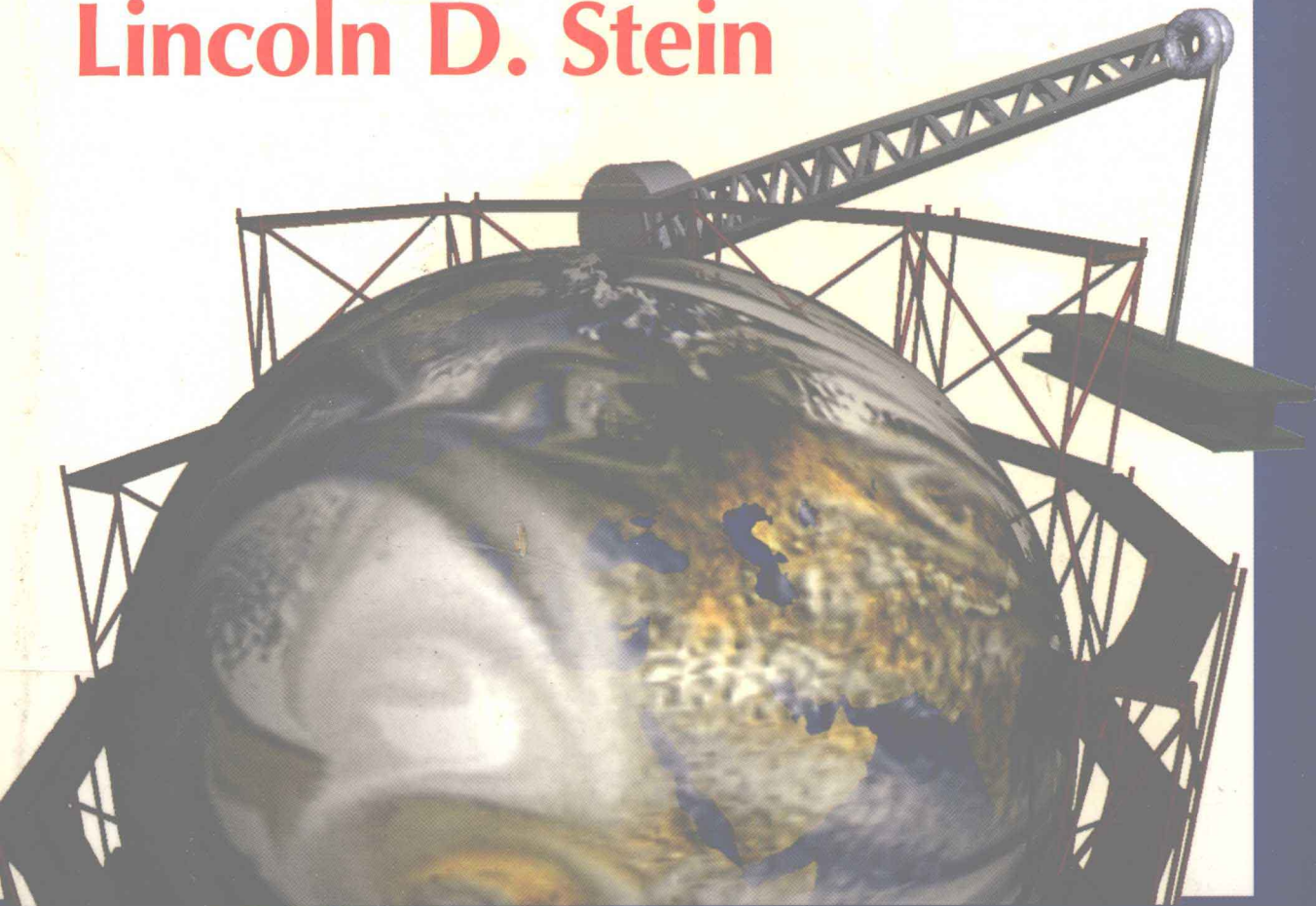


SECOND EDITION

How to Set Up and Maintain a Web Site

Lincoln D. Stein



CD-ROM with ready-to-use web page samples,
hot-linked throughout and web-ready

HOW TO SET UP AND MAINTAIN A

WEB SITE

Lincoln D. Stein



ADDISON-WESLEY

An imprint of Addison Wesley Longman, Inc.

Reading, Massachusetts • Harlow, England • Menlo Park, California
Berkeley, California • Don Mills, Ontario • Sydney
Bonn • Amsterdam • Tokyo • Mexico City

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 Longman, Inc. was aware of a trademark claim, the designations have been printed with initial capital letters.

The publisher offers discounts on this book when ordered in quantity for special sales.

The authors and publishers 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.

For more information, please contact:

Corporate & Professional Publishing Group
Addison Wesley Longman, Inc.
One Jacob Way
Reading, Massachusetts 01867

Library of Congress Cataloging-in-Publication Data

Stein, Lincoln D., 1960-

How to set up and maintain a Web site/Lincoln D. Stein.

-Second Edition. p. cm.

Includes index.

ISBN 0-201-63462-7 (pbk. : alk. paper)

1. World Wide Web (Information retrieval system) 2. Web sites-Design. 3. World Wide Web servers. I. Title.

TK5105.888.S74 1997

005.2'76--dc21

96-46116

CIP

Copyright © 1997 by Addison Wesley Longman, Inc.

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.

Text printed on recycled and acid-free paper.

ISBN 0-201-63462-7

6 7 8 9 1011 MA 02 01 00 99

6th Printing April 1999

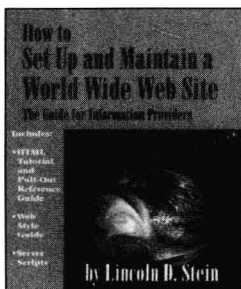
HOW TO SET UP AND MAINTAIN A

WEB SITE

P R A I S E F O R

How to Set Up and Maintain a World Wide Web Site

First Edition



"Lincoln Stein's *How to Set Up and Maintain a World Wide Web Site* is a wonderful book. If you are even thinking about becoming a webmaster, run out and get this book."

— Richard Morin
Unix Review

"This is an excellent book and thorough primer for anyone pondering the mysteries of Web publishing. After reading this book, the mysteries are revealed and the reader is well on his way to setting up shop on the Web."

— Chris Shipley, Editor
Computer Life

"When the Internet hit the mainstream, the flood of 'how to' books left new and old Internet users baffled by hundreds of mostly redundant titles. The same torrent promises to follow the popularity of the World Wide Web, but Lincoln Stein has done all of us an immense service by writing the definitive guide to web weaving before the deluge — a clear, complete, and hype-free book which anyone who wants to set up a web site (or just understand how a web site works) should read."

— Lyman Chapin, Chief Network Architect
BBN Systems and Technologies

"This book is useful for anyone with any level of interest in the WWW phenomenon. Whether you are an individual interested in authoring your own home page, an administrator interested in setting up a secure WWW server, or even just a bystander interested in 'how it works,' this is the book for you."

— Mukesh Kacker, Internet Engineering
Sun Microsystems Inc. (SunSoft)

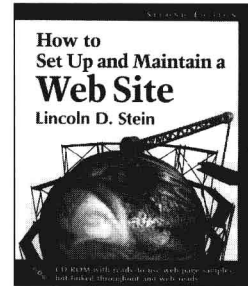
P R A I S E F O R

How to Set Up and Maintain a Web Site

Second Edition

This second edition includes coverage of:

- HTML 3.2
- Java™
- JavaScript
- CGI and Perl 5
- VRML
- UNIX®, Windows NT™, and Macintosh®
- Web Security FAQ



"Kudos for covering file uploads. This topic is often overlooked in books. I had never had occasion to examine the subject before, and had no difficulty understanding it after reading this book's discussion of it."

— Tom Boutell, Author of *CGI Programming in C and Perl*

"I think that this book would be a good starter kit for anyone interested in JavaScript. It would be good for everyone who wants a good kick start in JavaScript programming."

— Sandeep K. Gopisetty, Advisory Engineer
IBM Almaden Research Center

"For a quick tutorial, the content presented in the chapter on Creating Hypertext Documents would make a good start for a person who wanted to learn HTML. If the rest of the book is as good as this chapter, I would have no hesitation in recommending this book for staff use here at Penn State."

— D. Jay Newman, Research Programmer/Instructional Developer
Penn State

"The chapter on VRML is comprehensive. Since the title refers to tools, the author delivers exactly that. There's a lot of good reference material in this chapter."

— Mike Macedonia, Vice President
Fraunhofer Center for Research in Computer Graphics

Preface

This is a guide for anyone who is planning to set up a World Wide Web server site, or who wants to enhance an existing one. It is intended to embrace a variety of needs: those of the corporate marketing department exec who needs to get the fall catalog on-line fast; of the system administrator nervous about system security; of the scientist who wants to make a database of experimental results available to her colleagues; or of the college student eager to share his insights on the city's best ice-cream parlors.

Why purchase a book on WWW administration when all the information is already out there, freely available, in glorious hypermedia form? In part this book grew out of my frustration with the hypertext style of documentation. The information is indeed out there, but scattered about the globe, often incomplete, sometimes contradictory, ever changing, and frequently hard to locate again at a later date. This book pulls together all the relevant information garnered from an individual's struggles in setting up and maintaining a Web site.

Part of the beauty of the Web system is that a rudimentary site can be set up in an afternoon and allowed to grow and bear fruit for a long time thereafter. This guide is intended to be useful during all the phases of a Web site's life span, from the first invocation of the server's *install* script to the last baroque frill on a Web gateway that has grown so complex that not even its creator can figure out how it works. You probably won't need to read the whole book to accomplish what you want to do, but it is a comfort to know that it's all there when you need it. The book starts with the nitty gritty of choosing and obtaining Web server software, installing it at your site, and configuring it to behave itself. Next there are chapters on how to get your information into Web-compatible form: how to write hypertext documents; what tools are available to convert existing text files into hypertext; and how to negotiate the alphabet soup of graphics, sound, and video standards. Security is a growing issue everywhere on the Internet, and this book devotes a chapter to that issue: both the problem of keeping the Web site secure, and the task of dealing with network security measures that prevent Web software from working the way it's supposed to. Chapters on cgi

scripts, Java, and JavaScript describe how to give your site searchable indexes, fill-out forms, clickable maps, animations, and gateways to other services. Finally, there's a Web style guide that tries to balance Web page aesthetics with practical considerations such as performance. (A breathtakingly beautiful Web page is not much good if no one has the patience to wait for it to download.)

What this book is *not* is a manual for World Wide Web browsers or a listing of neat places to visit on the Web. Nor is it a guide to running all possible servers on all possible operating systems (there are more than 60 servers and counting!) Instead I've chosen a single popular server from each of the Unix, Macintosh, and Windows operating systems. There's enough similarity among the various servers that once you understand how one works, you pretty much understand them all.

I hope that you enjoy opening up a Web site as much as I have, and I look forward to seeing you on the net.

What's New Since the First Edition

A lot has changed in the year or so since the first edition of this book was written. The Web has increased in size more than 20-fold, and businesses have jumped into this exploding market with a bewildering offering of browsers, servers, HTML editors, and site management tools. Large parts of this book have been completely rewritten to keep up with the changing times, and new chapters have been added. Here are the highlights of what's new:

- Detailed instructions for setting up Windows and Macintosh servers
- Greatly expanded coverage of secure servers, particularly SSL servers
- Instructions on setting up virtual hosts
- HTML 3.2
- VRML, the Virtual Reality Modeling Language
- A completely new JavaScript chapter
- A completely new Java chapter
- Rewritten and expanded examples in the CGI chapters
- Coverage of such new HTTP features as cookies and virtual hosts
- Coverage of new HTML editors

Because the CERN server is no longer a supported product, I've removed it from this edition.

About This Book

Typographical Conventions

The code examples given in this book, including the contents of configuration files, paths, executable scripts and the source code for HTML, are in monospaced font. A **bold monospaced font** is used to indicate user input, as in:

```
zorro % date
Sun Aug 11 11:06:38 EDT 1996
zorro %
```

An *italic font* is used for URLs, the names of system commands, and for lowercase program names.

URLs

URLs (the ubiquitous “Uniform Resource Locators” that uniquely identify each document on the Web) are used everywhere in this book. Unfortunately print is a static medium and URLs change constantly. Some of the URLs in this book will have changed between the time it went to press and the time it appeared on bookstore shelves. It is hoped that the Webmasters responsible for these changed URLs left forwarding addresses telling you where the new versions can be found. If not, I can only apologize and suggest that you try to track down the new location using one of the Web’s many subject guides or keyword search services. The Web resource guide at www.genome.wi.mit.edu will also contain updated addresses.

Example HTML Documents and Scripts

You’ll find the source code for all the example HTML documents and executable scripts given in this book at

<http://www.genome.wi.mit.edu/WWW/>

Follow the links to *examples*. Here you’ll find pointers to the examples from each chapter. All of the example code is in the public domain. You are welcome to use all or part of a piece of code as a template for your own projects. At this location you will also find working versions of the executable scripts in Chapter 9, as well as errata and bug fixes.

For your convenience, I’ve also placed a copy of all the code examples on the companion CD-ROM.

Tools and Other Resources

The book refers to huge numbers of Web resources, including icons, tools, executable scripts, code libraries, and sundry utilities. Typically, each resource has a home site where its most recent version can be found. I’ve gathered up the most useful tools and placed copies of them in a subdirectory of <http://www.genome.wi.mit.edu/WWW/>. Follow the links to the resource guide.

Also check the CD-ROM, where many of the resources can be found. Some resources cannot be redistributed because of licensing agreements, but I’ve put copies of all the others into the subdirectory *tools*. If you see a

noncommercial or shareware tool mentioned in this book, chances are good that a copy of it is on the CD-ROM.

Since tools get updated frequently, you should also check a resource's home site to obtain the newest version.

Freeware, Shareware, and Other Beasties

Lots of software is available via the Internet, and although much of it is "freely available," not all of it is free. Truly free is software that has been explicitly placed in the public domain by its authors. This software can be used for any purpose whatsoever, including modifying and redistributing it. Several of the Web servers described in this book fall into this category. In contrast is a broad class of software loosely called "freeware." This is software whose authors have not given up copyright, but who allow you to use the software without payment. This software may have various restrictions placed on it, such as noncommercial use only or limitations on your ability to bundle it with other software products. Then there is "shareware," whose authors allow you to use the software for a trial period, after which you're honor-bound to discard the software or to pay a licensing fee. Finally, there's commercial demo software, which is usually a crippled version of the real thing.

Whenever I mention a piece of software, I try to report whether it is public domain, freeware, or shareware. Sometimes, however, I haven't been able to determine what the status of a utility is, or its status has changed. Before using any tool, you should make sure that you understand its author's intent.

Organization

Chapters 1 and 2 introduce the Web and explain how it works. You'll want to read Chapter 1 and the introductory sections of Chapter 2 whether you're more interested in administering Web server software, authoring hypertext documents, or developing executable scripts that create dynamic documents. Script developers will probably want to read through the esoterica at the end of Chapter 2 as well, because many clever tricks are possible when you understand the protocol in detail.

Chapters 3 and 4 are of most interest to the Web server administrator. They explain how to set up the server software, configure it, and make it secure.

Chapters 5 through 7 will be of most interest to the Web author. Together they explain how to write hypertext documents; provide pointers to tools for interconverting text, graphics, and animation files; and provide a style guide for making documents both effective and attractive.

Chapters 8 through 11 are for Web script developers and authors who are interested in learning to write executable scripts or to incorporate Java and JavaScript applications into their pages. These chapters also contain pointers to scripts and applets that you can pop into your site without any programming.

Acknowledgments

A surprising number of people have helped, directly or indirectly, with this book. I'm extremely grateful to the members of my lab at the Whitehead Institute. Robert Dredge, Robert Nahf, Richard Resnick, Steve Rozen, and Peter Young all offered invaluable assistance in installing, evaluating, and debugging Web software tools. Nadeem Vaidya worked nonstop to get the contents of the CD-ROM organized in time. Lois Bennett patiently kept the network running despite wave after wave of experimentation with increasingly esoteric aspects of Web administration. Andre Marquis deserves special thanks for introducing me to the Web and getting the lab's server up and running in the first place. Thanks as well to Drucilla Roberts and Cassia Herman, who provided the livestock photos.

I'd like to thank my reviewers, Ken Arnold, Thomas Boutell, Don Brutzman, Dan Connolly, Vansanthan S. Dasan, Mark Ellis, Doug Felteau, Lisa Friendly, Sandeep Gopisetty, Arlen Hall, Mukesh Kacker, Doug Kramer, Jerry Latimer, Mike Macedonia, Nick Manousos, Michael Moncur, Jay Newman, Scott Redmon, Kenn Scribner, Win Treese, Andrew Wooldridge, and Tony Zawilski for their insightful suggestions and for the many bloopers they collectively identified and nipped in the bud for this second edition.

Also, I'm grateful to the people who reviewed the first edition of this book, Steven Bellovin, A. Lyman Chapin, Robert Fleischman, Barry Margolin, Craig Partridge, and Clifford Skolnick.

My particular thanks to my editor, Carol Long, and her assistant, Mary Harrington, for their unflagging energy and encouragement throughout this project.

Lincoln D. Stein

lstein@genome.wi.mit.edu

<http://www.genome.wi.mit.edu/~lstein>

Contents

Preface	xvii
---------------	------

Chapter 1

Introduction to the Web	1
A Little History	1
Key Web Concepts	3
The Web and the Internet • Web Browsers and Servers • Support for Multiple Protocols • A Uniform Naming Scheme • Pages and Compound Pages • Hypertext • High-Level Page Description Language • Multimedia • Extensibility	
What Can You Do with the Web?	9
Distributing Scientific Data • Commerce • Education Courseware and Technical Manuals • Hypertext Books • Public Service Information • Guides to the Internet	
Steps to Creating a Web Site	13
Internet or Intranet? • Your Own Server or Your Service Provider's? • Finding an Internet Service Provider • Establish a Connection to the Internet • Choose a Good Domain Name • Choose the Server Hardware and OS • Select and Install Server Software • Write Your Site's Web Pages • Enhancing Your Site with Executable Scripts and Applets • Allocate Responsibility • Publicize Your Site • Keep the Site Up to Date	
Doing Business over the Web	24
Charging a Subscription Fee for Access to Your Site • Billing for Usage by Monitoring Access Logs • Accepting Payment over the Web	

Chapter 2

Unraveling the Web: How It All Works	27
Network Basics	27
The TCP/IP Protocol • IP Addresses • Domain Names • Clients and Servers • Ports • Daemons and <i>Inetd</i>	

Uniform Resource Locators	32
Complete Versus Partial URLs • Specific URLs	
The MIME Typing System	42
Introduction to MIME • MIME Types and Helper Applications • MIME Types and File Extensions	
Other Web Server Features	46
Virtual Hosts • Proxy Servers	
The HTTP Protocol	47
An Interview with the Daemon • The Phases of the HTTP Protocol	
BOX: HTTP/1.1 and HTTP-ng	56

Chapter 3

<i>Installing and Configuring a Web Server</i>	59
Choosing Web Server Software	59
Supported Operating System(s) • Administrative Features • Document Tree Features • Communications Features • Advanced HTTP Features • Scripting • Security Features • Miscellaneous Server Features	
Specific Servers	68
Installing and Configuring Apache for Unix Systems	68
System Requirements for Apache • Downloading Apache • Compiling Apache • Setting Up the Server and Document Roots • Basic Configuration • Starting and Stopping Apache • Customizing Apache • Configuring General Settings with <i>httpd.conf</i> • Configuring the Document Tree with <i>srm.conf</i> • Customizing Features in Individual Directories • Image Maps, CGI Scripts, and Server-Side Includes • Optional Apache Modules • Virtual Hosts on an Apache Server • Arranging for Apache <i>httpd</i> to Be Started Automatically • Managing Apache's Log Files • Monitoring Performance	
BOX: Improving Server Performance	116
Installing and Configuring WebSite for Windows 95/NT	119
System Requirements • Installing WebSite • The WebSite Server Root • Administering the WebSite Server • Customizing WebSite • WebSite Administration	
BOX: Analyzing Server Log Files	148
Installing and Configuring WebSTAR for Macintosh	150
WebSTAR System Requirements • Installing WebSTAR • Customizing a WebSTAR Site • Other WebSTAR Administrative Tasks	
BOX: Web Robots	176

<i>Chapter 4</i>	<i>Web Security</i>	179
	Planning the Security at Your Site	180
	Basic Security Measures	180
	Securing the Host Machine • Turning Off Insecure Web Server Features	
	• Other Basic Security Techniques	
	BOX: There's a Hole in My Server	184
	Web Server Security Features	185
	Some Security Scenarios	
	How Access Control Works	188
	Configuring Access Control	190
	Access Control in the Unix Apache Server • Access Control in WebSite for	
	Windows NT/95 • Access Control in WebSTAR for Macintosh	
	BOX: Keeping Log Files Private	202
	Running a Web Server in a Network with a Firewall	215
	How Firewalls Work • Providing Internet Service to Web Browsers Within the	
	Inner Network • SOCKS • Providing Web Services to Browsers in the Outside	
	World • Configuring the Apache Server to Act as a Caching Proxy	
	BOX: Running a Unix Server in a Change Root Environment ..	227
	Running an Encrypting Web Server	228
	How Encryption Works • Encryption Implementations • ApacheSSL	
 <i>Chapter 5</i>	 <i>Creating Hypertext Documents</i>	 241
	Basic HTML Tags	241
	Header Tags • Paragraphs and Line Breaks • Creating Hypertext Links	
	• Adding In-Line Images to Pages • The ALT Attribute	
	• Combining Links with Images	
	BOX: HTML and SGML	245
	BOX: A Brief History of HTML	251
	Advanced Tags	252
	Controlling Text Styles	252
	Logical Styles • Physical Styles • Controlling Font Size and Color	
	Special Text Characters	256
	Control over Paragraph Formatting	256
	Alignment • Special-Purpose Paragraph Styles • Preformatted Text	
	• Horizontal Rules	

Lists	260
Ordered Lists • Unordered Lists • Adjusting the Appearance of Ordered and Unordered Lists • Definition Lists	
Controlling the Size and Position of In-Line Images	266
Image Size • Image Alignment • Changing the Amount of Space Around an In-Line Image • Image Borders • Image Attributes for Clickable Image Maps	
Controlling the Global Appearance of the Document	270
Changing the Page's Background Color and Pattern • Changing the Color of Text and Links	
Tables	271
Miscellaneous HTML Tags	275
Comments • ISINDEX • BASE • META • LINK • Tags for Defining Fill-Out Forms and Clickable Image Maps	
Putting It All Together	277
Netscape-Specific HTML Extensions	283
Controlling Line Breaks • New Special Character Mnemonics	
Frames	283
Laying Out a Frameset • Creating the Contents of a Frameset • Targeting Frames	
Microsoft Internet Explorer-Specific HTML Extensions	292
BGSOUND • MARQUEE • IMG	
BOX: HTML Specifications	295
 <i>Chapter 6</i>	
<i>Software Tools for Text, Graphics, Sound, and Video</i>	297
HTML Editors	297
HTML Macros • WYSIWYG HTML Editors	
HTML Syntax Checkers	306
Htmlchek • Weblint • MOMSpider • The Online HTML Validation Service	
Converting Other Text Formats into HTML	309
Microsoft Word/RTF • WordPerfect • FrameMaker • LaTeX • troff • Other Word Processors • Electronic Mailboxes • Plain Text	
Using Graphic Images in Your Pages	318
Graphics Conversion Software for UNIX Systems • Graphics Conversion Software for MS-DOS • Graphics Conversion Software for Windows NT/95 • Graphics Conversion Software for Macintosh • Common Graphics Manipulation Tasks • Adobe Portable Document Format	

BOX: The Guff on GIF	323
Using Sound in Your Pages	340
Converting Sound Formats • Sound Converters for MS-DOS and Windows • Sound Converters for the Apple Macintosh • Sound Converters for UNIX Systems • Sound Creation and Editing Software • Sound Editors for MS-DOS and Windows • Sound Editors for the Macintosh	
Using Animation in Your Pages	345
Animation Formats	
VRML	351
VRML Plug-Ins • Unix VRML Viewers • Macintosh VRML Viewers • Windows VRML Viewers • Creating VRML Models from Scratch • Windows 95/NT • Macintosh • More Information	
 <i>Chapter 7</i> <i>A Web Style Guide</i>	 359
Lost in Hyperspace	359
To Hyper or Not to Hyper • In Hyperspace, Which Way Is Up? • Creating a Sense of Time and Place	
Making the Most of Your Pages	372
Bringing the Important Points Up to the Front • Drawing the Reader's Eye • Page Length • Working with HTML, Not Against It	
Making Hypertext Links Meaningful	378
Links to Graphics, Sounds, and Other Nontext Documents	380
Handling In-Line Graphics	383
Using Tables and Frames Effectively	385
Clickable Image Maps: Uses and Abuses	388
Optimizing Performance	391
Testing Your Pages	393
Managing a Changing Site	394
Under Construction • Working Off-Line • Concurrent Version Management	
Mirroring Other Sites	396
The Web and Copyrights	398
Respecting the Copyright of Others • Protecting Your Own Copyright	

Chapter 8	<i>Working with Server Scripts</i>	401
	Script Basics	401
	Communicating with Scripts	402
	Passing Parameters to Scripts in the URL • Passing Document Path Information to Scripts • Keyword Searches and Fill-Out Forms • Creating a Searchable Document • Creating an Input Document with Fill-Out Forms • Example: Creating a User Feedback Form	
	BOX: The CGI Interface	406
	Creating Clickable Image Maps	420
	Server-Side Image Maps • Client-Side Image Maps • Combining Server-Side and Client-Side Image Maps • Figuring Out Image Map Coordinates	
	Gateways to Other Services	430
	Remote Versus Local Gateways • A Gateway to <i>finger</i> • A Gateway to <i>Archie</i> • A Gateway to E-Mail • Mail Gateways for Macintosh and Windows Servers • Mailmerge, a Flexible E-Mail Gateway and Guestbook Generator	
	Fast Text-Based Searching for Documents at Your Site	439
	Obtaining and Compiling SWISH • Configuring SWISH • Creating the Index • Searching the Index from the Command Line • Creating Multiple Indexes • Searching SWISH over the Web • Configuring <i>wwwais</i> • Testing <i>wwwais</i> • Building a Custom Search Form with <i>wwwais</i> • Bold-Facing the Matched Words in Retrieved Documents • SWISH for Windows • Building a Custom Search Front Form with <i>Webfind</i> • Keeping Your Index Up to Date • Other Text Search Engines	
	Other Gateway Scripts	457
	Gateways to the Sybase Relational Databases • A Gateway to Oracle • Gateways to the <i>mSQL</i> Database • Gateways to Macintosh Databases • A Mailing List Gateway	
	BOX: Letting External Viewers Do the Work	460
	Simple Scripting with Server-Side Includes	461
	<i>echo</i> • <i>include</i> • <i>fsize</i> and <i>flastmod</i> • <i>exec</i> • <i>daycnt</i> , <i>totcnt</i> , and <i>lastzero</i> • <i>config</i> • Server-Side Includes on the Macintosh	
Chapter 9	<i>Writing Server Scripts</i>	467
	Introduction	467
	BOX: A Whirlwind Introduction to Perl	469
	Basic Scripts	472
	A CGI Calendar	