks based on the TCP/IP protocol.

Invisible Web. Those pages that are not indexed by Web search engines and therefore cannot be retrieved by means of a search on those engines.

Java. A programming language designed for use on networks, particularly the Internet, that allows programs to be downloaded and run on a variety of platforms. Java is incorporated into Web pages with small applications programs called "applets" that provide features such as animation, calculators, games, etc.

JavaScript. A computer language used to write "scripts" for use in browsers to allow creation of such features as scrolling marquees, etc.

metasearch engines. Search services that search several individual search engines and then combine the results.

metasites. Small, specialized Web directories providing a collection of related links on a specific topic, also know as cyberguides, resource pages, special directories, etc.

metatags. The portion (field) of the HTML coding for a Web page that allows the person creating the page to enter text describing the content of the page. The content of metatags is not shown on the page itself when the page is viewed in a browser window.

NEAR. A proximity connector that is used between two words to specify that a page should be retrieved only when those words are near each other in the page.

nesting. The use of parentheses to specify the way in which terms in a Boolean expression should be grouped, i.e., the order of the operations.

newsgroup. An online discussion group. A group of people and the messages they communicate on a specific topic of interest. More narrowly, the term refers to such a discussion group on Usenet.

NOT. The Boolean operator (connector) that, when used with a term, eliminates the records containing that term.

OR. The Boolean operator (connector) that is used between two terms to retrieve all records that contain either term.

portal. A site that serves as a "gateway" or "starting point" for a collection of Web resources. Portals typically have a variety of tools (such as a search engine, directory, news, etc.) all on a single page designed so that users can designate that page as their "start page" for their browser. Portals are often personalizable regarding content, layout, etc.

precision. In information retrieval, the degree to which a group of retrieved records actually matches the searcher's needs. More technically, precision is the ratio of the number of relevant items retrieved to the total number of items retrieved (multiplied by 100% in order to express the ratio as a percentage). For example, if a query produced 10 records and six of them were judged relevant, the precision would be 60 percent. This is sometimes referred to as relevance.

proximity. The nearness of two terms. Some search engines provide proximity operators, such as NEAR, which allow a user to specify how close two terms must be in order for a record containing those terms to be retrieved.

ranking. The process that determines the order in which retrieved records are displayed. Search engines use algorithms to evaluate records and assign a "score" to records indicating the relative "relevance" of each record. The retrieved records can then be ranked and listed on the basis of those scores.

recall. In information retrieval, the degree to which a search has actually managed to find all the relevant records in the database. More technically, it is the ratio of the number of relevant records that were retrieved to the total number of relevant records in the database (multiplied by 100 percent in order to express the ratio as a percentage). For example, if a query retrieved four relevant records, but there were 10 relevant records in the database, the recall for that search would be 40 percent. Recall is usually difficult to measure because the number of relevant records in a database is often very difficult to determine.

record. The unit of information in a database that contains items of related data. In an address book database, for example, each single record might be the collection of information about one individual person, such as name, address, ZIP code, phone, etc. In the databases of Web search engines, each record is the collection of information that describes a single Web page.

relevance. The degree to which a record matches the user's query (or the user's needs as expressed in a query.) Search engines often assign relevance "scores" to each retrieved record with the scores representing an estimate of the relevance of that record.

search engines. Programs that accept a user's query, search a database, and return to the user the records that match the query. The term is often used more broadly to refer not only to the information retrieval program itself, but also to the interface and associated features, programs, and services.

spider. Programs that search the World Wide Web in order to identify new (or changed) pages for the purpose of adding those pages to a search service's ("search engine's") database.

start page. The page that loads automatically when you open your browser. Also sometimes, confusingly, called your "home page." You select what you want your start page to be by using the "Edit > Preferences" or "Tools > Internet Options" choices on your browser's menu.

stopwords. Small or frequently occurring words that an information retrieval program does not bother to index (ostensibly because the words are "insignificant," but more likely because the indexing of those words would take up too much storage space or require too much processing).

submitted URLs. URLs (Internet addresses) that a person directly submits to a search engine service in order to have that address and its associated Web page added to the service's database.

syntax. The specific order of elements, notations, etc., in which instructions must be submitted to a computer system.

TCP/IP. Transfer Control Protocol/Internet Protocol. The collection of computer data transfer protocols (set of instructions) used on the Internet.

Telnet. A program that lets you log on to and access a remote computer using a text-based interface.

thesaurus. A listing of terms usually showing the relationship between terms, such as whether one term is narrower or broader than another. Thesauri are used in information retrieval to identify related terms to be searched.

thread. Within a group (newsgroup, discussion group, etc.), the series of messages on one specific topic consisting of the original message, replies to that message, replies to those replies, etc.

timeout. The amount of time a system will work on a task or wait for results before ceasing either the task or the waiting.

truncation. Feature in information retrieval systems that allows you to search using the stem or root of a word and automatically retrieve records with all terms that begin with that string of characters. Truncation is usually specified

using a symbol such as an asterisk. For example, in some Web search engines, *town** would retrieve *town, towns, township*, etc.

URL (Uniform Resource Locator). The address by which a Web page can be located on the World Wide Web. URLs consist of several parts separated by periods and, sometimes, slashes.

Usenet. The world's largest system of Internet discussion groups (also called newsgroups).

videotext. Systems, developed in the 1970s, allowing interactive delivery of text and images on television or computer screens. One of the first applications was the delivery of newspaper content.

vortal. A specialized portal. (from Vertical Market Portal).

Web (World Wide Web, WWW). That portion of the Internet that uses the Hypertext Transfer Protocol (http) and its variations to transmit files. The files involved are typically written in some variation of HTML (HyperText Markup Language), thereby viewable using browser software, allowing a GUI (Graphical User Interface), incorporation of hypertext point-and-click navigation of text, and extensive incorporation of images and other types of media and formats.

Web logs. Web sites, usually created by individuals, that are updated frequently, usually provide links to news items elsewhere on the Web and often contain commentary, etc., on a very specific topic.

http://www.extremesearcher.com

Chapter 1

A Brief History of the Internet, version 3.1
http://www.isoc.org/internet-history

Internet History and Growth
http://www.isoc.org/internet/history/2002_0918_Internet_History_and_Growth.ppt

Hobbes' Internet Timeline
http://www.zakon.org/robert/internet/timeline

The Virtual Chase:
Evaluating the Quality of Information on the Internet
http://www.virtualchase.com/quality

Evaluating the Quality of World Wide Web Resources
http://www.valpo.edu/library/evaluation.html

Wayback Machine—Internet Archive
http://www.archive.org

Direct Search
http://www.freepint.com/gary/direct.htm

invisible-web.net
http://www.invisible-web.net

CompletePlanet
http://completeplanet.com

United States Copyright Office
http://lcweb.loc.gov/copyright

Copyright Web Site
http://www.benedict.com

Copyright and the Internet
http://mason.gmu.edu/~montecin/copyright-internet.htm

Karla's Guide to Citation Style Guides
http://bailiwick.lib.uiowa.edu/journalism/cite.html

Style Sheets for Citing Internet & Electronic Resources
http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Style.html

The Resource Shelf
http://resourceshelf.blogspot.com

FreePint
http://www.freepint.com

ResearchBuzz
http://www.researchbuzz.com

Internet Resources Newsletter
http://www.hw.ac.uk/libwww/irn

The Scout Report
http://scout.wisc.edu

Chapter 2

Yahoo!
http://yahoo.com

Open Directory
http://dmoz.org

LookSmart

http://looksmart.com

Librarians' Index to the Internet

http://lii.org

Search Engine Colossus

http://www.searchenginecolossus.com

MSN

http://msn.com

Netscape

http://netscape.com

Excite

http://Excite.com

Lycos

http://lycos.com

Voila!

http://www.voila.fr

Traffick: The Guide to Portals and Search Engines. Frequently Asked Questions about Portals.

http://www.traffick.com/article.asp?aID=9#what

Chapter 3

The WWW Virtual Library

http://vlib.org

Search Engine Guide

http://www.searchengineguide.com

Internet Public Library Reference Ready Reference
http://www.ipl.org/ref/RR

refdesk.com
http://refdesk.com

InfoMine
http://infomine.ucr.edu

BUBL LINK
http://bubl.ac.uk/link

Project Gutenberg
http://www.promo.net/pg

Library of Congress Gateway to Library Catalogs
http://lcweb.loc.gov/z3950/gateway.html

Social Science Information Gateway
http://sosig.esrc.bris.ac.uk

Tennessee Tech History Web Site
http://www2.tntech.edu/history

Virtual Religion Index
http://religion.rutgers.edu/vri

ChemDex
http://www.chemdex.org

HealthFinder
http://www.healthfinder.gov

MEDLINE Plus Health Topics
http://www.nlm.nih.gov/medlineplus/healthtopics.html

EEVL: The Internet Guide to Engineering, Mathematics, and Computing
http://www.eevl.ac.uk

New York Times Cybertimes—A Selective Guide to Internet Business, Financial, and Investing Resources
http://www.nytimes.com/library/cyber/reference/busconn.html

CEOExpress
http://ceoexpress.com

Virtual International Business and Economic Sources
http://libweb.uncc.edu/ref-bus/vibehome.htm

Resources for Economists on the Internet
http://rfe.wustl.edu

WebEc
http://www.helsinki.fi/WebEc

I^3—Internet Intelligence Index
http://www.fuld.com/i3

Governments on the WWW
http://www.gksoft.com/govt

Foreign Government Resources on the Web
http://www.lib.umich.edu/govdocs/foreign.html

FirstGov
http://firstgov.gov

UK Online
http://www.open.gov.uk

Political Resources on the Net
http://www.politicalresources.net

FindLaw
http://www.findlaw.com

Kathy Schrock's Guide for Educators
http://school.discovery.com/schrockguide

Education World
http://education-world.com

Education Index
http://www/educationindex.com

Kidon Media-Link
http://www.kidon.com/media-link

Cyndi's List of Genealogy Sites on the Internet
http://www.cyndislist.com

Chapter 4

AllTheWeb
http://alltheweb.com

AltaVista
http://altavista.com *or* http://av.com

Google
http://www.google.com

HotBot
http://hotbot.com

Teoma
http://teoma.com

Lycos
http://lycos.com

WiseNut
http://www.wisenut.com

MSN Search
http://search.msn.com

Search Engine Watch
http://searchenginewatch.com

Chapter 5

Google Groups
http://groups.google.com

Yahoo! Groups
http://groups.yahoo.com

Delphi Forums
http://www.delphiforums.com

ezboard
http://www.ezboard.com

Topica
http://topica.com

Publicly Accessible Mailing Lists
http://paml.net

L-Soft CataList, the Official Catalog of LISTSERV® lists
http://www.lsoft.com/lists/listref.html

Chapter 6

Encyclopedia.com
http://encyclopedia.com

Encarta
http://encarta.msn.com

Voila Encyclopédie avec Hachette
http://encyclo.voila.fr

Encyclopedia Britannica
http://britannica.com

YourDictionary.com
http://www.yourdictionary.com

Merriam-Webster Online
http://www.m-w.com

Dictionnaire Universel Francophone En Ligne
http://www.francophonie.hachette-livre.fr

diccionarios.com
http://www.diccionarios.com

LEO—Link Everything Online
http://dict.leo.org

InfoPlease
http://www.infoplease.com

Wayp International White and Yellow Pages
http://www.wayp.com

Yahoo! People Search
http://people.yahoo.com

AnyWho
http://www.anywho.com

Quote Links
http://www.quotationspage.com

Bartleby
http://www.bartleby.com

Yahoo! Finance—Currency Conversion
http://finance.yahoo.com/m3?u

Weather Underground
http://wunderground.com

The Perry-Castañeda Library Map Collection
http://www.lib.utexas.edu/maps

Global Gazetteer
http://www.world-gazetteer.com

U.S. Postal Service
http://www.usps.com/zip4

CNNMoney
http://money.cnn.com

Statistical Resources on the Web—Comprehensive Subjects
http://www.lib.umich.edu/govdocs/stcomp.html

Statistics.com
http://www.statistics.com

The Directory of Online Statistics Sources
http://www.berinsteinresearch.com/stats.htm

USA Statistics in Brief
http://www.census.gov/statab/www/brief.html

FedStats
http://www.fedstats.gov

Amazon.com
http://www.amazon.com

Barnes & Noble
http://www.barnesandnoble.com

Library of Congress
http://catalog.loc.gov

British Library
http://blpc.bl.uk

The Online Books Page
http://digital.library.upenn.edu/books

Project Gutenberg
http://www.promo.net/pg

Bartleby.com
http://www.bartleby.com

EuroDocs: Primary Historical Documents from Western Europe
http://library.byu.edu/~rdh/eurodocs

A Chronology of U.S. Historical Documents
http://www.law.ou.edu/hist

University of Virginia Hypertext Collection
http://xroads.virginia.edu/~HYPER/hypertex.html

Governments on the WWW
http://www.gksoft.com/govt

Foreign Government Resources on the Web
http://www.lib.umich.edu/govdocs/foreign.html

CIA World Factbook
http://www.odci.gov/cia/publications/factbook

CountryWatch.com
http://www.countrywatch.com

FirstGov
http://firstgov.gov

GPO Access
http://www.gpoaccess..gov

THOMAS: Legislative Information on the Internet
http://thomas.loc.gov

Library of Congress—State and Local Governments
http://lcweb.loc.gov/global/state/stategov.html

UK Online
http://www.open.gov.uk

CorporateInformation
http://www.corporateinformation.com

Hoovers
http://www.hoovers.com

D&B Small Business Solutions
http://sbs.dnb.com

D&B Express Online
http://www.dnbsearch.com

Thomas Register
http://www2.thomasregister.com

American Society of Association Executives Gateway to Associations
http://info.asaenet.org/gateway/OnlineAssocSlist.html

AMA Physician Select—Online Doctor Finder
http://www.ama-assn.org/aps/amahg.htm

Lawyers.com
http://lawyers.com

Direct Search
http://www.freepint.com/gary/direct.htm

A grab bag of (mainly) free bibliographies and bibliographic databases on the Web
http://www.leidenuniv.nl/ub/biv/freebase.htm

ingenta
http://www.ingenta.com

Peterson's
http://petersons.com

College Search
http://www.collegeboard.com/csearch

Fodors.com
http://www.fodors.com

Lonely Planet Online
http://www.lonelyplanet.com

Travelocity.com
http://travelocity.com

Expedia.com
http://expedia.com

Orbitz
http://orbitz.com

Internet Movie Database
http://www.imdb.com

Chapter 7

Finding Images Online: Directory of Web Image Sites
http://www.berinsteinresearch.com/fiolinks.htm

Digital Librarian:
A Librarian's Choice of the Best of the Web: Images
http://www.digital-librarian.com/images.html

BUBL LINK: Image Collections
http://bubl.ac.uk/link/types/images.htm

Google
http://google.com

AltaVista
http://altavista.com

AllTheWeb
http://alltheweb.com

Corbis
http://corbis.com

American Memory Project
http://memory.loc.gov

WebMuseum, Paris
http://ibiblio.org/wm (more specifically,
http://www.ibiblio.org/wm/paint)

Barry's Clipart
http://barrysclipart.com

Yahoo! Directory > Graphics > Clipart
http://dir.yahoo.com/Computers_and_Internet/Graphics/Clip_Art

World Wide Web Virtual Library: Audio
http://archive.museophile.sbu.ac.uk/audio

Radio-Locator (formerly The MIT List of Radio Stations on the Internet)
http://www.radio-locator.com

The History Channel: Speeches
http://www.historychannel.com/speeches

The Movie Sounds Page
http://www.moviesounds.com

BUBL LINK / 5:15 Catalogue of Internet Resources: Video
http://bubl.ac.uk/link/v/video.htm

Chapter 8

Kidon Media-Link
http://www.kidon.com/media-link

NewsLink
http://newslink.org

Metagrid
http://www.metagrid.com

BBC
http://news.bbc.co.uk

CNN
http://www.cnn.com

CNNMoney
http://money.cnn.com

MSNBC
http://www.msnbc.com

Reuters

http://reuters.com

Radio-Locator (formerly The MIT List of Radio Stations on the Internet)

http://www.radio-locator.com

NPR

http://www.npr.org

World News Network

http://www.wn.com

Moreover.com

http://moreover.com

Newsnow.co.uk

http://newsnow.co.uk

AllTheWeb News Search

http://alltheweb.com

AltaVista News Search

http://altavista.com

Google News Search

http://news.google.com

Open Directory: Computers: Internet: On the Web: Weblogs

http://dmoz.org/Computers/Internet/On_the_Web/Weblogs

NewsAlert

http://www.newsalert.com

Google News Alerts

http://www.google.com/newsalerts

Chapter 9

ShoppingSpot
http://Shoppingspot.com

ThomasRegister
http://www.thomasregister.com

Yahoo! Shopping
http://shopping.yahoo.com

Amazon.com
http://amazon.com

eBay
http://ebay.com

Froogle
http://froogle.com

Open Directory: Consumer Information: Price Comparisons
http://dmoz.org/Home/Consumer_Information/Price_Comparisons

My Simon
http://www.mysimon.com

Epinions
http://epinions.com

BizRate.com
http://bizrate.com

Consumer Reports
http://consumerreports.org

ConsumerReview.com
http://consumerreview.com

Consumer Search
http://consumersearch.com

safeshopping.org
http://safeshopping.org

E-Commerce and the Internet
http://www.ftc.gov/bcp/menu-internet.htm.

eConsumer
http://eConsumer.gov

Chapter 10

Open Directory
http://dmoz.org

GeoCities
http://geocities.com

Tripod
http://tripod.com

Angelfire
http://angelfire.com

Webmonkey
http://hotwired.lycos.com/webmonkey

Reallybig.com:
The Complete Resource for All Web Builders
http://Reallybig.com

About.com: Web Design
http://webdesign.about.com

Blogger
http://Blogger.com

This page intentionally left blank

RANDOLPH HOCK, PH.D.

Randolph "Ran" Hock is the principal of Online Strategies, a company that specializes in creating and delivering customized courses on Web research. His courses have been delivered to large and small corporations, government agencies, nongovernmental organizations, universities, and associations. He has trained searchers throughout the U.S., Austria, France, Hungary, Portugal, Switzerland, and the U.K. Ran has been a chemistry teacher and a chemistry librarian (at MIT), and was the first Data Services Librarian at the University of Pennsylvania. For many years he held training and management positions at DIALOG Information Services and Knight-Ridder Information. Ran is the author of *The Extreme Searcher's Guide to Web Search Engines* (CyberAge Books. First edition, 1999. Second edition, 2001). He lives in Vienna, Virginia, with his wife and two children, and hopes to someday have time to again pursue his hobby of genealogy.

This page intentionally left blank

B

C

D

E

F

G

H

I

J

K

L

M

N

O

S

T

U

V

W

Y

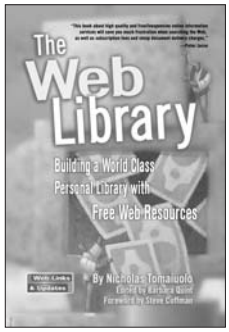
Z

This page intentionally left blank

More CyberAge Books from Information Today, Inc.

The Web Library

Building a World Class Personal Library with Free Web Resources



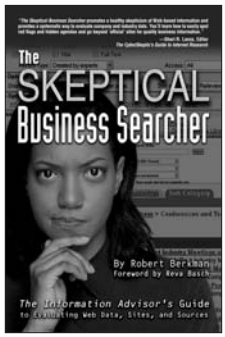
By Nicholas G. Tomaiuolo • Edited by Barbara Quint

With this remarkable, eye-opening book and its companion Web site, Nicholas G. (Nick) Tomaiuolo shows how anyone can create a comprehensive personal library using no-cost Web resources. And when Nick say "library," he's not talking about a dictionary and thesaurus on your desktop: He means a vast, rich collection of data, documents, and images that—if you follow his instructions to the letter—can rival the holdings of many traditional libraries. If you were to calculate the expense of purchasing the hundreds of print and fee-based electronic publications that are available for free with "the Web Library" you'd quickly recognize the potential of this book to save you thousands, if not millions, of dollars. (Fortunately, Nick does the calculating for you!) This is an easy-to-use guide, with chapters organized into sections corresponding to departments in a physical library. *The Web Library* provides a wealth of URLs and examples of free material you can start using right away, but best of all it offers techniques for finding and collecting new content as the Web evolves. Start building your personal Web library today!

2003/440 pp/softbound/ISBN 0-910965-67-6 • $29.95

The Skeptical Business Searcher

The Information Advisor's Guide to Evaluating Web Data, Sites, and Sources



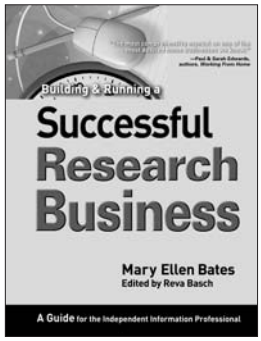
By Robert Berkman • Foreword by Reva Basch

This is the experts' guide to finding high-quality company and industry data on the free Web. Information guru Robert Berkman offers business Internet users effective strategies for identifying and evaluating no-cost online information sources, emphasizing easy-to-use techniques for recognizing bias and misinformation. You'll learn where to go for company backgrounders, sales and earnings data, SEC filings and stockholder reports, public records, market research, competitive intelligence, staff directories, executive biographies, survey/poll data, news stories, and hard-to-find information about small businesses and niche markets. The author's unique table of "Internet Information Credibility Indicators" allows readers to systematically evaluate Web site reliability. Supported by a Web page.

2003/300 pp/softbound/ISBN 0-910965-66-8 • $29.95

Building and Running a Successful Research Business

A Guide for the Independent Information Professional



By Mary Ellen Bates • Edited by Reva Basch

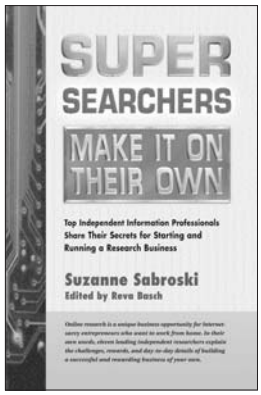
This is the handbook every aspiring independent information professional needs to launch, manage, and build a research business. Organized into four sections, "Getting Started," "Running the Business," "Marketing," and "Researching," the book walks you through every step of the process. Author and long-time independent researcher Mary Ellen Bates covers everything from "is this right for you?" to closing the sale, managing clients, promoting your business, and tapping into powerful information sources.

"The most comprehensive manual on one of the most desired home businesses we know."
—Paul & Sarah Edwards,
authors, *Working From Home*

2003/360 pp/softbound/ISBN 0-910965-62-5 • $29.95

Super Searchers Make It On Their Own

Top Independent Information Professionals Share Their Secrets for Starting and Running a Research Business



By Suzanne Sabroski • Edited by Reva Basch

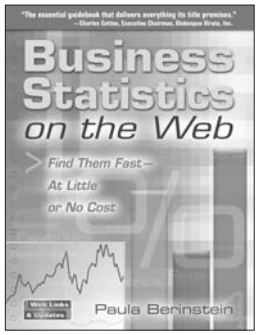
If you want to start and run a successful Information Age business, read this book. Here, for the first time anywhere, 11 of the world's top research entrepreneurs share their strategies for starting a business, developing a niche, finding clients, doing the research, networking with peers, and staying up-to-date with Web resources and technologies. You'll learn how these super searchers use the Internet to find, organize, analyze, and package information for their clients. Most importantly, you'll discover their secrets for building a profitable research business.

"What do you get when the Information Age meets Free Agent Nation? A new breed of entrepreneurs called 'independent information professionals.' In Super Searchers Make It On Their Own, *Suzanne Sabroski beautifully captures the wisdom and spirit of these pioneers. Her smart and useful book is the ideal guide to succeeding in this exciting new field."*
—Daniel H. Pink,
author of *Free Agent Nation: The Future of Working for Yourself*

2002/336 pp/softbound/ISBN 0-910965-59-5 • $24.95

Business Statistics on the Web

Find Them Fast—At Little or No Cost



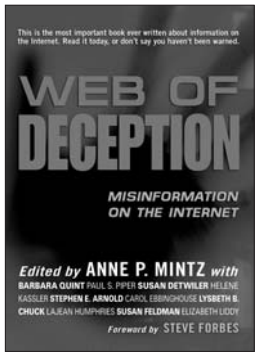
By Paula Berinstein

Statistics are a critical component of business and marketing plans, press releases, surveys, economic analyses, presentations, proposals, and more—yet good statistics are notoriously hard to find. In this practical guide, statistics guru Paula Berinstein (author of six previous books including *Finding Statistics Online* and *The Statistical Handbook on Technology*) shows readers how to use the Net to find statistics about companies, markets, and industries, how to organize and present statistics, and how to evaluate them for reliability. Here are dozens of easy-to-use tips and techniques for manuevering around obstacles to find the statistics you need. Supported by a Web page.

2003/240 pp/softbound/ISBN 0-910965-65-X • $29.95

Web of Deception

Misinformation on the Internet



Edited by Anne P. Mintz • Foreword by Steve Forbes

Intentionally misleading or erroneous information on the Web can wreak havoc on your health, privacy, investments, business decisions, online purchases, legal affairs, and more. Until now, the breadth and significance of this growing problem for Internet users had yet to be fully explored. In *Web of Deception*, Anne P. Mintz (Director of Knowledge Management at Forbes, Inc.) brings together 10 information industry gurus to illuminate the issues and help you recognize and deal with the flood of deception and misinformation in a range of critical subject areas. A must-read for any Internet searcher who needs to evaluate online information sources and avoid Web traps.

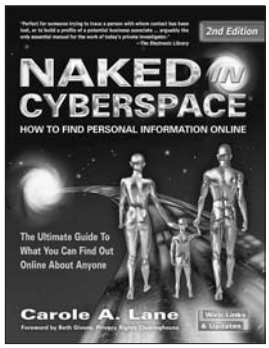
"Experts here walk you through the risks and traps of the Web world and tell you how to avoid them or to fight back … Anne Mintz and her collaborators have done us a genuine service."

—Steve Forbes,
from the foreword

2002/278 pp/softbound/ISBN 0-910965-60-9 • $24.95

Naked in Cyberspace

How to Find Personal Information Online, 2nd Edition



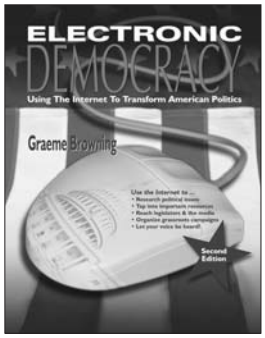
By Carole A. Lane • Foreword by Beth Givens

In this fully revised and updated second edition of her bestselling guide, author Carole A. Lane surveys the types of personal records that are available on the Internet and online services. Lane explains how researchers find and use personal data, identifies the most useful sources of information about people, and offers advice for readers with privacy concerns. You'll learn how to use online tools and databases to gain competitive intelligence, locate and investigate people, access public records, identify experts, find new customers, recruit employees, search for assets, uncover criminal records, conduct genealogical research, and much more. Supported by a Web page.

2002/586 pp/softbound/ISBN 0-910965-50-1 • $29.95

Electronic Democracy, 2nd Edition

Using the Internet to Transform American Politics



By Graeme Browning
Foreword by Adam Clayton Powell III

In this new edition of *Electronic Democracy,* award-winning journalist and author Graeme Browning details the colorful history of politics and the Net, describes the key Web-based sources of political information, offers practical techniques for influencing legislation online, and provides a fascinating, realistic vision of the future.

"By harnessing the power of the Internet to inform, organize, and advocate, Americans can use technology to broaden and deepen their role in our representative democracy. Combining political savvy with computer know-how, Graeme Browning shows us how."

—Bill Bradley

2001/260 pp/softbound ISBN 0-910965-41-2 • $19.95