

# The **INTERNET PUBLISHING HANDBOOK**

For World-Wide Web,  
Gopher, and WAIS

- Choose the methods, hardware, and software that best suit your needs
- Make sense of the confusing mess of protocols, platforms, and data types
- Get up to speed on the technical, administrative, financial, and legal issues involved in Internet publishing

**MIKE FRANKS**



Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Addison-Wesley was aware of a trademark claim, the designations have been printed in initial capital letters or all capital letters.

The author and publisher have taken care in preparation of this book, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

### **Library of Congress Cataloging-in-Publication Data**

Franks, Mike.

The Internet publishing handbook : for World-Wide Web, Gopher, and  
WAIS / Mike Franks.

p. cm.

Includes bibliographical reference (p.349) and index.

ISBN 0-201-48317-3 (pbk.)

1. Internet (Computer network) 2. Gopher servers. 3. World-Wide  
Web (Information retrieval system) 4. WAIS (Information retrieval  
system) I. Title.

TK5105.875. I57F769 1995

025.04—dc20

95-30895

CIP

Copyright © 1995 by Mike Franks

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America. Published simultaneously in Canada.

*Sponsoring Editor:* Claire Horne

*Project Editor:* Sarah Weaver

*Production Coordinator:* Erin Sweeney

*Cover design:* Robert Sweeney

*Text design:* Greg Johnson, Art Directions

Set in 11.5 point Minion by Greg Johnson, Art Directions

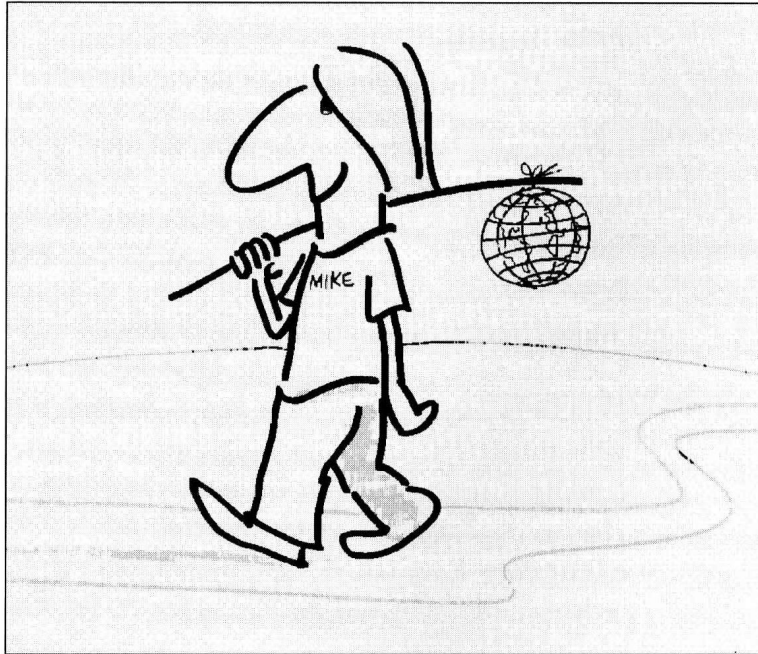
1 2 3 4 5 6 7 8 9 -MA- 9998979695

*First printing, October 1995*

Addison-Wesley books are available for bulk purchases by corporations, institutions, and other organizations. For more information please contact the Corporate, Government, and Special Sales Department at (800) 238-9682.

Find us on the World-Wide Web at: <http://www.aw.com/devpress/>

*To Riley, Nicholas, and Naomi. I'm a lucky guy.*



Courtesy of Mino  
Fortaleza, Ceará, BRASIL

# Acknowledgments

I'd like to thank Richard Chabran, Phil McAbee, and John Powell for their constant support and encouragement. In addition, I wouldn't have been able to write this book if it wasn't for the advice, assistance, and interesting environment created by my coworkers at UCLA's Social Sciences Computing: G.J. Chen, Mary Johnson, Carlos Tobar, Lily Sun, Eleanor Chan, Michael Hwa, Julie Chen, Wendy Wright, and our past and present directors, Barry Gerber, Professor William Mason, and Tom Phelan. I'd also like to thank Gary Bolles for indirectly getting me started on this project. And, of course, I'd like to thank my parents whose lifelong support and sense of humor made this all possible.

Although any errorz are mine, there would have been a great many more without the help of many people. I'd like to thank Dave Brooks, William Zhang, Michelle Lew, Andy Pho, Karie Masterson, Wayne Miller, George Bing, Gloria Werner, Don Worth, Pieter Lechner, Kent Wada, and Kelly Stack for their assistance and advice. I'd also like to thank all the patient souls who allowed me to interview them by phone, via e-mail, or at crowded conferences. Clifford Lynch contributed greatly to my understanding of Z39.50. Claire Horne, my editor, was always a pleasure to work with. And copy editor Polly Kummel added a great deal of professionalism to my writing and toned down my excesses. I'd also like to thank the production crew: Sarah Weaver, Erin Sweeney, John Webber, and Greg Johnson. Finally, I'd like to thank all those on the Internet who made it possible to learn so much online.

# Introduction

*In America, freedom of the press is largely reserved for those who own one.*

—A. J. Liebling

**T**his book is for those of you who want to publish on the Internet. I assume that you have browsed the Internet (at least a little) and are now ready to put information of your own on the Internet. This book will provide you with the information you need to set up shop, and it will enable you to reach millions of users who connect to the Internet looking for information, services, and products.

This book will take you through the process of Internet publishing from beginning to end, using examples and advice gathered from Internet publishers around the world. It will bring you up to speed on the issues, techniques, and problems that you will encounter when you join the growing legions of educational institutions, government organizations, manufacturers and retailers, service organizations, large corporations and small businesses, those in the arts and sciences, and others who are tapping in to the Internet to publish information and quickly reach an expanding, worldwide audience, thanks to tools such as Gopher, the World-Wide Web, and the Wide Area Information Server.

As of May 1995 there were more than 30,000 Internet publishing sites: more than 7,000 Gopher servers in at least 47 countries, at least 23,000 World-Wide Web (WWW) servers in 60 countries, and about 137 Wide Area Information Service (WAIS) hosts serving 420 databases (although this ignores the role that WAIS plays behind the scenes at many Gopher and WWW sites). The beauty of publishing on the Internet is that it is just as simple for the user or viewer (the person who reads what you publish) to get information from a site in Hong Kong as it is from the computer down the hall. The computer down the hall may respond faster, but the server in Hong Kong, Rio de Janeiro, or Finland may have special information that you want. And whether you're dealing with customers around the world or around the corner, you can use this system to provide your

clients with up-to-date information that can be accessed with only a few keystrokes.

## **Commercial Services versus the Internet**

---

Are America Online, CompuServe, and Prodigy the same as the Internet?

No, they are commercial online services (or bulletin boards) that charge their users for access by the hour. Large as they are, they are much, much smaller than the Internet. In many ways they are the opposite of the Internet:

### **INTERNET**

- An international collection of interconnected computers and networks
- Independent control of the equipment at each site (be it university, organization, or company)
- A tradition of academic and research activities (noncommercial), with commercial activities in the early stages
- User access to other computers and their services, largely for free, although some commercial databases are becoming available
- A strong tradition of free speech and cooperation between sites
- Access problems, although these are diminishing because of Gopher, WWW, and WAIS
- An audience of 13.5 million users (those who can browse Internet), a figure that has doubled every year for the last six years (according to an October 1994 estimate by Matrix Information and Directory Services of Austin, Texas, online at [<http://www.tic.com/mids/midshome.html>](http://www.tic.com/mids/midshome.html)).

### **COMMERCIAL ONLINE SERVICES**

- Centralized computers limited to the United States (except for CompuServe)
- Central control by each online service

- Fee structures, usually with per minute charges (in addition to the basic monthly fee) for access to popular features
- Competition for users
- Some censorship
- Relative ease of use
- 2.8 million CompuServe subscribers
- 2 million America Online members
- 1 million Prodigy members

The real story is that all three of these commercial online services have begun to connect their users to the Internet, first by allowing e-mail to cross over, next by providing Gopher access, and now by providing their users with access to the Web.

## **What Is Internet Publishing?**

---

Internet publishing is making material available for viewing on a computer connected to the Internet, so that others on the Internet can browse or search the information at will. It is not restricted to those who publish magazines, newspapers, or books. Information published on the Internet comes in many forms: plain text, formatted text, hypertext (links to other files), pictures, sounds, video, databases, interactive maps, and anything else you can store on a computer. The variety of content and purposes of Gopher, WWW, and WAIS servers demonstrates clearly the diversity of interests and philosophies on the Internet. Until now it was impossible for individuals and organizations to inexpensively and easily publish information all over the world. This new capability has even led some to see the vast expansion of Gopher and Web servers on the Internet as a rebirth of democracy. Although it has not reached the state envisioned by WWW creator Tim Berners-Lee, who believes publishing information should be as easy as browsing it, publishing has become much less expensive and more open to more people and therefore much more democratic.

Internet “publishers” range from the mighty to the small, from IBM to the florist down the street. Many individuals are leaving their mark on

the Internet with “home pages” that might give a brief personal description of themselves and even include their children’s baby pictures.

The wide variety of methods of publishing on the Internet include FTP servers, **finger daemons** (or programs), Hyper-G, Usenet News, e-mail mailing lists (list servers) and a variety of telephone and e-mail address directory services (see the glossary for definitions of these terms). But Gopher, WWW, and WAIS are our models, because they provide the easiest methods for getting your information on the Net. They are also frequently used in combination with each other. IBM, for example, has both a Gopher and a WWW server to display product information. And a WAIS server often provides the searching mechanism, hidden behind Gopher and WWW.

## What You Should Get from This Book

---

I want you to come away from this book with three things:

- Excitement about the possibilities and prospects of publishing on the Internet
- Familiarity with the issues and choices involved
- Knowledge of where to go on the Internet to continue your education

## Connecting to the Internet

---

This book assumes that you have already browsed the Internet to a certain extent. But if you haven’t, here are some ways to start. In Chapter 2 I’ll talk more about the kind of Internet access you’ll need for running Gopher, WWW, and WAIS servers and Internet publishing.

Your Internet access options will vary considerably depending on whether you are affiliated with a college or university, or whether your company or organization pays for Internet access or you have to pay for it entirely on your own. Generally, universities and colleges offer greatly subsidized Internet access to their students and faculty, because the Internet has long been primarily a research resource.



There are two kinds of home access, which can be simplified as text-only and graphical. Text-only can be accomplished with old PCs and 2,400-baud (slow) modems and a dial-up connection to an Internet service provider (ISP). Graphical Internet access usually requires a Macintosh or Windows-capable PC, a 14,400-baud modem, and either a SLIP or PPP connection to an Internet service provider. This usually costs more but can still be found in metropolitan areas for \$20 to \$30 per month. Where possible you'll want graphical access, but don't despair if that's out of your reach for now. Gopher, WWW, and WAIS are almost completely accessible via text-only browsers.

## **Finding an Internet Service Provider**

To figure out what your options are for gaining access to the Internet:

- Check with any educational institution with which you are affiliated. This includes elementary and high schools. In the United States more and more schools are starting to offer at least limited Internet access, sometimes through special machines in libraries or computer labs. Ask if it is possible to dial in from home.
- Check with your employer. Be sure you are aware of any restrictions the business may place on that access. Check for dial-in access as well.
- Ask friends and acquaintances if they know anyone who has Internet access and get information from them.
- Check for the existence of a FreeNet in your area. This is the movement to offer free Internet access by geographical area.
- Look in the Yellow Pages for Internet service providers and be sure to call all of them because rates and services vary widely. Where possible you want to avoid paying by the hour.
- Read the ads in local computer magazines (the kind that are given away at computer stores) for local and national ISPs.
- Check for local computer bulletin boards that also offer Internet access. This is a growing field.
- Check out such products as Internet-in-a-Box, OS/2, and others that offer the software to automatically connect you to national ISPs.

- Check with commercial online services to see what Internet access they offer. In this case you're paying for much more than just the Internet access, but you may find you enjoy those aspects as well. Prodigy, America Online, CompuServe, and others all offer some form of Internet access and plan to expand this service.

Remember that these methods are designed to get you inexpensive Internet access for browsing the Internet. That is a much less expensive option than the 24-hour dedicated connection you will require when you start publishing on the Internet.

## How Best to Read This Book

---

All the fun stuff is in Chapter 10. That is where you'll find example after example of different kinds of publishing on the Internet. Look there for ideas, but remember it is only a small sampling of what is on the Internet. Whenever you get bored or confused, go to Chapter 10 or browse the Internet some more to renew your enthusiasm.

Second, this book is written for several types of readers. If you find material that you think is too technical, skip it and move ahead. Look for a key sentence or two in those sections so you can see what the important point is. Later, after you have read and browsed the Internet some more and have begun to actually set up your server, review these sections because they are likely to make more sense to you. If they don't, write and tell me what you didn't understand so I can make the next edition better.

Third, *follow the links*. This book is not a step-by-step guide. In almost every section I refer you to sites on the Internet, known as links or URLs (Uniform Resource Locators), for more detailed information. The URLs are scattered throughout this book like footnotes out to the Internet. You can follow these links and see what they have to say if you have a program for browsing the World-Wide Web (such as Mosaic, Netscape, or Lynx). They appear in italics at the ends of sentences and paragraphs, and look like this: *<<http://sunsite.unc.edu/usenet-b/home.html>>*. Things are changing so fast that you should get in the habit of checking the Internet regularly.

Pay particular attention to the resource tables in most chapters. They include books, magazines, Internet sites, mailing lists, and Usenet newsgroups (described in Chapter 2) that are particularly useful in keeping up to date. Prices are given in U.S. dollars.

## What If a Link Doesn't Work?

---

Although I've done the best I can to make sure that all the links are correct and in working order, the Internet is constantly changing, so you may find that some links don't work. If that happens, try again later to make sure that it wasn't just busy when you tried it. Then try shortening the link, removing words one by one from the right end to get somewhere on the site. Once you've made some connection, look around for links that will get you what you want. Finally, if it's been changed completely, do a search on that subject through one of the Internet indexers listed in Table 2-1. You will almost always find related information that way, even if you don't find the original site, which may have gone out of existence.

## Quick Vocabulary Lesson

---

Several words and concepts with which you may already be familiar are basic to understanding the Web and the Internet, so let's review them briefly:

**Client/servers** are tools, such as **Gopher**, **WWW**, and **WAIS** (pronounced *ways*), for accessing the Internet through one computer. A **server** offers information or a connection to other computers, called **clients** (or browsers). The program that you run on your computer to do the searching and browsing is the client program; the computers with the information (Gopher, WWW, and WAIS) are the servers. The client and server can also be separate programs running on the same machine.

**Browsers** are the client half of the client/server relationship on the Internet. Browsers are programs that people run to search out information on the Internet. Browsers enable you to explore the Internet,

because they know how to “talk” to computers that act as information servers on the Internet. Browsers should not be confused with the servers to which they connect you. For example, NCSA Mosaic is a well-known WWW browser. Other WWW browsers include Lynx, Arena, Netscape, and many more. Gopher browsers include TurboGopher, WSGopher, HGopher, and Minuet. Browsers are only half of the Internet equation.

**Servers** are the other half of the Internet equation. They provide the information that client programs read. Servers come in different types: Gopher, WWW, WAIS, and many others. Basically, servers wait for clients to request information. Then they send the information if they have it.

## Organization of the Book

---

Chapter 1, Overview of Internet Publishing, WWW, Gopher, and WAIS, takes you through the last few years on the Internet and shows how Gopher, WWW, and WAIS were developed (and what problems they solved).

Chapter 2, Groundwork, gets down to business; it deals with all the steps that you must take, no matter which system or systems you decide to use. It will make you start to think about the issues and decisions you will have to face, as well as the planning, administration, and resources that publishing on the Internet requires.

Chapter 3, Gopher and Gopher+, describes exactly what these servers allow you to do. Among the advantages detailed in this chapter is the relative simplicity of using Gopher.

Chapter 4, World-Wide Web (more properly known as the HyperText Transfer Protocol or HTTP), describes the servers that are among the fastest-growing on the Internet. This chapter describes different versions and their relative advantages. The chapter also reviews the steps involved in setting up a server, learning to write and design HyperText Markup Language (HTML) documents (the basic text of WWW servers), and announcing, advertising, and

indexing your server. The chapter describes text conversion utilities and HTML editors, as well as approaches to hypertext layout and maintenance.

Chapter 5, WAIS and Indexing Alternatives, describes the extremely powerful searching capabilities behind many Gopher and WWW servers. This chapter describes various implementations (for example, FreeWAIS and WAIS Inc.) and techniques for data filtering (identifying the parts of your data files so that WAIS can index them correctly).

Chapter 6, Other Tools of the Trade, describes the use of Finger, CSOs, White Pages, Usenet News, MOOs, and Hyper-G for publishing on the Internet. These tools, often used in conjunction with Gopher, WWW, and WAIS, can provide exactly the right mix of service and flexibility.

Chapter 7, Internet Commerce, explores the developing security and charging techniques used today or in development, and Chapter 8, Hiring Out the Work, offers options if you don't have the appropriate Internet connection or computer equipment or just don't want to manage a Gopher, WWW, or WAIS information server yourself.

Chapter 9, Copyright, addresses some different aspects of publishing in this new medium including copyright and legal issues.

Chapter 10, Putting It All Together, offers case histories and interviews that show some of the creative ways people are putting information on the Internet.

The final chapter, Chapter 11, The Future of Internet Publishing, looks to the future of Gopher, WWW, and WAIS and discusses the likely resolutions of such Internet issues as network overloading, electronic money and digital signatures, document authentication, and implications for political and social issues.

The appendixes include a bibliography and a glossary, which you should check out sooner rather than later. Links to every example site or resource mentioned in the book can be found on the Internet in the Addison-Wesley Web Server, <<http://www.aw.com/devpress/>>.



## Notes on Writing This Book

---

As you may already know from browsing the Internet, it is a constantly changing landscape that is growing at a breakneck pace. For me the excitement has been in finding that everything has changed since the last time I checked. Happily, it's now possible to surf the Internet while I'm composing this literary masterpiece. Unhappily, it is far too easy to get distracted by new sites or by new information at old ones. So although I've done my best to capture the essence of what Internet publishing is all about, I'm afraid that some material may seem dated because the details change so quickly. I've tried to give you information and perspective that will be useful no matter what software or hardware you use.

Wherever possible throughout the book I included the names of the individuals who have written programs, run sites, or have contributed to the Internet in some way. I did this for two reasons. First, because the Internet is a community of sorts, it is helpful to know the people who make up the community. You'll see their names in e-mail and on newsgroups as you wander about the Internet. It is truly possible to make your reputation on the Internet in this way (for good or ill). Second, I wanted to include names as a way to thank those who've contributed so much to the Internet. Often their contribution is entirely voluntary, and it is this spirit—more than cables or wiring—that truly makes the Internet the fascinating entity that it is.

The Internet is an unusually chaotic, active area with hundreds, if not thousands, of communities springing up around an idea, program, or controversy. Like any community, it has its rules or accepted ways of doing things, but unlike many communities the Internet is extremely open to change. New information servers are greeted with delight and a great deal of interest, and you may find unexpected rewards once you start publishing on the Internet. I'd like to take this opportunity to encourage you to be as creative as you can and not to hesitate to put your, or your organization's, personality into your work. Above all, the Internet is about the exchange of ideas, so don't hide yours.

# Contents

<b>INTRODUCTION</b>	<b>ix</b>
Commercial Services versus the Internet	x
What Is Internet Publishing?	xi
What You Should Get from This Book	xii
Connecting to the Internet	xii
How Best to Read This Book	xiv
What If a Link Doesn't Work?	xv
Quick Vocabulary Lesson	xv
Organization of the Book	xvi
Notes on Writing This Book	xviii
 <b>CHAPTER 1 Overview of Internet Publishing, WWW, Gopher, and WAIS</b>	 <b>1</b>
A Quick History of Internet Publishing Techniques	5
Summary	17
 <b>CHAPTER 2 Groundwork</b>	 <b>19</b>
Joining the Internet Community	19
Identifying Your Competition	38
Defining Your Goals	39
Identifying Your Audience	40
Identifying Your Data	42
Identifying Your Needs	45
Identifying Available Resources	48
Determining How to Make Up the Difference	49
Identifying Charging Mechanisms, If Any	49
Identifying Security Risks	50
Planning the Presentation of Your Data	52
Preparing Your Data	55
Indexing Your Data	56
Registering or Announcing Your Server	58
Maintaining Reliability	59
Evaluating Server Performance and Audience Response	60
Using What You've Learned	63
Summary	63
 <b>CHAPTER 3 Gopher and Gopher+</b>	 <b>65</b>
Gopher: Pros and Cons	66
How Gopher Works	67
Gopher and Gopher+ Features	68
Gopher Resources	70
Types of Items Gopher Handles	73
Gopher Servers	77
Installing Gopher Server Software	82
Setting Up the Gopher Data Directory	89

Useful Techniques for Adding Information and Files	93
Gopher Tools	97
Gopher Menu Layout	98
Password Protection	100
What Logs Can Tell You	100
Setting Up Forms Using ASK Blocks	102
Security Holes and How to Prevent Them	103
Running Scripts or Programs from a Gopher Menu	104
Registering Your Gopher Server	105
Indexing	106
Summary	109
<b>CHAPTER 4 World-Wide Web</b>	<b>113</b>
WWW: Pros and Cons	113
How WWW Works	115
WWW Features	117
WWW Resources	118
Web Data Types	118
HTML—HyperText Markup Language	127
HTML Editors, Filters, and Converters	133
Web (HTTP) Servers	135
UNIX	137
VMS	140
Macintosh	141
Windows NT	143
DOS and Windows	145
Other Platforms	146
Deciding Who Runs Your Web Server	147
Designing Your Web Pages	151
Adding Information and Files	157
Web Tools	159
Advanced Web Techniques	159
Images	159
CGI Scripts	165
Forms	167
Web Indexing	172
External Indexers (Web Crawlers, Spiders, and Worms)	175
Advertising and Registering Your Web Server	177
Summary	178
<b>CHAPTER 5 WAIS and Indexing Alternatives</b>	<b>181</b>
WAIS: Pros and Cons	182
How WAIS Works	183
When WAIS is Useful	183
WAIS and Z39.50	184
Searching Capabilities and Limitations	185

Fielded Searching	188
WAIS Servers	188
Indexing Alternatives to WAIS	191
WAIS Resources	193
Estimating Computer Requirements	193
Installing a WAIS Server	193
WAIS Fundamentals	195
WAIS File Types	196
Indexing with WAIS	203
WAIS Tools	204
Advertising Your WAIS Sources	205
Summary	206
<b>CHAPTER 6 Other Tools of the Trade</b>	<b>207</b>
Perl	207
Pyphon	208
Glimpse	208
Harvest	209
FTP Daemons	209
Finger	210
CSO	210
List Servers	211
MUDs and MOOs	212
Hyper-G	213
HyperNews	214
Internet Relay Chat (IRC)	214
WebChat	215
Summary	215
<b>CHAPTER 7 Internet Commerce</b>	<b>217</b>
Advertisers: You're Not in Kansas Anymore	217
Demographics	218
Opportunities	219
Is Anyone Making Money on the Net?	220
Economic Models	220
Charging Issues and Concepts	223
Survey of Charging Techniques	226
Offline Payment	242
Summary	242
<b>CHAPTER 8 Hiring Out the Work</b>	<b>245</b>
Reasons to Hire Out	245
What You Pay For	250
Services and Consultants	253
Summary	264