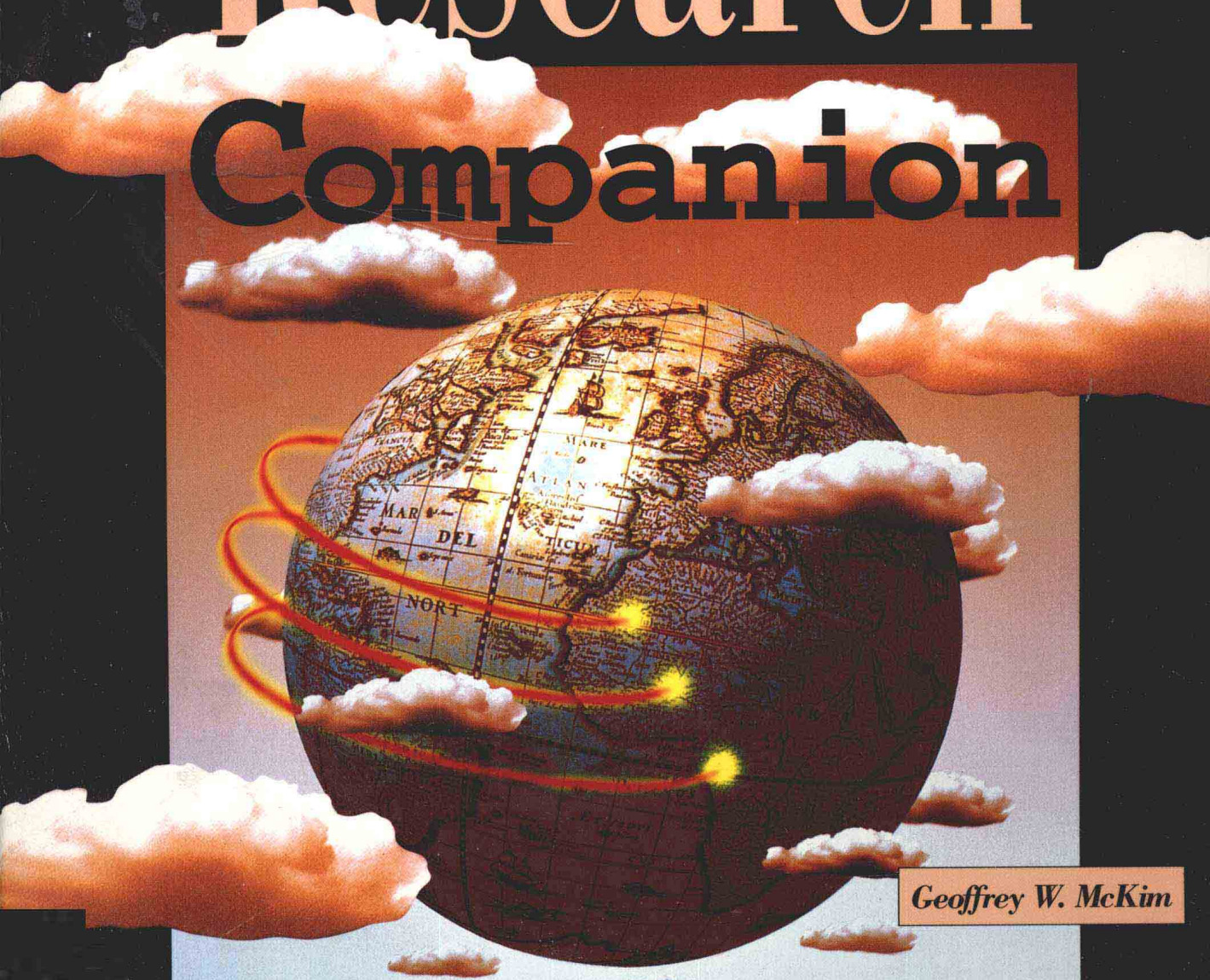
Internet Research Companion



Geoffrey W. McKim

Making the right connections and discoveries

Internet Research Companion

Geoffrey W. McKim
Indiana University

Internet Research Companion

Copyright © 1996 by Que® Education & Training

All rights reserved. Printed in the United States of America. No part of this book may be used or reproduced in any form or by any means, or stored in a database or retrieval system, without prior written permission of the publisher except in the case of brief quotations embodied in critical articles and reviews. Making copies of any part of this book for any purpose other than your own personal use is a violation of United States copyright laws. For information, address Que College, Macmillan Computer Publishing, 201 W. 103rd Street, Indianapolis, IN 46290.

Library of Congress Catalog No.: 95-74875

ISBN: 1-57576-050-9

This book is sold *as is*, without warranty of any kind, either express or implied, respecting the contents of this book, including but not limited to implied warranties for the book's quality, performance, merchantability, or fitness for any particular purpose. Neither Que Corporation nor its dealers or distributors shall be liable to the purchaser or any other person or entity with respect to any liability, loss, or damage caused or alleged to be caused directly or indirectly by this book.

98 97 96 4 3 2 1

Interpretation of the printing code: the rightmost double-digit number is the year of the book's printing; the rightmost single-digit number, the number of the book's printing. For example, a printing code of 96-1 shows that the first printing of the book occurred in 1996.

Screens reproduced in this book were created using Collage Plus from Inner Media, Inc., Hollis, NH.

Publisher: David P. Ewing

Associate Publisher: Chris Katsaropoulos

Product Marketing Manager: Susan J. Dollman

Managing Editor: Sheila Cunningham

Acquisitions Editor
Satbir Bedi

Acquisitions Coordinator
Elizabeth D. Brown

Production Editor
Virginia Noble

Technical Editor
Mary Beth Maddox

Cover Designer
Anne Jones

Production Team
Mona Brown
Chris Cleveland
Mike Dietsch
Jason Hand
Sonja Hart
Ayanna Lacey
Clint Lahnen
Paula Lowell
Steph Mineart
Bobbi Satterfield
SA Springer
Andy Stone
Mark Walche

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Que cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark. Microsoft, Microsoft Windows, and MS-DOS are registered trademarks of Microsoft Corporation. UNIX is a registered trademark of UNIX System Laboratories, Inc. Mac and the Macintosh logo are registered trademarks of Apple Computer, Inc.

About the Author

Geoffrey McKim has developed, managed, and taught Internet-based information resources for over five years. He holds a degree in Library and Information Science from the Indiana University School of Library and Information Science. He continues to teach courses and workshops on UNIX, computer skills, Internet skills, digital libraries, and World Wide Web disign and programming. Mr. McKim is currently the Director of Information Technology at the Indiana University School of Library and Information Science, and his research interests include Computer Supported Cooperative Work.

Acknowledgments

I would first and foremost like to thank Amy Cornell for all of her advice, patience, and constant support. Without her editing skill, incisive questioning, and love, this book would not have been possible.

I would also like to thank Julie Fore, also without whom this project would not have happened. Her Internet expertise, her trenchant and perceptive understanding of user behavior, and her friendship have all been invaluable throughout the process.

A generous measure of thanks also goes to the editing team at Que Education and Training. Thanks to Satbir Bedi for managing the whole project, for keeping me on task, and for giving me the opportunity to write this book. Thanks also to Ginny Noble for making sure that what I write makes sense to anyone else, and for her always enjoyable editor's comments. Finally, thanks to Mary Beth Maddox, both for her technical editing and for introducing me to Que Education and Training.

I would like to thank my colleagues at the Indiana University School of Library and Information Science, in particular Blaise Cronin and Howard Rosenbaum, for providing a dynamic and exciting environment, in which the information and communications provided by the Internet play an integral role.

Finally, I would like to thank my parents, William and Karen McKim, for always having challenged me and for having provided a wonderful home environment, in which learning was promoted and supported. Not a day goes by in which I don't recognize their influence in what I do.

Composed in *Sabon* and *MCPdigital* by Que Corporation

Table of Contents

Introduction 1
How This Book Is Organized ... 3
Who This Book Is For ... 4
Conventions Used in This Book ... 4

Part I Tools of the Trade: The Technologies of the Internet 6

1 Internet Access Tools 6
Issues to Consider ... 7
Telnet—Terminal Access to the Internet ... 9
FTP—Transferring Files over the Internet ... 13
Gopher—Accessing Networked Information through Menus ... 14
Referring to Internet Resources through Uniform Resource Locators (URLs) ... 15
World Wide Web—Networked Hypermedia ... 17

2 Data Formats and Media 25
Beyond Text—Media on the Internet ... 26
Graphics on the Internet ... 26
Sound on the Internet ... 28
Page Description Languages ... 28
Compression and Data Files ... 30

Part II Retrieving and Working with Information on the Internet 33

3 Using the Web to Retrieve Information from the Internet 33
Using a URL to Retrieve Information on the Web ... 34
Moving Around in a Web Document ... 37
Searching within a Web Document ... 44
Saving Documents Retrieved from the Web ... 47
Printing Documents Retrieved from the Web ... 50
Retrieving Non-Web Resources with a Web Browser ... 51

4 How to Find It Again: Bookmarks 54
The Basics of Bookmarks ... 55
Creating Bookmarks with Netscape ... 56
Creating Bookmarks with Lynx ... 58

5 Using FTP to Retrieve Information from the Internet 62
FTP Clients ... 63
Retrieving a File with FTP ... 63

此为试读,需要完整PDF请访问: www.ertongbook.com

6 Finding Information on the World Wide Web: Indexes and Searching 71
Search Engines on the Web ... 73
Searching and Browsing ... 73
Indexes and Indexing ... 74
Doing a Sample Index Search with WebCrawler ... 76
Full-Text Indexing versus Title-Only Indexing ... 81
Free-Text Indexing versus Controlled-Vocabulary Classification ... 82
Doing Sample Directory Searches with Yahoo and the WWW Virtual Library ... 83

7 Refining Internet Searches: Search Engines and Boolean Logic 89
AND Queries ... 90
OR Queries ... 92
NOT Queries ... 94
Combining Boolean Queries ... 95
Truncation ... 96
Beyond Boolean: Relevance Rankings ... 98
WAIS Databases ... 99

8 Search Engines for Searching on the Internet 100
How Search Engines Differ ... 101
Web Directories of Resources ... 101
Web-Based Search Engines ... 116
A Gopher Search Engine ... 125
Common Problems of Searching on the Internet ... 131

9 Think Globally, Act Locally: Subject Guides to the Internet 134
Resource Guides to Subject-Specific Information ... 135
Lists of Frequently Asked Questions (FAQ Lists) ... 140
Home Pages for Organizations ... 141
Strategies for Finding Useful Resources on the Internet ... 142

10 Using Library Catalogs on the Internet 144
Why You May Want to Use a Library Catalog ... 145
What an Online Library Catalog Can and Cannot Do ... 145
Using a Library Catalog ... 146
Accessing an Online Library Catalog ... 147
Locating Library Catalogs with HYTELNET ... 153
Using HYTELNET to Access Library Catalogs ... 155

11 Useful Commercial and Government Resources 161
UnCover: A Document Delivery Service ... 162
LEXIS-NEXIS ... 168
Dialog ... 169
Dow Jones News/Retrieval ... 169

Library of Congress Services 170
FedWorld 172

Part III People as Resources 174

12 Using E-Mail on the Internet 174
Basics of Internet Mail 175
About E-Mail Addresses 176
Basics of the PINE E-Mail Program 178
E-Mail Etiquette 193

13 Locating People 195
The Basic Strategy 196
The NetFind Tool 203

14 Electronic Discussion Groups 205
Introducing LISTSERV and Majordomo 206
Working with LISTSERVs 207
Working with Majordomo Lists 210
Locating the Right Discussion Group 212
Choosing an Appropriate Discussion Group 218
Guidelines for Participating in Electronic Discussion Groups 219
Another Source for Finding Discussion Groups: The Inter-Links/Dartmouth List 220
Terms and Abbreviations Used in Discussion Groups 220

15 Usenet Newsgroups: Electronic Bulletin Boards 223
The Structure of Usenet 224
Accessing Usenet Newsgroups 226
Using Netscape to Read Usenet News 227
Proper Conduct on Usenet 233
Frequently Asked Questions on Usenet 234
Finding Appropriate Usenet Newsgroups 235
Using Other Web Browsers to Access Usenet Newsgroups 235

Part IV Evaluation and Citation 237

16 Evaluation of Internet Resources 237
The Importance of Evaluating Internet Resources 238
Everyone a Librarian 239
Preevaluated Resources: Library Collections and E-Journals 242

17 Giving Credit: Citing Internet Resources in Your Projects 247
What Is a Citation and Why Give It? 248
Internet Citation Styles 249

Appendix: Where to Find It 253

Introduction

Electronic information on the Internet has recently become useful for college students who do research and write term papers. These Internet-based resources may serve as primary material, providing new and original research, or as reference material, bringing together existing materials in the form of annotated bibliographies, directories, atlases, and resource lists. Some instructors are even recommending that students consult certain Internet sites for information.

For college students, Internet-based information has many advantages over traditional paper-based materials. First, electronic resources have the potential of being up-to-date in a way that was impossible in the print world, even with periodicals. Students now have the opportunity to do original and innovative research, beyond what was traditionally expected of them. Second, electronic resources are searchable, enabling students to access and retrieve easily the exact information they need. Third, as the Internet becomes more accessible in both the home and the

residence hall, students have more flexibility in where and when they do research. Finally, the Internet enables students to determine the scope and breadth of a particular subject or discipline with ease, giving them a snapshot of the relevant issues, documents, and centers of activity.

Along with the advantages, however, come some disadvantages. First, although Internet resources have the potential to be current, they frequently are not. Quite often, the managers of Internet-based resources maintain them for fun or out of professional interest, so the managers have little incentive to keep the resources up-to-date. This problem is magnified because you can't always tell when an electronic resource is up-to-date. You can't simply check the date of publication, as you can with a published work.

The most serious problem is the lack of quality control in resources found on the Internet. Although inaccuracy is not a stranger to the print world, the publisher of print materials at least puts an *imprimatur*, or seal of quality, on the work; in fact, the publisher's reputation depends on the quality of the material published. But on the Internet anyone can be a publisher, from a large media company to a high school student. Because this publisher seal of quality no longer exists, the student researcher must play the unfamiliar role of evaluator. The student must be responsible for assessing the quality of the resources.

Certain academic disciplines and areas of study present students with a rich and useful set of information resources on the Internet. The fields of government, computing and computer technology, and popular culture are especially notable. However, students in all areas of study—from postmodern literary studies to medieval history to genetics to political science—can benefit academically from the vast number of resources on the Internet.

Of course, the challenge to the student who wants to make use of these new sources of information is to *find* them. Greg Rawlins, in an influential paper (published on the Internet), referred to three emerging roles—what he called "mapmakers, filters, and ferrets"—that people and programs will have to assume in order to make use of information in the new networked, electronic world. This book is about the third role, that of ferret. Students now have both a great opportunity and a great incentive to become "information ferrets" on the Internet, to be able to search for and retrieve information that they can read, synthesize, and incorporate in their research and term papers.

Becoming such seekers of information will require that students pick up a new set of skills. First, they must learn to use new computer tools, ones that enable them to access electronic resources. Second, students must learn how to search

for information, particularly when the potential number of sources is open-ended. Many students leave high school never knowing how to use information sources beyond the encyclopedia and *Readers' Guide to Periodical Literature*. Finally, students must learn to evaluate the quality of the resources they find on the Internet. If the quality is poor, students will not want to rely on such resources for papers.

Students do not need to learn these skills by themselves. Many sources can help. Books are available on the search and access tools. The first two chapters of this book, in fact, discuss some of these tools. College and university librarians can be an invaluable resource for students wanting to learn to use Internet-based tools and materials. Although librarians can help students use the tools, these professionals can be of most help in knowing where to find information, in guiding students to the best resources and avoiding those of poor quality. Many libraries have even hired librarians whose sole job is to work with Internet-based resources. Every student should think of the librarian as one of the primary sources of assistance in finding, retrieving, evaluating, and using information on the Internet.

How This Book Is Organized

This book is divided into four parts. Part I, "Tools of the Trade: The Technologies of the Internet," describes the programs that enable you, the student researcher, to access information on the Internet. These tools include Telnet, FTP, Gopher, and the World Wide Web. Because the World Wide Web is clearly the most useful tool, much of the discussion focuses on the Web.

Part II, "Retrieving and Working with Information on the Internet," shows you how to find and retrieve documents that may be of use in researching and writing papers.

Part III, "People as Resources," focuses also on finding information on the Internet, but the focus is not on documents. It is on getting information from people—an equally (if not more) important information resource. Tools such as e-mail, LISTSERVs, electronic discussion groups, and Usenet newsgroups (electronic bulletin boards) are discussed in Part III.

Part IV, "Evaluation and Citation," deals with what you do after you have the information. You learn how to evaluate Internet information and how to incorporate citations to such Internet information in your bibliographies and footnotes.

An appendix is also included, which lists the Internet addresses of the research tools discussed in the book.

Who This Book Is For

This book is not a comprehensive guide to using the Internet, or even to using a particular part of it. Instead, this book is aimed at students who want to make use of the Internet as a research tool in their studies. Students who want to become technologists or Internet "gurus" will probably not benefit from this book; in fact, they may know most of the information in it. This book is for the rest of us, those who want to use the Internet for research but may be a little bewildered by the scope and breadth of information available, as well as the difficulty in learning the necessary tools.

In writing this book, I followed two guidelines:

- Don't "dumb down" explanations, but don't include unnecessary technical detail.
- Present the most valuable techniques and resources, and don't include those of limited value or those whose difficulty of use outweighs any potential value.

If your favorite search tool is not covered in this book, keep in mind that I included only those resources and tools that I believe are of the most value for the effort expended in learning them.

I also maintained a scholastic focus throughout the book. There are already plenty of books available on recreational materials on the Internet, and most students will find these on their own. The techniques and tools presented in this book will increase your academic productivity and give you access to a variety of materials and resources for use in researching and writing papers.

Conventions Used in This Book

Certain conventions are followed in this book to help you easily understand the information presented.

Words or phrases defined for the first time appear in *italic*. Words or phrases that you are asked to type are in **boldface**. Screen displays and on-screen messages appear in a special `monospace` typeface. Internet addresses appear in **`monospace and boldface`**.

此为试读,需要完整PDF请访问: www.ertongbook.com

Keys are represented as they appear on your keyboard. Key combinations, such as Ctrl+B and Alt+Enter, are connected by a plus sign (+). In menu commands and dialog box options, access keys are in boldface and underlined. An example is the Open Location command on the File menu.

The Enter key is often mentioned in this book. Note that this key may be labeled differently on your keyboard. On Macintosh keyboards and on many older PC keyboards, this key is labeled Return. If you are using such a keyboard, simply press the Return key whenever you are instructed to press the Enter key.

Although the Macintosh keyboard has a key labeled Control, this key isn't used as frequently as it is on the PC. Most often, Macintosh commands require that you use the Command key instead of the Control key. Please note that the Command key is often called the "apple" key or the "cloverleaf" (⌘) key.

1

Internet Access Tools

Objectives

After reading this chapter, you will be able to

- Understand the differences between host-based and direct Internet connections
- Recognize the basic Internet tools: Telnet, FTP, Gopher, and the World Wide Web
- Understand how Uniform Resource Locators (URLs) are used to refer to Internet resources

This book is not solely about the technologies of the Internet, as there are many excellent books already on the market about that topic. However, an understanding of a few key technological issues is critical to making full use of this book. This chapter presents several technological issues and introduces four key technologies: FTP (File Transfer Program), Telnet, Gopher, and the World Wide Web. Although some of these technologies are

newer than others (in particular, the World Wide Web has become dominant), all of them can be useful at different times for student research.

Issues to Consider

To make use of the Internet and its technologies, you first need to consider three important issues. These issues affect how the Internet tools *look* to you and thus how you *read* some of the text. This discussion covers the following key issues: the type of Internet connection you use, the Internet software available to you, and the type of computer you have.

Direct Internet Access versus Host-Based Internet Access

You can access the Internet in one of two modes: host-based mode and direct-access mode. In host-based mode, all interaction with the Internet is done in the form of lines of text. Generally, this mode is associated with accessing the Internet through other online systems, such as local bulletin boards, campus computers running UNIX, and some commercial online services. Direct-access mode, which is becoming more common, has the distinct advantage of enabling you to view multimedia information, such as sound and movies, as well as text. Interaction with the Internet in direct-access mode is often in the form of mouse clicks instead of typed text. In addition, you usually use graphical user interfaces such as Microsoft Windows, Microsoft Windows 95, and the Macintosh.

Internet information access programs such as Mosaic, Netscape, TurboGopher, WinGopher, and HGopher require that you have access to a direct Internet connection of some sort. Because much of the information available on the Internet is composed of multiple media, having graphical access is an advantage. Although the information available is generally the same with host-based access, you won't be able to view any graphics or movies in host-based mode (at least not with ease); nor will you be able to listen to sounds.

To make use of direct-access Internet tools, you need a couple of things. First, you need to be using an operating system that provides a graphical environment, such as Windows 95 or the Macintosh. (The X Window environment on UNIX machines will also work, but if you are using that, you probably already know how to use the Internet!) Second, you need some sort of network connection to the Internet.

If you are accessing the Internet from an on-campus lab or classroom, or if you are in a residence hall that provides a network connection, you probably have a direct network connection. Otherwise, you can still have a network connection even if you are dialing in to a network with a modem over the telephone lines from home. In that case, you need to be using either PPP (point-to-point protocol) or SLIP (serial-line Internet protocol). Both of these systems enable your computer to pretend that it actually has a network connection to the Internet. Your campus computing staff should be able to tell you whether SLIP or PPP access is available at your site.

If you don't have access to a network connection to the Internet, you are probably dialing in through some sort of terminal software (like ProComm, Terminal, Smartcom, or Zterm). In addition, you are most likely using the Internet tools on a computer running a computer operating system called UNIX. (This doesn't mean that you have a computer running UNIX sitting on your desk; it just means that the computer you are dialing into is running UNIX.) In this case, you are limited to text-only tools, although you still should have access to all the resources.

NOTE ***This book is directed to both direct-access and host-based Internet users. Although direct-access and host-based modes may look quite different (primarily because direct-access mode is graphics-based), all the examples in this book illustrate basic principles that are valid no matter which mode you are using. Don't panic if things don't look exactly the same for you—you should still benefit from the basic tasks presented in these chapters.***

The Internet Tools Available to You

Not all Internet tools are available on all systems. If you have a direct connection to the Internet, you are capable of running all the available tools. In that case, there is no reason why you shouldn't have one of each type of tool discussed in this book. (The appendix describes where to obtain various types of Internet software.)

If, however, you have a host-based connection to the Internet, usually through a UNIX computer, you will have available only what the system administrator put on the system. If you are lucky enough to have a World Wide Web browser (usually Lynx—if you have access through a UNIX computer), you can access almost any Internet resource. Much of this book deals with World Wide Web–based resources. Yet even if you cannot access the World Wide Web but have

access to all the other tools (Telnet, FTP, and Gopher), there is still much that you can do, using this book for assistance.

The Type of Computer You Use

Although it would make life easier if we all had the same kind of computer, that is not the nature of the marketplace, and ultimately, we are probably the better for it. However, writing computer books is certainly more challenging. The fact is that some people use Windows-based computers to access the Internet, others use Macintoshes, and still others use nongraphical UNIX accounts.

Fortunately, many of the tools for the Macintosh and for Windows (particularly the World Wide Web browsers) look very much the same. The tools used with host-based Internet access, of course, look quite different. Throughout this book, Windows examples are used, and the screen displays are taken from a Windows-based computer. Macintosh users should be able to get the same information from these examples, even though their screens might look slightly different. This book also includes many text-only examples, particularly when the text-only tools are different in substance as well as in look.

The following sections discuss four tools used in the Internet. These core tools are Telnet, FTP, Gopher, and the World Wide Web. Later in the book, you learn about a couple of other tools: electronic mail and Usenet news.

Telnet—Terminal Access to the Internet

Telnet was one of the earliest technologies used to access the Internet. This technology reflects the way computers were used in the days before personal computers, when people used terminals to access central shared computers. A terminal generally looks like a computer: It has a screen and a keyboard but has no processor, which is the "brain" of a computer. All that a terminal can do is display data sent to it from another computer, allow the user to type data at the keyboard, and then send the data back to that computer. Terminals tend to be limited to text only (with some exceptions). The key distinguishing feature of a terminal, though, is that it doesn't actually do any processing—it just connects to a central shared computer that does the processing.

Nowadays, few people use terminals; they instead use personal computers of various kinds. However, many electronic resources on the Internet were developed for terminal access, and often you will need to access these resources as though you were using a terminal. To access these resources, you need a

program called Telnet, which makes your personal computer look like a terminal and connects it to a remote shared computer.

Using Telnet to Connect to Internet Resources

Many different versions and brands of Telnet software are available, enabling you to connect to remote shared computers to access terminal-based Internet resources. Some names that you might encounter are NCSA Telnet, PCTCP, CUTCP, WinQVT for IBM PC compatibles, and NCSA Telnet for the Macintosh. In addition, every UNIX system comes with a Telnet program simply called telnet. Almost everyone who has access to the Internet has access to a UNIX account, so the UNIX telnet program is used in the following example. Keep in mind that the basic principles are the same, no matter which Telnet program you use or what your Telnet program looks like.

Before you can connect to a remote computer using Telnet, you need to know the Internet address of that computer. Internet addresses can be in one of two forms: a series of letters or words connected by periods (such as **ezinfo.ucs.indiana.edu** or **velcome.iupui.edu**), or a series of four numbers connected by periods (such as **129.79.33.75**). Ultimately, all computers on the Internet can be addressed through a series of four numbers connected by periods, called an *IP number*. However, because people don't tend to work well with such numbers, computers on the Internet can be addressed with addresses of the first form as well. A system on the Internet known as Domain Name Service (DNS) translates between the names and numbers that computers go by. DNS can be thought of as a telephone directory for computers. These are all technical details, however. Simply put, to connect to a remote computer using Telnet, you need to know the address of the remote computer.

In the following example, you use the UNIX version of Telnet to connect to a remote computer at the address **infogate.ucs.indiana.edu**. This is the address of the library catalog at Indiana University. Before you begin, take a quick look at this Internet address. The last two elements, **indiana** and **edu**, tell you something about where the resource is located. **indiana** means that the resource is at Indiana University. (If **purdue** were in that location, the resource would be located at Purdue University.) Each institution, organization, and business with Internet resources has its own identifying suffix on the Internet. The **edu** part of the address identifies the computer as being located at an educational institution. Other types of institutions have identifying suffixes in their Internet addresses: **com** for commercial organizations, **org** for not-for-profit organizations, and **gov** for government agencies.

此为试读,需要完整PDF请访问: www.ertongbook.com