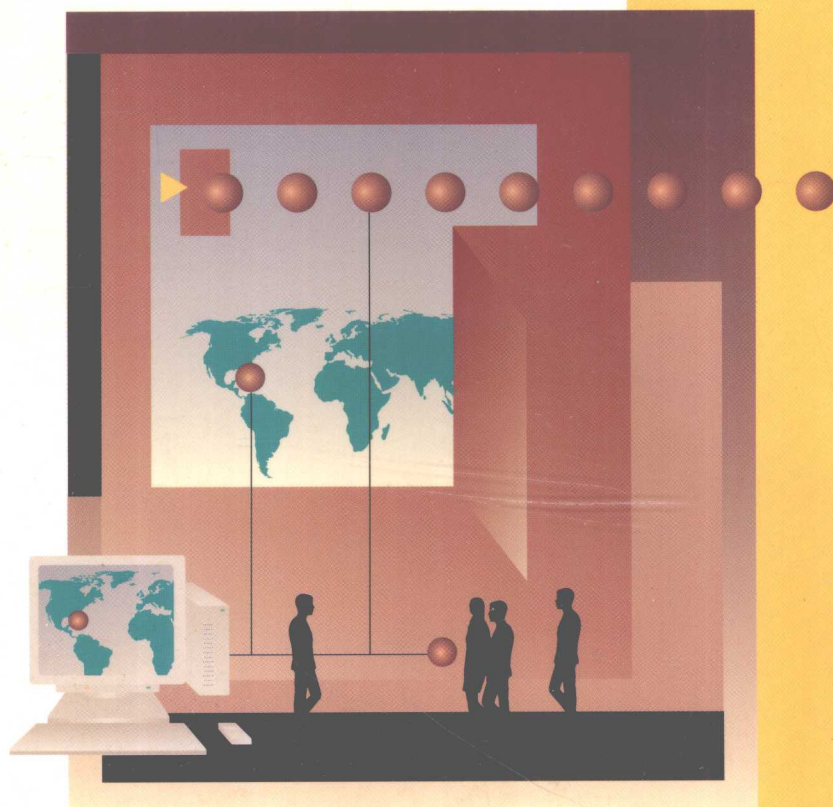


C H E R Y L H A R R I S



An Internet Education

ED

USED

A Guide to Doing Research on the Internet

An Internet Education

A Guide to Doing Research on the Internet

Cheryl Harris

California State University-Fullerton



An Imprint of Wadsworth Publishing Company

I(T)P® An International Thomson Publishing Company

Belmont • Albany • Bonn • Boston • Cincinnati • Detroit • London • Madrid • Melbourne
Mexico City • New York • Paris • San Francisco • Singapore • Tokyo • Toronto • Washington

Technology Editor: Kathy Shields
Assistant Editor: Tamara Huggins
Production: Ruth Cottrell
Print Buyer: Karen Hunt
Copy Editor: Ruth Cottrell
Interior Design: Image House
Cover Design: Image House
Compositor: Conch Composition
Printer: Malloy Lithographing, Inc.

COPYRIGHT © 1996

By Wadsworth Publishing Company

A Division of International Thomson Publishing Inc.

ITP The ITP logo is a registered trademark used under license.

Printed in the United States of America

1 2 3 4 5 6 7 8 9 10

For more information, contact Wadsworth Publishing Company.

Wadsworth Publishing Company
10 Davis Drive
Belmont, California 94002
USA

International Thomson Editores
Campos Eliseos 385, Piso 7
Col. Polanco
11560 México D. F. México

International Thomson Publishing Europe
Berkshire House 168-173
High Holborn
London, WC1V 7AA, England

International Thomson Publishing GmbH
Königswinterer Strasse 418
53227 Bonn, Germany

Thomas Nelson Australia
102 Dodds Street
South Melbourne 3205
Victoria, Australia

International Thomson Publishing Asia
221 Henderson Road
#05-10 Henderson Building
Singapore 0315

Nelson Canada
1120 Birchmount Road
Scarborough, Ontario
Canada M1K 5G4

International Thomson Publishing - Japan
Hirakawacho Kyowa Building, 3F
2-2-1 Hirakawacho
Chioda-ku, Tokyo 102, Japan

All rights reserved. No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without the written permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Harris, Cheryl, 1961—

An Internet education : a guide to doing research on the Internet

/ Cheryl Harris.

p. cm.

Includes bibliographical references and index.

ISBN 0-534-25851-4

1. Internet (Computer network) 2. Computer network resources.

3. Information services—User education. I. Title.

TK5105.875.I57H366 1995

001.4'0285'5713—dc20

95-42045

LIST OF FIGURES

- 1 Example of a Shell Account
- 2 The Trumpet News Reader and Usenet
- 3 Example of an IRC Discussion
- 4 FTP Session
- 5 Setting Up an FTP Connection
- 6 Searching with Archie
- 7 A Telnet Session
- 8 A Gopher Session
- 9 A Visit to a WWW Server
- 10 Windows 3.1 Control Panel
- 11 Selecting the Ports Icon
- 12 Settings for COM 2
- 13 The Trumpet Winsock TCP Manager (TCPMAN)
- 14 Selecting the setup.cmd
- 15 Locating the TCP Manager
- 16 Switching to the TCP Manager
- 17 TCP Manager Setup

P R E F A C E

PURPOSE

This book is designed to introduce the process of using the Internet for research with a special emphasis on acquiring the skills needed to use Internet tools efficiently and effectively. Since the Internet is essentially chaotic — without centralized organization or control — it can be frustrating to use without a guide. This book will show you how to find and use the Internet's vast resources on demand. In addition, within the discussion of each Internet tool there is a special section that illustrates how to use the entire Internet, even if your connection is an electronic mail only one.

PREREQUISITES

In most cases, Internet applications work identically in both the Windows and Macintosh operating environments. Where there are differences, however, the author has illustrated the Windows solution. The book presumes that the user is familiar with basic computer functions such as saving files, printing files, and using dialog boxes and pull-down menus.

PROJECTS

It is strongly recommended that you reinforce your understanding of all the Internet tools through practice. Ask your instructor for research questions, follow through on issues about which you are curious, explore and “surf” the Net freely (make sure you have a firm command of “netiquette” principles before you do, however!). An excellent way to get a feel for the overall Internet is to become a regular visitor to the “What’s New” sections that are part of the WWW’s many site directories.

GLOSSARY/INDEX

A glossary of research and Internet-related terms is included at the end of the book as well as an index.

COMMAND AND OTHER CONVENTIONS

Commands and site addresses are separated from the text by their placement on a new line and in a different typeface. They represent examples of what you must type to communicate with your software and with other systems. The *tool* to be used (i.e., Telnet, Gopher, WWW browser, etc.) is indicated by text above the command line.

TECHNOLOGY WATCH

The Internet is a moving target — what is there today may not be there tomorrow. Already too vast to be explored in a single lifetime, it still grows at a phenomenal pace. Resources are constantly moving, being reorganized, upgraded, or replaced. For that reason, resource and site addresses have been rarely and cautiously mentioned in this text. The serious Internet student is advised to seek out one of the regularly updated directories of resources (such as Hahn's annual *The Internet Yellow Pages*, McGraw Hill), widely available at bookstores, or to routinely turn to online subject directories such as Yahoo (<http://www.yahoo.com>). It is wise to be prepared for change in online environments, so stay flexible and be willing to troubleshoot if resources you need are not where you expect them.

TO THE INSTRUCTOR

An online, hypertext instructor's manual with suggested projects and a comprehensive links list is available to accompany this text.

ACKNOWLEDGMENTS

Special thanks go to the following for assistance in developing this book: Kathy Shields and Tamara Huggins at Wadsworth, Ruth Cottrell (whose editorial expertise and patience are legendary), my colleagues at California State University, and the Fondation d'Art de la Napoule, France, for reminding me that offline life still exists.

TABLE OF CONTENTS

List of Figures vii

Preface ix

PART

1

Introduction 1

What Is the Internet? 2

The History of the Internet 3

Current Status of the Internet 5

The Future of the Internet 6

Implications of the Internet for You 11

Network Ethics 12

Privacy 13

Purpose of This Book and How to Use It 14

PART

2

Conducting Research on the Internet 19

Getting and Using an Internet Connection 19

Acceptable Use Policies 21

The Logic of Internet Structure 21

Common UNIX Commands 22

Research Basics 23

Research Goals 24

Areas of the Internet for Research 25

Databases 25

Libraries 26

Usenet: Discussion Groups 26

E-Mail and Listservs 26

Remote Computing Resources 26

Weaknesses of the Internet for Research 27

Research Methods 29

The Importance of Choosing the Right Tool 29

Sources of Error in Research 29

Primary Research	33
Content Analysis and Discourse Analysis	33
Interviews	34
Surveys	35
Electronic Focus Groups	36
Archival Research Issues	36
The Ethics of Data Collection	378
Ethics and the Internet Culture	38
Spamming	41
The Cancelmoose and Cancelbots	42
Research Collaboration	44
Data Analysis: A Creative Process	45
Evaluating the Quality of Data	46
Examining Patterns	48
Fair Use, Copyright, and Attribution	49
Research Success Stories	52

PART**3****Using Resources 65**

Power Tools and Internet Research	65
Basics of Electronic Mail	66
Domains and Addresses	69
The E-Mail Signature	71
Search Strategies: Finger, Netfind, Knowbot, Whols, etc.	72
Listservs	74
Other Mailing Lists	76
Emoticons	77
Acronyms	77
Privacy	78
Anonymous Re-Mailers	79
Style and Etiquette Issues	80
Fax Services Available Through E-Mail	81
Mailing List Archives	81
Usenet	81
Power Uses	83
Basics	83
Frequently Asked Questions Files (FAQs)	85
Style and Etiquette	85
Search Strategies	85
Usenet Available by E-Mail	86
Internet Relay Chat (IRC)	86
Power Uses	87
Basics	87
Bots, Lurking, and Other IRC Phenomena	90
Style and Etiquette	91

File Transfer Protocol (FTP)	92
Power Uses	93
Basics	83
File Compression and Decompression Software	94
Launching FTP	95
Getting Things Done	96
Transferring Files	97
Netiquette	99
Estimating Connect Time	99
Mirrors	100
Search Strategies	100
Telnet	105
Basics	105
Gopher	107
Power Uses	107
Basics	108
Netiquette	109
Search Strategies: Veronica and Jughead	110
Gopher Available by E-Mail	110
The World Wide Web (WWW)	111
Power Uses	112
Basics	113
Search Strategies	116
Publishing Your Own Homepage	118
WWW Available by E-Mail	119
WWW Available by Fax	119
Applying the Tools	120
Online Libraries: What Libraries Have to Offer	120
Online Communications	124
Electronic Mail: How E-Mail and Discussion Groups Can Be Power	
Research Tools	124
More on Internet Culture	125
The Internet Treasure Hunt: Learning with the Net	126
RFCs: Getting More Technical Information	126
Entertainment and the Internet: Games and MUDs	127
Notes on Configuring TCP Manager and Windows 3.1 for a SLIP/PPP Connection	127
Initial Software Installation	128
Checking Windows' Configuration	128
Configuring the TCP Manager	130
Ending the Session/Disconnecting	132
Troubleshooting	133
Logging In	133
No IP Address	133
Two Important Points	134

PART The Ever-Mutating Net: Keeping Up with
4 Changes 137

Online Resources 138

InterNIC 138

WWW "What's New" Services 139

Yanoff List 141

Alex: A Catalog of Electronic Texts 141

Internet-on-a-Disk Newsletter 141

Scott's Hotlist (Subject Guide) 141

Internet Monthly Report 142

Final Considerations: How to Protect Yourself from Online Hazards 142

Viruses 143

Crackers and Unauthorized Use 143

Putting It All Together: Harnessing the Total Internet 144

PART Glossary 145

5

PART References 155

6

Index 163

Introduction

Someday a technology called **smart agents**¹ will patiently analyze a complete profile of our needs, our likes and dislikes, our habits, our tastes, and our exact requirements. Then we can spend time on the beach and the **agent** will do what we do not have the time or the inclination to do, such as “search every corner of the information universe.”² Until that time, however, each of us needs to become successful and efficient in a new “information economy” that redefines our roles as essentially information workers no matter what our job title or career aspirations. Promises of unlimited leisure time have been coincident with the introduction of almost every new technological development in the twentieth century (and for the most part have not materialized), and until the halcyon days of smart agents arrive, we have our work cut out for us.

The past decade has been characterized by an explosion of information: by some estimates medical and scientific journals have increased tenfold. Every morning the average office worker is faced with a barrage of faxes, electronic mail, voice mail, Federal Express and other overnight packages and, of course, the “regular” mail. To stay current in any one field requires the expenditure of hours probably well in excess of a normal 40 hour workweek. Barry Diller, media mogul and chairman of QVC, reported recently that every morning he receives “a fat packet of at least 80 articles” from many sources, and “that has been whittled down three times before I see it.”³

In short, we live in an information jungle. Of course we must learn to manage the jungle in order to survive, but beyond that, true expertise in information management brings with it enormous power, personal advancement, and satisfaction.

The purpose of this book is to introduce some of the powerful techniques for finding and analyzing information through the Internet. The Internet is only one of the important information resources today, and many of the techniques discussed here can also be applied to other venues.

What Is the Internet?

By late 1994 it seemed the word **Internet** was on everyone’s lips. Newspapers ran front page headlines about the **information superhighway**, and in fact nearly 20,000 stories were printed in 1993–1994 using the words *information highway* or *superhighway*.⁴ Giant corporations vied to cut deals that would place them in a favorable position on the infobahn. The whole concept of a new electronic culture heralded by the Internet represented the outer edge of technological promise, of rekindled hope and awe, the way space travel did in the 1960s. We are now a culture that creates and must process massive amounts of information. It is also increasingly clear that we are becoming a global culture. To survive in this environment, all of us must learn how to navigate within this new global information system—use it, manipulate it, communicate with and through it, and evolve with it. Still, in June 1994, a national Harris Poll revealed that *two-thirds* of Americans said they had not “seen, heard or read anything about the Information Superhighway.”⁵ Even worse, of the 34% who were aware of the topic, most claimed to have very little understanding of it.

There are good reasons for this confusion. The Internet is an entity with millions of users and millions of hosts. It has no true central

organizing body or even what might be called a true organizing principle. It reaches 137 countries, and every 10 minutes a new network connects to it. In 1994, traffic on one part of the Internet, the **World Wide Web (WWW)**, grew an astounding 341,634%. The breadth of the **network** is very difficult to grasp. Moreover, although the Internet offers a vast array of fascinating and useful resources, learning how to find and use the Internet's tools is difficult for most new users. There is a good reason for this because the Internet was not originally developed for the novice or even for average-skill-level computer users, but for highly skilled defense engineers and scientists who wanted a secure communications system in case of nuclear attack (**ARPAnet**) and who also needed to share remote computing resources with each other. As a result, until quite recently moving around the Internet required more than a passing knowledge of **UNIX** as well as other operating systems and command languages and quite a bit of patience.

So what is the Internet (or Net, as it is often affectionately called), really? It is by far the largest, most comprehensive collection of human knowledge and experience we have, encompassing the collections of many libraries and a myriad of other information sources. It is global in scope and exists outside many recognizable boundaries of time and place. Despite the fact that there is a large amount of unedited, unevaluated data along with the "jewels," it is an incredibly rich, diverse, and valuable tool.

Notice that I say it is a *tool*. The Net is a collection of resources with many diverting aspects—games, play areas, and all kinds of entertainment—but it is unmatched as a resource for real work and research. All Net users should be strongly warned that, if not carefully handled, the Internet is without doubt a strong promoter of procrastination. There is far too much to see and do, much of it without obvious purpose, and all of it available whenever you have a deadline to meet. Using this resource demands planning, skill, and judicious use. In addition, the Internet's amazing collection of data is more often than not unedited, and each resource must be evaluated in terms of its accuracy, validity, and reliability. This book is intended to equip new users with many of the techniques required to use the Net efficiently for research. But first, let's look at Net history and consider some of the significant future implications of the Internet.

The History of the Internet

Scientists began experimenting with linking computers to each other and to institutional sites via telephone hookups in the 1960s. This

research was funded by the U.S. Defense Department's **Advanced Research Projects Agency (ARPA)**. ARPA wanted to see if computers in different locations could be connected efficiently using a new technology known as **packet switching**, which would allow several users to share one communications line. Packet switching cuts up data into discrete units, each one identified by a code or label, which can be sent over high-speed telephone lines. Each packet was given the computer equivalent of a postal address so that it could be sent to the right destination, to be reassembled shortly before reaching its destination into a message the computer could use and, finally, a human could understand.

One of the purposes of this new system, beyond allowing researchers and scientists to talk to each other, was the creation of a communications network that the government hoped would be secure in case of nuclear or other emergency. Packet switching and other protocols were designed so that if one or more **nodes** on the network went down, messages could be flexibly routed around them. The eventual development of reliable **electronic mail (E-mail)** protocols meant that even lengthy messages and files could be exchanged faster and less expensively than they could by ordinary phone calls. E-mail messages can be transmitted faster than two people can say the same things over the telephone.

As ARPA became ARPAnet (based at UCLA from 1969), users developed a way to conduct online conferences through the network. These began as discussions focusing on sciences, but eventually their scope broadened to include hundreds, and now even thousands, of subject areas. Scientists, researchers, students, professors, and government workers were attracted to these conferences as a way of exchanging information quickly and efficiently.

In the 1970s, ARPAnet split into military and civilian networks, mostly universities or defense contractors.⁶ ARPA helped support the development of standards for transferring data between different types of computer networks, so that networks would reliably recognize each other and communicate. These Internet (from internetworking) protocols made it possible to develop the worldwide Internet we have today; without them, connecting to other computers would be a frustrating case of hit or miss. Within the decade, many other countries accepted the ARPAnet standards and joined the Internet. Effectively, the world was linked together via computers.

The 1980s were a time of exponential network growth. Most of this growth was in the educational sector, with thousands of colleges and universities worldwide joining the network, primarily because the National Science Foundation (overseeing the network) insisted that

access be expanded and offered to as many institutions as possible. Toward the end of the decade, however, commercial organizations and large corporations began to hook their own internal networks to the Internet for worldwide access and as a way of handling messages for traveling employees. In addition to communications, hundreds of thousands of databases, files, library catalogs, and other information resources were added to the network by its participants throughout the 1980s. Access by nonuniversity or military organizations to the Internet's electronic mail and messaging systems became more widespread by 1990.

Current Status of the Internet

The Internet is at a critical turning point in its history. Just two years after policies banning commercial use were tacitly dropped, the Network has become something of a moving target, with huge gains in new users and **hosts** (**servers** that connect to the network) every month. There were times in 1994 when new users doubled every month. Today's Net is a global forum of many networks with more than 25 million users—a loose collection of millions of computers at about 500,000 sites. More than 300 gigabytes of information—the equivalent of a half-million 250 page books—flow through the system daily.

About 33% to 40% of that growth is attributed to commercial, or “.com” **domains**, a jump of 628% in just three years according to the Internet Society. Commercial use is centered on a part of the Internet known as the World Wide Web (WWW), which links together many different types of resources for improved search and retrieval and features a graphic interface with multimedia capabilities. The World Wide Web was estimated in late 1994 to have grown in excess of 356,000% in one year.⁷ Nor is the growth curve over. Vinton Cerf, one of the original pioneers of the Net, believes that the Internet may soon offer more than 600 million networks (connected hosts) and be the model for the future “information superhighway.”⁸ By the time you read this, it is certain that the Internet will have expanded even further.

This kind of popularity does not come without a price. The frenzy to join the Internet party by businesses and hobbyists has been perceived by many long-term Net denizens as a serious threat to the cherished culture of the network, long founded on the belief that information must be freely shared, that respect for each other's privacy and opinions is paramount, and that the resources of the entire network should be used responsibly and conservatively for the greater good. In the absence of a central governing body or authority,

the network ran surprisingly smoothly due to the self-governance of the Internet community. For years the Internet was arguably the only “mass media” that successfully bypassed the gatekeeping functions of the commercial press and the publishing and broadcasting industries. The large numbers of new users strain resources to their limits, and advertising efforts through the network result in intense **flame wars** (battles played through electronic mail exchanges) that in some cases bring network traffic to a near standstill. The culture clash of the experienced Internet citizens, sharing common goals and ethics, and the many “newbies,” not yet “Internet socialized,” is still very much underway.

Who are these new users? Demographic studies of Internet users suffer from sampling limitations due to self-selection, and they are also unable to keep up with the pace of the network’s growth.⁹ However, most preliminary work profiling the network’s membership suggests that it is a young, affluent, and primarily male (90% by some accounts) world.¹⁰ Although gender on the Internet is surprisingly fluid and subjective, as users shift in and out of gender roles at will in “chat forums” such as the **Internet Relay Chat (IRC)** and other discussion groups, the network has been heavily criticized for being hostile to women. Net content has also been critiqued as reflecting the interests of its peculiar demographic makeup rather than the wider audience now being attracted to it. It is said that this bias is responsible for the somewhat spotty and skewed nature of available resources.¹¹

The Future of the Internet

We can speak of the Internet as being simultaneously “the new hope of the dispossessed . . . and the ultimate tool of Big Brother.”¹² A networked world could be the means to the creation of a true global consciousness, bringing down the boundaries to communication that governments try to erect, but the information superhighway also has the potential to change our daily lives in ways we may not find so appealing.

Within a few years, the Internet will be part of a technological matrix in which computers, telephony, television, and other communications media have converged.¹³ This highly interactive interface will provide opportunities, given enough bandwidth and computing power, to work and live in virtual spaces (environments created by **virtual reality** technology) significantly different from our experience of the world today. For some time businesses have been interested in the Internet and other new technologies as a means of dissolving or downsizing their

“real space” operations in favor of their Net presence—the advertising agencies Chiat Day and Winkler McManus, the Virtual Online University, and a number of law firms come to mind as examples. Some organizations do not exist in the traditional physical sense except as the Internet links their personnel and resources together. Think tanks predict that businesses and governments will be attracted to the virtual workplace model because it allows them to reduce spending on infrastructure (buildings, equipment, etc.) and also overcomes the traditional objections to home-based workers because completely networked employees are easily monitored. This workforce would also be flexible because many, if not most, would be essentially freelancers hired on a project, contract, or seasonal basis. Moreover, virtual workers have the advantage of easy collaboration with colleagues half a world away.

True virtual reality promises to provide a seamless impression of physical presence; for example, you could experience working in an office, even seeing your office mates when you turn your head, although they are actually in Bangkok or Mogadishu—not sitting next to you.¹⁴ On the negative side, dismantling unneeded physical infrastructure (home-based, networked workers do not need more highways, cars, office buildings, and so on) will cause the loss of many jobs.¹⁵ The popularity of interactive, online shopping services will further contribute to the erosion of local retail economies.¹⁶

Even college education can be conducted online through a handful of global universities that have no actual campuses other than online forums and other meeting places. Courses could be very specific and targeted to the student's information and skill needs, and they could be given in the time frame, format, and at the pace of the individual student. Experimental workshops and courses offered online have already generated an enormous response. One such course offered through the University of Pittsburgh drew 864 people from more than 20 countries in its first term. Later this workshop attracted 19,994 students representing 54 countries.¹⁷ Furthermore, “unlike traditional distance education systems which relied heavily on print-based materials supported by audiotape, telephone contact, videotape, color slides, study pictures, or kits containing samples, the Internet gives increased access to graphics, sound, and video files via software like **Mosaic**, as well as real-time communications. Innovative computer and telecommunication technologies expand and enhance traditional distance education by adding additional means of communication.”¹⁸ Interestingly, this educational delivery system changes the nature of student-teacher dynamics significantly: “The virtual classroom was found to be a positive yet different type of communications from the traditional classroom. This change in communication